ELSEVIER
NORTH
HOLLAND

Handbook
of
THE HISTORY
OF LOGIC

Volume 7

LOGIC AND THE MODALITIES
IN THE TWENTIETH CENTURY

Edited by

Dov M. Gabbay
John Woods

Handbook of the History of Logic

Volume 7
Logic and the Modalities
in the Twentieth Century

Handbook of the History of Logic

Volume 7
Logic and the Modalities
in the Twentieth Century

ELSEVIER
NORTH
HOLLAND

Handbook of the History of Logic

Volume 7
Logic and the Modalities
in the Twentieth Century

Edited by

Dov M. Gabbay
Department of Computer Science
King's College London
Strand, London, WC2R 2LS, UK

and

John Woods
Philosophy Department
University of British Columbia
Vancouver, BC Canada, V6T 1Z1
and
Department of Computer Science
King's College London
Strand, London, WC2R 2LS, UK
and
Department of Philosophy
University of Lethbridge
Lethbridge, Canada

ELSEVIER
NORTH
HOLLAND

Amsterdam - Boston - Heidelberg - London - New York - Oxford - Paris
San Diego - San Francisco - Singapore - Sydney - Tokyo

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons
or property as a matter of products liability, negligence or otherwise, or from any use
or operation of any methods, products, instructions or ideas contained in the material
herein. Because of rapid advances in the medical sciences, in particular, independent
verification of diagnoses and drug dosages should be made

For information on all Elsevier publications
visit our website at books.elsevier.com

CONTENTS

This Page is Intentionally Left Blank

PREFACE

In its traditional sense, a modal logic is one whose logical vocabulary contains the modal expressions "possibly", "necessarily" and "contingently", construed as sentence operators. If the first half of the twentieth century can lay fair claim to having produced the deep and definitive accounts of classical logic, perhaps the dominant achievement in the next quarter century was the attainment of a firm semantic grip on a hefty plurality of modal systems, marked by the seminal works of Hintikka, Kanger and Kripke. The semanticizing of modal sentences — apart from the importance intrinsic to such an achievement — opened up an important tension between modal and classical logics. Perhaps the most significant difference is that, whereas classical systems are extensional, modal setups are intensional, a happenstance which various philosophers of logic have greeted with suspicion and — in some cases — incredulity. Some of the skeptics — Quine and Harman are two — raised doubts about whether a modal system could have the *bona fides* of a genuine logic. This notwithstanding, the great burst of energy in the modal research programmes since the 1950s has proved irresistible, and the central semantic idea of accessibility relations on alternative possible worlds has had a philosophical influence well beyond the confines of logic, especially in the philosophy of language.

One of the byproducts of the modal groundswell is that there are a great many more interpreted systems of the modals "possibly" and "necessarily" than there are different meanings of these terms in ordinary English. It is easy to see that they are ambiguous in English, that "possibly" encompasses the quite different senses of logical, physical, causal, and practical possibility (ditto "necessarily"). But the sheer plurality of logical systems in which these terms are centrally at issue greatly exceeds this rather modest number of ordinary-language meanings. It is not wholly clear how to understand this proliferation. One possibility is that logic has a greater capacity to identify different concepts of possibility than do native speakers of languages such as English. It is also possible that the multiplication of heretofore unrecognized concepts of possibility is more a matter of the free creation of the theoretical logician. Whatever is to be said for these and other options, it is safe to say that, in having taken the modal turn, logic took on a more experimental character than was evident in the classical heyday.[1] Here, too,

---

[1]Intimations of the experimental proclivities of modal logicians are evident in Aristotle's modal logic, of which there are up to five distinct treatments of logical necessity. Then, too, the stream of systems produced by C.I. Lewis from 1912 into the 1930s encompasses vastly different axiomatizations.

it is not entirely clear what to make of this. Of the golden age of classical logic it can be said with some confidence that logicians took themselves to be doing one of two things. Taking the implication relation as an example, either they were formalizing the pre-existing concept of implication or they were originating a concept designed to facilitate some larger purpose, such as the construction of logically precise languages adequate for science or capable of supporting the reductive burdens of logicism. Part of the answer may be that the model theoretic apparatus needed for the interpretation of modal systems is more complex, and admits of greater recombinations of its elements, than do the standard models of classical logic. Accordingly, it may be more natural for the modal theorist to reconfigure a possible worlds semantics and wait for the kind of, e.g., implication relation it embeds to "fall out".

We should take care not to over-press the contrast between analyzing pre-existing concepts and fashioning new ones. The distinction is present in Kant's pre-critical writings, and persists in the works of his maturity. Kant saw analysis as *making concepts clear*, a job for philosophy. Synthesis was the business of *making clear concepts*, a task that falls to the mathematician. Since logic's great classical interlude arose from the contributions of philosophers and mathematicians alike, we cannot say, especially in the aftermath of the paradox that dethroned intuitive (or "analytic") set theory, that classical logic is synthesis-free. Far from it. What is more, apart from the local disputes within classical logic itself, there were early rivals, such as intuitionism. Even so, classical logic has not been especially pluralistic, whereas modal logic is vigorously so. It is moreover a rather pacific and non-antagonistic pluralism, which attests further to its readiness to view logic as the exploration of mathematically interesting languages and model theoretic structures, with a focus that is a good deal less analytical or philosophical than most classical variations on classical semantics.

It also bears on this issue that the semantics of some of the 20th century's earliest axiom systems — Lewis' S2 and S3 for example — must stretch themselves beyond ordinary recognition in order to keep up with the axioms. S2 and S3 cannot be semanticized in a normal worlds approach; so nonnormal worlds were postulated (more experimentation still), giving inadvertent anticipation of somewhat later paraconsistent developments. A further case in point is the attempt by relevance logicians to impose relevance constraints on the implication relation, so as to evade the classical theorem that everything whatever is implied by an inconsistency. In most relevant approaches, the disjunctive syllogism rule is demoted from a valid to a merely admissible rule. On its face a rather slight adjustment, this actually strips these logics of the truth-functional character of their classical predecessors. More intensionality still.

An attraction of the traditional alethic modals is the ease with which they bear new interpretations in the breakthrough work of Hintikka, von Wright and others in epistemic and deontic logic. These were significant developments twice-over. On the one hand, the new logics of knowledge and belief, and of obligation and permission, were able to retain much of the syntactic and semantic machinery of

their alethic predecessors, showing that all these logics are to a degree variations of each other. On the other hand, however, the logics of knowledge and obligation had taken yet another step away from classical logic. Not only are these newer logics intensional and more experimentally oriented than their classical vis-à-vis, but there is now the looming presence of *agents* operating *in time*. We say "looming" rather than "overt" inasmuch as neither agents nor times are much developed, if at all, in the semantics of these particular systems.

The ephemeral presence of agents and times is important in another respect. It indicates that logic was developing in ways that would satisfy a broader interpretation of the adjective "modal". If one were to consult *The New Oxford Dictionary of English*, corrected reprint, 2001, it would be seen that in the entry for modal logic the first reference is not to a logic of possibility and necessity, but rather to a logic in which sentences are subject to "some qualification". If, then, we were to accept the trichotomy of the basic modes of language introduced by the linguist C.W. Morris — the trichotomony between syntax, semantics and pragmatics — we would be reminded that a pragmatic approach to language is one that takes expressly into account the roles of language-users and the contexts in which they operate. By these lights, the developments in the 1950s and 1960s within the epistemic and deontic adaptations of alethic modal logic mark the transition of logic from a purely syntacto-semantic enterprise, to a "Morrisean" enterprise in which agents, times and situations have a load-bearing role. This was the pragmatic turn in logic.[2] In the broad sense of "modal", a pragmatic logic counts as modal, a happenstance that the Editors have allowed themselves to be guided by in organizing the present volume. In addition to chapters on the traditional modal logics, their epistemic and deontic variations and relevant logic, there are chapters on systems in which the *times* of utterance are taken note of, in which temporal *change* is tracked, in which an utterer's *situation* is taken into account, in which *interpersonal* utterance is acknowledged, and in which agents *compete* with one another in the furtherance of their interests. In each case, the sentences of the logic are modified by these other factors — time, change, agents, situations, dialogue roles, and procedural strategies. Modal logics in the broad sense reflect another change in logic's conception of itself. In the classical heyday, logicians were preoccupied with the analysis of properties (such as implication and logical truth) of abstractly linguistic constructions or of linguistic artifacts in relation to abstractly set theoretic structures. If such logicians gave any thought to the ins-and-outs of human reasoning in the here and now, it was much the received view that the classical laws were also norms of reasoning, albeit in a highly idealized form. Even so, the attention to reasoning was at best an afterthought.[3]

With the rise of modern modal logic, the emphasis began to shift. Under press

---

[2]The word "pragmatic" invites confusion. In its logico-epistemological sense, it is the Quinean doctrine that no principle of logic is immune from overthrow. In its linguistic meaning, it is a logic that takes express note of the role of linguistic agents. The second sense is intended here.

[3]The pretensions of so-called natural deduction systems to be more "natural" than axiomatic systems reflected rather more a distrust of the epistemic privilege that logicians sought to extend to their axioms than to a burning interest in getting reasoning on the ground right.

of developments in computer science and argumentation theory (chiefly dialogue logic), logic started a shift toward a greater emphasis on reasoning. What we find in the chapters of this volume is an attempt, to the extent possible, to lodge pragmatic developments affecting agents and situations in the methodology and principal attainments of the classical analyses of implication and the like. No one thinks that modalizing the implication relation either narrowly or broadly will leave the classical analysis untouched. But, for the most part, there is a widespread desire on the part of modal logicians to retain as much of classical logic as comports with their modal ambitions. What we see in the proliferation of modal logics is not, therefore, a revolution in logic but a development. It is a development very much in progress as we write. But it is already wholly clear that it has broken the research programme in logic wide-open, and has given rise to questions and challenges that are not likely to be settled with any definiteness for some time to come.

Once again the Editors are deeply and most gratefully in the debt of the volume's able authors. The Editors also warmly thank the following persons: Professor Margaret Schabas, Head of the Philosophy Department, and Professor Nancy Gallini, Dean of the Faculty of Arts, at the University of British Columbia; Professor Bryson Brown, Chair of the Philosophy Department and his successor Michael Stingl, and Professor Christopher Nicol, Dean of the Faculty of Arts and Science, at the University of Lethbridge; Professor Alan Gibbons, Head of the Computer Science Department, and his successor Andrew Jones, at King's College London; Jane Spurr, Publications Administrator in London; Dawn Collins and Carol Woods, Production Associates in Lethbridge and Vancouver, respectively; and our colleagues at Elsevier, Senior Publisher, Arjen Sevenster, and Production Associate, Andy Deelen.

<div align="right">

Dov M. Gabbay
King's College London

John Woods
University of British Columbia
and
King's College London
and
University of Lethbridge

</div>

CONTRIBUTORS

Keith Devlin
CSLI, Stanford University, 210 Panama Street, Stanford, CA 94305-4115, USA.
devlin@csli.stanford.edu

Jan van Eijck
CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.
jve@cwi.nl

Dov M. Gabbay
Department of Computer Science, King's College London, Strand, London WC2R 2LS, UK
dg@dcs.kcl.ac.uk

Paul Gochet
Université de Liège, Institut Montefiore, B28, B-4000 Liège, Belgium.
pgochet@ulg.ac.be

Robert Goldblatt
School of Mathematics, Statistics and Computer Science, Victoria University, P. O. Box 600, Wellington, New Zealand.
Rob.Goldblatt@vuw.ac.nz

Pascal Gribomont
Université de Liège, Institut Montefiore, B28, B-4000 Liège, Belgium.
gribomont@montefiore.ulg.ac.be

Per F. V. Hasle
Department of Communication
Aalborg University, Denmark.
phasle@hum.aau.dk

Erik C. W. Krabbe
Department of Theoretical Philosophy, Faculty of Philosophy, Rijksuniversiteit Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands.
E.C.W.Krabbe@rug.nl

Paul McNamara
Philosophy Department, Hamilton Smith Hall, University of New Hampshire, Durham, NH 03824, USA.
paulm@cisunix.unh.edu

Peter Øhrstrøm
Department of Communication, Aalborg University, Denmark.
poe@hum.aau.dk

Greg Restall
Department of Philosophy, University of Melbourne, Victoria 3010, Australia.
restall@unimelb.edu.au

Martin Stokhof
ILLC / Department of Philosophy, Faculty of Humanities, Universiteit van Amsterdam,
Nieuwe Doelenstraat 15, 1012 CP Amsterdam, The Netherlands.
m.j.b.stokhof@uva.nl

John Woods
Philosophy Department, University of British Columbia, Vancouver, British Columbia
Canada, V6T 1Z1
jhwoods@interchange.ubc.ca

MATHEMATICAL MODAL LOGIC:
A VIEW OF ITS EVOLUTION

Robert Goldblatt

> ...there is no one fundamental logical notion of *necessity*, nor consequently of *possibility*. If this conclusion is valid, the subject of modality ought to be banished from logic, since propositions are simply true or false ...
> [Russell, 1905]

1 INTRODUCTION

Modal logic was originally conceived as the logic of *necessary* and *possible* truths. It is now viewed more broadly as the study of many linguistic constructions that qualify the truth conditions of statements, including statements concerning knowledge, belief, temporal discourse, and ethics. Most recently, modal symbolism and model theory have been put to use in computer science, to formalise reasoning about the way programs behave and to express dynamical properties of transitions between states.

Over a period of three decades or so from the early 1930's there evolved two kinds of mathematical semantics for modal logic. *Algebraic* semantics interprets modal connectives as operators on Boolean algebras. *Relational* semantics uses relational structures, often called *Kripke models*, whose elements are thought of variously as being possible worlds, moments of time, evidential situations, or states of a computer. The two approaches are intimately related: the subsets of a relational structure form a modal algebra (Boolean algebra with operators), while conversely any modal algebra can be embedded into an algebra of subsets of a relational structure via extensions of Stone's Boolean representation theory. Techniques from both kinds of semantics have been used to explore the nature of modal logic and to clarify its relationship to other formalisms, particularly first and second order monadic predicate logic.

The aim of this article is to review these developments in a way that provides some insight into how the present came to be as it is. The pervading theme is the *mathematics* underlying modal logic, and this has at least three dimensions. To begin with there are the new mathematical ideas: when and why they were

introduced, and how they interacted and evolved. Then there is the use of methods and results from other areas of mathematical logic, algebra and topology in the analysis of modal systems. Finally, there is the application of modal syntax and semantics to study notions of mathematical and computational interest.

There has been some mild controversy about priorities in the origin of relational model theory, and space is devoted to this issue in section 4. An attempt is made to record in one place a sufficiently full account of what was said and done by early contributors to allow readers to make their own assessment (although the author does give his).

Despite its length, the article does not purport to give an encyclopaedic coverage of the field. For instance, there is much about temporal logic (see [Gabbay *et al.*, 1994]) and logics of knowledge (see [Fagin *et al.*, 1995]) that is not reported here, while the surface of modal *predicate* logic is barely scratched, and proof theory is not discussed at all. I have not attempted to survey the work of the present younger generation of modal logicians (see [Chagrov and Zakharyaschev, 1997], [Kracht, 1999], and [Marx and Venema, 1997], for example). There has been little by way of historical review of work on intensional semantics over the last century, and no doubt there remains room for more.

Several people have provided information, comments and corrections, both historical and editorial. For such assistance I am grateful to Wim Blok, Max Cresswell, John Dawson, Allen Emerson, Saul Kripke, Neil Leslie, Ed Mares, Robin Milner, Hiroakira Ono, Amir Pnueli, Lawrence Pedersen, Vaughan Pratt, Colin Stirling and Paul van Ulsen.

This article originally appeared as [Goldblatt, 2003c]. As well as corrections and minor adjustments, there are two significant additions to this version. The last part of section 6.6 has been rewritten in the light of the discovery in 2003 of a solution of what was described in the first version as a "perplexing open question". This was the question of whether a logic validated by its canonical frame must be characterised by a first-order definable class of frames. Also, a new section 7.7 has been added to describe recent work in theoretical computer science on modal logics for "coalgebras".

2   BEGINNINGS

2.1   *What is a Modality?*

Modal logic began with Aristotle's analysis of statements containing the words "necessary" and "possible".[1] These are but two of a wide range of *modal connectives*, or *modalities* that are abundant in natural and technical languages. Briefly, a modality is any word or phrase that can be applied to a given statement S to create a new statement that makes an assertion about the *mode of truth* of S:

---

[1] For the early history of modal logic, including the work of Greek and medieval scholars, see [Bochenski, 1961] and [Kneale and Kneale, 1962]. The Historical Introduction to [Lemmon and Scott, 1966] gives a brief but informative sketch.

about when, where or how S is true, or about the circumstances under which S may be true. Here are some examples, grouped according to the subject they are naturally associated with

tense logic:	henceforth, eventually, hitherto, previously, now, tomorrow, yesterday, since, until, inevitably, finally, ultimately, endlessly, it will have been, it is being ...
deontic logic:	it is obligatory/forbidden/permitted/unlawful that
epistemic logic:	it is known to X that, it is common knowledge that
doxastic logic:	it is believed that
dynamic logic:	after the program/computation/action finishes, the program enables, throughout the computation
geometric logic:	it is locally the case that
metalogic:	it is valid/satisfiable/provable/consistent that

The key to understanding the *relational* modal semantics is that many modalities come in dual pairs, with one of the pair having an interpretation as a universal quantifier ("in all...") and the other as an existential quantifier ("in some..."). This is illustrated by the following interpretations, the first being famously attributed to Leibniz (see section 4).

necessarily	in all possible worlds
possibly	in some possible world
henceforth	at all future times
eventually	at some future time
it is valid that	in all models
it is satisfiable that	in some model
after the program finishes	after all terminating executions
the program enables	there is a terminating execution such that

It is now common to use the symbol □ for a modality of universal character, and ◇ for its existential dual. In systems based on classical truth-functional logic, □ is equivalent to ¬◇¬, and ◇ to ¬□¬, where ¬ is the negation connective. Thus "necessarily" means "not possibly not", "eventually" means "not henceforth not", a statement is valid when its negation is not satisfiable, etc.

Notation

Rather than trying to accommodate all the notations used for truth-functional connectives by different authors over the years, we will fix on the symbols ∧, ∨, ¬, → and ↔ for conjunction, disjunction, negation, (material) implication, and (material) equivalence. The symbol ⊤ is used for a constant true formula, equivalent to any tautology, while ⊥ is a constant false formula, equivalent to ¬⊤. We also use ⊤ and ⊥ as symbols for truth values.

The standard syntax for propositional modal logic is based on a countably infinite list $p_0, p_1, \ldots$ of *propositional variables*, for which we typically use the letters p, q, r. Formulas are generated from these variables by means of the above connectives and the symbols $\Box$ and $\Diamond$. There are of course a number of options about which of these to take as primitive symbols, and which to define in terms of primitives. When describing the work of different authors we will sometimes use their original symbols for modalities, such as M for *possibly*, L or N for *necessarily*, and other conventions for deontic and tense logics.

The symbol $\Box^n$ stands for a sequence $\Box\Box\cdots\Box$ of n copies of $\Box$, and likewise $\Diamond^n$ for $\Diamond\Diamond\cdots\Diamond$ (n times).

A systematic notation will also be employed for Boolean algebras: the symbols $+, \cdot, -$ denote the operations of sum (join), product (meet), and complement in a Boolean algebra, and 0 and 1 are the greatest and least elements under the ordering $\leq$ given by $x \leq y$ iff $x \cdot y = x$. The supremum (sum) and infimum (product) of a set X of elements will be denoted $\sum X$ and $\prod X$ (when they exist).

2.2 MacColl's Iterated Modalities

The first substantial algebraic analysis of modalised statements was carried out by Hugh MacColl, in a series of papers that appeared in *Mind* between 1880 and 1906 under the title *Symbolical Reasoning*,[2] as well as in other papers and his book of [1906]. MacColl symbolised the conjunction of two statements a and b by their concatenation ab, used $a+b$ for their disjunction, and wrote $a : b$ for the statement "a implies b", which he said could be read "if a is true, then b must be true", or "whenever a is true, b is also true". The equation $a = b$ was used for the assertion that a and b are equivalent, meaning that each implies the other. Thus $a = b$ is itself equivalent to the "compound implication" $(a : b)(b : a)$, an observation that was rendered symbolically by the equation $(a = b) = (a : b)(b : a)$.

MacColl wrote a' for the "denial" or "negative" of statement a, and stated that $(a' + b)'$ is equivalent to ab'. However, while $a' + b$ is a "necessary consequence" of $a : b$ (written $(a : b) : a' + b$), he argued that the two formulas are not equivalent because their denials are not equivalent, claiming that the denial of $a : b$ "only asserts the *possibility* of the combination ab' ", while the denial of $a' + b$ "asserts the *certainty* of the same combination".[3]

Boole had written $a = 1$ and $a = 0$ for "a is true" and "a is false", giving a temporal reading of these as *always true* and *always false* respectively [Boole, 1854, ch. XI]. MacColl invoked the letters ϵ and η to stand for certainty and impossibility, initially describing them as replacements for 1 and 0, and then introduced a third letter θ to denote a statement that was neither certain nor impossible, and hence

---

[2]A listing of these papers is given in the Bibliography of [Lewis, 1918] and on p. 132 of Church's bibliography in volume 1 of *The Journal of Symbolic Logic*. A comprehensive bibliography of MacColl's works is given in [Astroh and Klüwer, 1998].

[3]This appears to conflict with his earlier claim that the denial of $a' + b$ is equivalent to ab'. "Actuality" may be a better word than "certainty" to express what he meant here (see [MacColl, 1880, p. 54].

was "a *variable* (neither always true nor always false)". He wrote the equations $(a = \epsilon)$, $(b = \eta)$ and $(c = \theta)$ to express that a is a certainty, b is an impossibility, and c is a variable. Then he changed these to the symbols a^ϵ, b^η, c^θ, and went on to write a^τ for "a is true" and a^ι for "a is false", noting that a true statement is "not necessarily a certainty" and a false one is "not necessarily impossible". In these terms he stated that $a : b$ is equivalent both to $(a.b')^\eta$ ("it is impossible that a and not b") and to $(a' + b)^\epsilon$ ("it is certain that either not a or b").

Once the step to this superscript notation had been taken, it was evident that it could be repeated, giving an easy notation for iterations of modalities. MacColl gave the example of $A^{\eta\iota\epsilon\epsilon}$ as "it is certain that it is certain that it is false that it is impossible that A", abbreviated this to "it is certain that a is certainly possible", and observed that

> Probably no reader—at least no English reader, born and brought up in England—can go through the full unabbreviated translation of this symbolic statement $A^{\eta\iota\epsilon\epsilon}$ into ordinary speech without being forcibly reminded of a certain nursery composition, whose ever-increasing accumulation of *thats* affords such pleasure to the infantile mind; I allude, of course, to "The House that Jack Built". But trivial matters in appearance often supply excellent illustrations of important general principles.[4]

There has been a recent revival of interest in MacColl, with a special issue of the *Nordic Journal of Philosophical Logic*[5] devoted to studies of his work. In particular the article [Read, 1998] analyses the principles of modal algebra proposed by MacColl and argues that together they correspond to the modal logic T, later developed by Feys and von Wright, that is described at the end of section 2.4 below.

2.3   The Lewis Systems

MacColl's papers are similar in style to earlier nineteenth century logicians. They give a descriptive account of the meanings and properties of logical operations but, in contrast to contemporary expectations, provide neither a formal definition of the class of formulas dealt with nor an axiomatisation of operations in the sense of a rigorous deduction of theorems from a given set of principles (axioms) by means of explicitly stated rules of inference. The first truly modern formal axiom systems for modal logic are due to C. I. Lewis, who defined five different ones, S1–S5, in Appendix II of the book *Symbolic Logic* [1932] that he wrote with C. H. Langford. Lewis had begun in [1912, p. 522] with a concern that

> the expositors of the algebra of logic have not always taken pains to indicate that there is a difference between the algebraic and ordinary meanings of implication.

---

[4] *Mind* (New Series), vol. 9, 1900, p. 75.
[5] Volume 3, number 1, December 1998, available at
`http://www.hf.uio.no/filosofi/njpl/vol3no1/index.html`.

He observed that the algebraic meaning, as used in the *Principia Mathematica* of Russell and Whitehead, leads to the "startling theorems" that a false proposition implies any proposition, and a true proposition is implied by any proposition. These so-called *paradoxes of material implication* take the symbolic forms

$$\neg\alpha \rightarrow (\alpha \rightarrow \beta)$$
$$\alpha \rightarrow (\beta \rightarrow \alpha).$$

For Lewis the ordinary meaning of "α implies β" is that β can be *validly inferred*[6] from α, or is *deducible*[7] from α, an interpretation that he considered was not subject to these paradoxes. Taking "α implies β" as synonymous with "either not-α or β", he distinguished *extensional* and *intensional* meanings of disjunction, providing two meanings for "implies". Extensional disjunction is the usual truth-functional "or", which gives the *material* (algebraic) implication synonymous with "it is false that α is true and β is false". Intensional disjunction

is such that at least one of the disjoined propositions is "necessarily" true.[8]

That reading gives Lewis' "ordinary" implication, which he also dubbed "strict", meaning that "it is *impossible* (or *logically inconceivable*[9]) that α is true and β is false".

The system of Lewis's book *A Survey of Symbolic Logic* [1918] used a primitive *impossibility* operator to define strict implication. This later became the system S3 of [Lewis and Langford, 1932], which introduced instead the symbol $\diamond$ for possibility, but Lewis decided that he wished S2 to be regarded as the correct system for strict implication. The systems were defined with negation, conjunction, and possibility as their primitive connectives, but he made no use of a symbol for the dual combination $\neg\diamond\neg$.[10] For strict implication the symbol $\dashv3$ was used, with $\alpha \dashv3 \beta$ being a definitional abbreviation for $\neg\diamond(\alpha \wedge \neg\beta)$. Strict equivalence $(\alpha = \beta)$ was defined as $(\alpha \dashv3 \beta) \wedge (\beta \dashv3 \alpha)$.

Here now are definitions of S1–S5 in Lewis's style, presented both to facilitate discussion of later developments and to convey some of the character of his

---

[6][Lewis, 1912, p. 527]
[7][Lewis and Langford, 1932, p. 122]
[8][Lewis, 1912, p. 523]
[9][Lewis and Langford, 1932, p. 161]
[10]The dual symbol $\square$ was later devised by F. B. Fitch and first appeared in print in 1946 in a paper of R. Barcan. See footnote 425 of [Hughes and Cresswell, 1968, fn. 425].

approach. System S1 has the axioms[11]

$$(p \wedge q) \strictif (q \wedge p)$$
$$(p \wedge q) \strictif p$$
$$p \strictif (p \wedge p)$$
$$((p \wedge q) \wedge r) \strictif (p \wedge (q \wedge r))$$
$$((p \strictif q) \wedge (q \strictif r)) \strictif (p \strictif r)$$
$$(p \wedge (p \strictif q)) \strictif q,$$

where p, q, r are propositional variables, and the following rules of inference.

- *Uniform substitution* of formulas for propositional variables.

- *Substitution of strict equivalents*: from $(\alpha = \beta)$ and γ infer any formula obtained from γ by substituting β for some occurrence(s) of α.

- *Adjunction*: from α and β infer $\alpha \wedge \beta$.

- *Strict detachment*: from α and $\alpha \strictif \beta$ infer β.[12]

System S2 is obtained by adding the axiom $\Diamond(p \wedge q) \strictif \Diamond p$ to the basis for S1. S3 is S1 plus the axiom $(p \strictif q) \strictif (\neg \Diamond q \strictif \neg \Diamond p)$. S4 is S1 plus $\Diamond \Diamond p \strictif \Diamond p$, or equivalently $\Box p \strictif \Box \Box p$. S5 is S1 plus $\Diamond p \strictif \Box \Diamond p$.

   The axioms for S4 and S5 were first proposed for consideration as further postulates in a paper of Oskar Becker [1930]. His motivation was to find axioms that reduced the number of logically non-equivalent combinations that could be formed from the connectives "not" and "impossible". He also considered the formula $p \strictif \neg \Diamond \neg \Diamond p$, and called it the "Brouwersche axiom". The connection with Brouwer is remote: if "not" is translated to "impossible" ($\neg \Diamond$), and "implies" to its strict version, then the intuitionistically acceptable principle $p \rightarrow \neg \neg p$ becomes the Brouwersche axiom.

2.4   Gödel on Provability as a Modality

Gödel in [1931] reviewed Becker's 1930 article. In reference to Becker's discussion of connections between modal logic and intuitionistic logic he wrote

> It seems doubtful, however, that the steps here taken to deal with this problem on a formal plane will lead to success.

He subsequently took up this problem himself with great success, and at the same time simplified the way that modal logics are presented. The Lewis systems contain all truth-functional tautologies as theorems, but it requires an extensive analysis

---

[11]Originally $p \strictif \neg \neg p$ was included as an axiom, but this was shown to be redundant by McKinsey in 1934.

[12]Lewis used the name "Inference" for the rule of strict detachment. He also used "assert" rather than "infer" in these rules.

to demonstrate this.[13] Such effort would be unnecessary if the systems were defined by directly extending a basis for the standard propositional calculus. That approach was first used in the note "An interpretation of the intuitionistic propositional calculus" [Gödel, 1933], published in the proceedings of Karl Menger's mathematical colloquium at the University of Vienna for 1931–1932. Gödel formalised assertions of *provability* by a propositional connective B (from "beweisbar"), reading $B\alpha$ as "α is provable". He defined a system which has, in addition to the axioms and rules of ordinary propositional calculus, the axioms

$$Bp \rightarrow p,$$
$$Bp \rightarrow (B(p \rightarrow q) \rightarrow Bq),$$
$$Bp \rightarrow BBp,$$

and the inference rule: *from α infer $B\alpha$*. He stated that this system is equivalent to Lewis' S4 when $B\alpha$ is translated as $\Box\alpha$.[14] Then he gave the following two translations of propositional formulas

p	p		p	p
$\neg\alpha$	$\neg B\alpha$		$\neg\alpha$	$B\neg B\alpha$
$\alpha \rightarrow \beta$	$B\alpha \rightarrow B\beta$		$\alpha \rightarrow \beta$	$B\alpha \rightarrow B\beta$
$\alpha \vee \beta$	$B\alpha \vee B\beta$		$\alpha \vee \beta$	$B\alpha \vee B\beta$
$\alpha \wedge \beta$	$\alpha \wedge \beta$		$\alpha \wedge \beta$	$B\alpha \wedge B\beta$

and asserted that in each case the translation of any theorem of Heyting's intuitionistic propositional calculus [15] is derivable in his system, adding that "presumably" the converse is true as well. He also asserted that the translation of $p \vee \neg p$ is not derivable, and that a formula of the form $B\alpha \vee B\beta$ is derivable only when one of $B\alpha$ and $B\beta$ is derivable. Proofs of these claims first appeared in [McKinsey and Tarski, 1948], as is discussed further in section 3.2.

Those familiar with later developments will recognise the pregnancy of this brief note of scarcely more than a page. Its translations provided an important connection between intuitionistic and modal logic that contributed to the development both of topological interpretations and of Kripke semantics for intuitionistic logic. Its ideas also formed the precursor to the substantial branch of modal logic concerned with the modality "it is provable in Peano arithmetic that". We will return to these matters below (see §3.2, 7.5, 7.6).

It is now standard practice to present modal logics in the axiomatic style of Gödel. The notion of *a logic* refers to any set Λ of formulas that includes all truth-functional tautologies and is closed under the rules of uniform substitution for variables and detachment for material implication. The formulas belonging to Λ are the *Λ-theorems*, and are also said to be *Λ-provable*. A logic is called *normal*

---

[13]See [Hughes and Cresswell, 1968, pp. 218–223]

[14]More precisely, he stated that it is equivalent to Lewis's System of Strict Implication supplemented by Becker's axiom $\Box p \dashv 3 \Box\Box p$. It is unlikely that he was aware of the name "S4" at that time.

[15]Heyting published this calculus in 1930.

if it includes Gödel's second axiom, which is usually presented (with $\Box$ in place of B) as

$$\Box(p \to q) \to (\Box p \to \Box q),$$

and has the rule of *Necessitation*: from α infer $\Box\alpha$. S5 can be defined as the normal logic obtained by adding the axiom $p \to \Box\Diamond p$ to Gödel's axiomatisation of S4. Following [Becker, 1930], $p \to \Box\Diamond p$ is called the *Brouwerian* axiom. The smallest normal logic is commonly called K, in honour of Kripke. The normal logic obtained by adding the first Gödel axiom $\Box p \to p$ to K is known as T. That system was first defined by Feys[16] in 1937 by dropping Gödel's third axiom from S4. T is equivalent to the system M of [von Wright, 1951]. The *Brouwerian System* B is the normal logic obtained by adding the Brouwerian axiom to T.

The first formulation of the *non-normal* systems S1–S3 in the Gödel style was made in [Lemmon, 1957], which also introduced a series of systems E1–E5 designed to be "epistemic" counterparts to S1–S5. These systems have no theorems of the form $\Box\alpha$, and in place of Necessitation they have the rule *from $\alpha \to \beta$ infer $\Box\alpha \to \Box\beta$*. Lemmon suggests that they capture the reading of $\Box$ as "it is scientifically but not logically necessary that".

3   MODAL ALGEBRAS

Modern propositional logic began as algebra, in the thought of Boole. We have seen that the same was true for modern modal logic, in the thought of MacColl. By the time that the Lewis systems appeared, algebra was well-established as a postulational science, and the study of the very notion of *an abstract algebra* was being pursued [Birkhoff, 1933; 1935]. Over the next few years, algebraic techniques were applied to the study of modal systems, using *modal algebras*: Boolean algebras with an additional operation to interpret $\Diamond$. During the same period, representation theories for various lattices with operators were developed, beginning with the Stone representation of Boolean algebras [1936], and these were to have a significant impact on semantical studies of modal logic.

3.1   McKinsey and the Finite Model Property

J. C. C. McKinsey in [1941] showed that there is an algorithm for deciding whether any given formula is a theorem of S2, and likewise for S4. His method was to show that if a formula is not a theorem of the logic, then it is falsified by some finite model which satisfies the logic. This property was dubbed the *finite model property* by Ronald Harrop [1958], who proved the general result that any finitely axiomatisable propositional logic Λ with the finite model property is decidable. The gist of Harrop's argument was that finite axiomatisability guarantees that Λ is effectively enumerable, while the two properties together guarantee the same for the complement of Λ. By enumerating the finite models and the formulas, and at

---

[16]Who called it "t".

the same time systematically testing formulas for satisfaction by these models, a list can be effectively generated of those formulas that are falsifed by some finite model which satisfies the axioms of Λ. By the finite model property this is just a listing of all the non-theorems of Λ.

McKinsey actually showed something stronger: the size of a falsifying model for a non-theorem α is bounded above by a number that depends computably on the size of α. Thus to decide if α is a theorem it suffices to generate all finite models up to a prescribed bound. However this did not yield a *feasible* algorithm: the proof for S2 gave an upper bound of $2^{2^{n+1}}$, doubly exponential in the number n of subformulas of α.

McKinsey's construction is worth outlining, since it was an important innovation that has been adapted numerous times to other propositional logics (as he suggested it might be), and has been generalised to other contexts, as we shall see. He used models of the form $(K, D, -, {}^{*}, \cdot)$, called *matrices*, where $-$, *, $\cdot$ are operations on a set K for evaluating the connectives $\neg$, $\diamond$, and $\wedge$, while D is a set of *designated* elements of K. A formula α is *satisfied* by such a matrix if every assignment of elements of K to the variables of α results in α being evaluated to a member of the subset D. These structures abstract from the tables of values, with designated elements, used to define propositional logics and prove the independence of axioms. Their use as a general method for constructing logical systems is due to Alfred Tarski.[17]

A logic is *characterised* by a matrix if the matrix satisfies the theorems of the logic and no other formulas. Structures of this kind had been developed for S2 by E. V. Huntington [1937], who gave the concrete example of K being the class of "propositions" and D the subclass of those that are "asserted" or "demonstrable", describing this subclass as "corresponding roughly to the Frege assertion sign".

A matrix is *normal* if

$$
\begin{array}{rcl}
x, y \in D & \text{implies} & x \cdot y \in D, \\
x, (x \Rightarrow y) \in D & \text{implies} & y \in D, \\
(x \Leftrightarrow y) \in D & \text{implies} & x = y,
\end{array}
$$

where $(x \Rightarrow y) = -^{*}(x \,\cdot\, - y)$ and $(x \Leftrightarrow y) = (x \Rightarrow y) \cdot (y \Rightarrow x)$ are the operations interpreting strict implication and strict equivalence in K. These closure conditions on D are intended to correspond to Lewis' deduction rules of adjunction, strict detachment, and substitution of strict equivalents. In a normal S2-matrix, $(K, -, \cdot)$ is a Boolean algebra in which D is a filter. Hence the greatest element 1 is always designated. McKinsey showed that there exists an infinite[18] normal matrix that characterises S2, using what he described as an unpublished method due to Lindenbaum that was explained to him by Tarski and which applies to any propositional calculus that has the rule of uniform substitution for variables. Taking $(K, -, {}^{*}, \cdot)$ as the algebra of formulas, with $-\alpha = \neg\alpha$, ${}^{*}a = \diamond\alpha$

---

[17]The historical origins of the "matrix method" are described in [Łukasiewicz and Tarski, 1930]. See footnotes on pages 40 and 43 of the English translation of this article in [Tarski, 1956].

[18]Dugundji [1940] had proved that none of S1–S5 has a *finite* characteristic matrix.

and $\alpha \cdot \beta = \alpha \wedge \beta$, and with D as the set of S2-theorems, gives a characteristic S2-matrix which satisfies all but the last normality condition on D. Since that condition is needed to make the matrix into a Boolean algebra, it is imposed by identifying formulas α, β whenever $(\alpha \Leftrightarrow \beta) \in D$. The resulting quotient matrix is the one desired, and is what is now widely known as the *Lindenbaum algebra* of the logic. Its designated elements are the equivalence classes of the theorems.

Now if α is a formula that not an S2-theorem, then there is some evaluation in this Lindenbaum algebra that fails to satisfy α. Let $x_1, \ldots, x_n$ be the values of all the subformulas of α in this evaluation, and let K_1 be the Boolean subalgebra generated by the $n + 1$ elements $x_1, \ldots, x_n, {}^*0$. Then K_1 has at most $2^{2^{n+1}}$ members. Define an element of K_1 to be designated iff it was designated in the ambient Lindenbaum algebra. McKinsey showed how to define an operation ${}^*_1$ on K_1 such that ${}^*_1 x = {}^*x$ whenever x and ${}^*x$ are both in K_1:

$${}^*_1 x = \prod \{{}^*y \in K_1 : x \le y \in K_1\}.$$

The upshot was to turn K_1 into a finite S2-matrix in which the original falsifying evaluation of α can be reproduced.

This same construction shows that S4 has the finite model property, with the minor simplification that the element ^{*0} does not have to be worried about, since ${}^*0 = 0$ in any normal S4-matrix (so the computable upper bound becomes 2^{2^n}). The Lindenbaum algebra for S4 has only its greatest element designated, i.e. D=$\{1\}$, because $(\alpha \dashv \beta) \wedge (\beta \dashv \alpha)$ is an S4-theorem whenever α and β are, putting all theorems into the same equivalence class. This is a fact that applies to any logic that has the rule of Necessitation, and it allows algebraic models for normal logics to be confined to those that just designate 1.

3.2   Topology for S4

Topological interpretations of modalities were given in a paper of Tang Tsao-Chen [1938], which proposed that "the algebraic postulates for the Lewis calculus of strict implication" be the axioms for a Boolean algebra with an additional operation x^∞ having $x^\infty \cdot x = x^\infty$ and $(x \cdot y)^\infty = x^\infty \cdot y^\infty$. The symbol $\diamond$ was used for the dual operation $\diamond x = -(-x)^\infty$. The notation $\vdash x$ was defined to mean that $1^\infty \le x$, and it was shown that $\vdash x$ holds whenever x is any evaluation of a theorem of S2. In effect this says that putting $D = \{x : 1^\infty \le x\}$ turns one of these algebras into an S2-matrix. In fact if $1^\infty = 1$, or equivalently $\diamond 0 = 0$, it also satisfies S4. But S4 was not mentioned in this paper.

A "geometric" meaning was proposed for the new operations by taking x^∞ to be the interior of a subset x of the Euclidean plane, in which case $\diamond x$ is the *topological closure* of x, i.e. the smallest closed superset of x. If the greatest element 1 of the algebra is the whole plane, or any open set, then in that case $1^\infty = 1$, but it is evident that Tang did not intend this, since the paper has a footnote explaining that another geometric meaning of x^∞ can be obtained by letting 1^∞ be some subset of the plane, possibly even a one-element subset, and defining x^∞ to be

$x \cdot 1^{\infty}$. (This construction could be carried out in any Boolean algebra by fixing 1^{∞} arbitrarily.) It appears then that the best way to understand Tang's first geometric meaning is that the ambient Boolean algebra should be the powerset algebra $\mathcal{P}(S)$ of all subsets of some subset S of the Euclidean plane, with "interior" and "closure" being taken in the subspace topology on S.

Now a well-known method, due to Kuratowski, for defining a topology on an arbitrary set S is to give a *closure operation* $X \mapsto \mathbf{C}X$ on subsets X of S, i.e. an operation satisfying $\mathbf{C}\emptyset = \emptyset$, $\mathbf{C}(X \cup Y) = \mathbf{C}X \cup \mathbf{C}Y$ and $X \subseteq \mathbf{C}X = \mathbf{C}\mathbf{C}X$. Then a set X is *closed* iff $\mathbf{C}X = X$, and *open* iff its complement in S is closed. Any topological space can be presented in this way, with $\mathbf{C}X$ being the topological closure of X.

McKinsey and Tarski in [1944] undertook an abstract algebraic study of closure operations by defining a *closure algebra* to be any Boolean algebra with a unary operation $\mathbf{C}$ satisfying Kuratowski's axioms. The operation * on an S4-matrix satisfies these axioms, and McKinsey had shown in his work [1941] on S4 that any *finite* normal S4-matrix can be represented as the closure algebra of all subsets of some topological space, using the representation of a finite Boolean algebra as the powerset algebra of its set of atoms. McKinsey and Tarski now extended this representation to arbitrary closure algebras. Combining the Stone representation of Boolean algebras with the idea of the $^*{}_1$-operation from McKinsey's finite model construction they showed that any closure algebra is isomorphic to a subalgebra of the closure algebra of subsets of some topological space. They gave a deep algebraic analysis of the class of closure algebras, including such results as the following.

1. The closure algebra of any zero-dimensional dense-in-itself subspace of a Euclidean space (e.g. Cantor's discontinuum or the space of points with rational coordinates) includes isomorphic copies of all finite closure algebras as subalgebras.

2. Every finite closure algebra is isomorphic embeddable into the closure algebra of subsets of some *open* subset of Euclidean space.

3. An equation that is satisfied by the closure algebra of any Euclidean space is satisfied by every closure algebra.

4. An equation that is satisfied by all finite closure algebras is satisfied by every closure algebra (this is an analogue of McKinsey's finite model property for S4).

5. If an equation of the form $\mathbf{C}\sigma \cdot \mathbf{C}\tau = 0$ is satisfied by all closure algebras, then so is one of the equations $\sigma = 0$ and $\tau = 0$.

The proof of result (5) involved taking the direct product of two closure algebras that each reject one of the equations $\sigma = 0$ and $\tau = 0$, and then embedding this direct product into another closure algebra that is *well-connected*, meaning that if

x and y are non-zero elements, then $\mathbf{C}x \cdot \mathbf{C}y \neq 0$. The result itself is equivalent to the assertion that if the equation $\mathbf{I}\sigma + \mathbf{I}\tau = 1$ is satisfied by all closure algebras, then so is one of the equations $\sigma = 1$ and $\tau = 1$, where $\mathbf{I} = -\mathbf{C}-$ is the abstract *interior* operator dual to $\mathbf{C}$. This is an algebraic version of one of the facts about S4 stated in [Gödel, 1933] (see later in this section).

In a sequel article [1946], McKinsey and Tarski studied the algebra of *closed* (i.e. $\mathbf{C}x = x$) elements of a closure algebra. These form a sublattice with operations $x \dot{-} y = \mathbf{C}(x \cdot -y)$ and $\ominus x = 1 \dot{-} x = \mathbf{C}-x$. An axiomatisation of these algebras was given in the form of an equational definition of certain *Brouwerian* algebras of the type $(K, +, \cdot, \dot{-}, 1)$, and a proof that every Brouwerian algebra is isomorphic to a subalgebra of the Brouwerian algebra of closed sets of some topological space. Results were proven for Brouwerian algebras that are analogous to results (1)–(5) above for closure algebras, with the analogue of (5) being:

1. If the equation $\sigma \cdot \tau = 0$ is satisfied by all Brouwerian algebras, then so is one of the equations $\sigma = 0$ and $\tau = 0$.

Brouwerian algebras are so named because they provide models of the intuitionistic propositional calculus IPC. This works in a way that is *dual* to the method that has been described for evaluating modal formulas, in that 0 is the unique designated element; $\wedge$ is interpreted as the lattice sum/join operation $+$; $\vee$ is interpreted as lattice product/meet $\cdot$; $\rightarrow$ is interpreted as the operation $\div$ defined by $x \div y = y \dot{-} x$; and $\neg$ is interpreted as the unary operation $x \div 1 = \ominus x$.

The algebra of *open* (i.e. $\mathbf{I}x = x$) elements of a closure algebra also form a sublattice that is a model of intuitionistic logic. It relates more naturally to the Boolean semantics in that 1 is designated and $\wedge$ and $\vee$ are interpreted as $\cdot$ and $+$. Implication is interpreted by the operation $x \Rightarrow y = \mathbf{I}(-x + y) = -\mathbf{C}(x \cdot -y)$ and negation by $-x = x \Rightarrow 0 = \mathbf{I}-x$. This topological interpretation had been developed in the mid-1930's by Tarski [1938] and Marshall Stone [1937–1938] who independently observed that the lattice $\mathcal{O}(S)$ of open subsets of a topological space S is a model of IPC under the operations just described. Tarski took this further to identify a large class of spaces, including all Euclidean spaces, for which $\mathcal{O}(S)$ exactly characterises IPC.

The abstract algebras $(K, +, \cdot, \Rightarrow, 0)$ that can be isomorphically embedded into ones of the type $\mathcal{O}(S)$ form an equationally defined class. They are commonly known as *Heyting algebras*, or *pseudo-Boolean algebras*. The relationship between Brouwerian and Heyting algebras as models is further clarified by the description of Kripke's semantics for IPC given in section 7.6.

McKinsey and Tarski applied their work on the algebra of topology to S4 and intuitionistic logic in their paper [1948], which uses closure algebras with just one designated to model S4, and Brouwerian algebras in the manner just explained to model Heyting's calculus. Using various of the results (1)–(4) above, it follows that S4 is characterised by the class of (finite) closure algebras, as well as the closure algebra of any Euclidean space, or of any zero-dimensional dense-in-itself subspace of Euclidean space. Hence in view of result (5), the claim of [Gödel,

1933] follows: if $\Box\alpha \vee \Box\beta$ is an S4-theorem, then so is one of α and β, therefore so is one of $\Box\alpha$ and $\Box\beta$ by the rule of Necessitation. Similarly, result (6) gives a proof of the *disjunction property* for IPC: if $\alpha \vee \beta$ is a theorem, then so is one of α and β. The final section of the paper uses the relationships between Brouwerian and closure algebras to verify the correctness of the two translations of IPC into S4 conjectured in Godël's paper, and introduced a new one:

$$
\begin{array}{c|l}
p & \Box p \\
\neg\alpha & \Box\neg\alpha \\
\alpha \rightarrow \beta & \Box(\alpha \rightarrow \beta) \quad \text{(i.e. } \alpha \dashv\!\!\!3\, \beta) \\
\alpha \vee \beta & \alpha \vee \beta \\
\alpha \wedge \beta & \alpha \wedge \beta.
\end{array}
$$

It is this translation that inspired Kripke [1965a] to derive his semantics for intuitionistic logic from his model theory for S4 (see section 7.6).

Another significant result of the 1948 paper is that S5 is characterised by the class of all closure algebras in which each closed element is also open. Structures of this kind were later dubbed *monadic algebras* by Halmos in his study of the algebraic properties of quantifiers [Halmos, 1962]. The connection is natural: the modalities $\Box$ and $\diamond$ have the same formal properties in S5 as do the quantifiers $\forall$ and $\exists$ in classical logic. The *polyadic* algebras of Halmos and the *cylindric* algebras of Tarski and his co-researchers [Henkin *et al.*, 1971] have a family of pairwise commuting closure operators for which each closed element is open.

Any Boolean algebra can be made into a monadic algebra by defining $\mathsf{C}0 = 0$ and otherwise $\mathsf{C}x = 1$. These are the *simple*[19] monadic algebras. Let $\mathfrak{A}_n$ be the simple monadic algebra defined on the finite Boolean algebra with n atoms, viewed as a matrix with only 1 designated. Then S5 is characterised by the set of all these $\mathfrak{A}_n$'s. This was shown by Schiller Joe Scroggs in his [1951], written as a Masters thesis under McKinsey's direction, whose analysis established that every finite monadic algebra is a direct product of $\mathfrak{A}_n$'s. Scroggs used this to prove that each proper extension of S5 is equal to the logic characterised by some $\mathfrak{A}_n$, and so has a finite characteristic matrix. By "extension" here is meant any logic that includes all S5-theorems and is closed under the rules of uniform substitution for variables and detachment for material implication. Scroggs was able to show from this characterisation that any such extension of S5 is closed under the Necessitation rule as well, and so is a normal logic.

Another notable paper on S5 algebras from this era is [Davis, 1954], based on a 1950 doctoral thesis supervised by Garrett Birkhoff. This describes the correspondence between equivalence relations on a set and S5 operations on its powerset Boolean algebra; a correspondence between algebras with two S5 operations and the *projective* algebras of Everett and Ulam [1946]; and the use of several S5 operators to provide a Boolean model of features of first-order logic.

---

[19]In the technical algebraic sense of having no non-trivial congruences.

3.3 BAO's: The Theory of Jónsson and Tarski

The notion of a *Boolean algebra with operators* (BAO) was introduced by Jónsson and Tarski in their abstract [1948], with the details of their announced results being presented in [1951]. That work contains representations of algebras that could immediately have been applied to give new characterisations of modal systems. But the paper was overlooked by modal logicians, who were still publishing rediscoveries of some of its results fifteen years later.

A unary function f on a Boolean algebra is an *operator* if it is additive, i.e. $f(x + y) = f(x) + f(y)$. f is *completely additive* if $f(\sum X) = \sum f(X)$ whenever $\sum X$ exists, and is *normal* if $f(0) = 0$. A function of more than one argument is an operator/is completely additive/is normal when it is has the corresponding property separately in each argument. A BAO is an algebra $\mathfrak{A} = (\mathfrak{B}, f_i : i \in I)$, where the f_i's are all operators on the Boolean algebra $\mathfrak{B}$.

The *Extension Theorem* of Jónsson and Tarski showed that any BAO $\mathfrak{A}$ can be embedded isomorphically into a complete and atomic BAO $\mathfrak{A}^\sigma$ which they called a *perfect extension* of $\mathfrak{A}$. The construction built on Stone's embedding of a Boolean algebra $\mathfrak{B}$ into a complete and atomic one $\mathfrak{B}^\sigma$, with each operator f_i of $\mathfrak{A}$ being extended to an operator f_i^σ on $\mathfrak{B}^\sigma$ that is *completely* additive, and is normal if f_i is normal. The notion of perfect extension was defined by three properties that determine $\mathfrak{A}^\sigma$ uniquely up to a unique isomorphism over $\mathfrak{A}$ and give an algebraic characterisation of the structures that arise from Stone's topological representation theory. These properties can be stated as follows.

(i) For any distinct atoms x, y of $\mathfrak{A}^\sigma$ there exists an element a of $\mathfrak{A}$ with $x \leq a$ and $y \leq -a$.

(ii) If a subset X of $\mathfrak{A}$ has $\sum X = 1$ in $\mathfrak{A}^\sigma$, then some finite subset X_0 of X has $\sum X_0 = 1$.

(iii) $f_i^\sigma(x) = \prod\{f_i(y) : x \leq y \in \mathfrak{A}^n\}$ when f_i is n-ary and the terms of the n-tuple x are atoms or 0.

Property (i) corresponds to the Hausdorff separation property of the Stone space of $\mathfrak{B}$, while (ii) is an algebraic formulation of the compactness of that space. The meaning of (iii) will be explained below.

Jónsson and Tarski showed that any equation satisfied by $\mathfrak{A}$ will also be satisfied by $\mathfrak{A}^\sigma$ if it does not involve Boolean complementation (i.e. refers only to $+, \cdot, 0, 1$ and the operators f_i). More generally, perfect extensions were shown to preserve any implication of the form $(t = 0 \rightarrow u = v)$ whose terms t, u, v do not involve complementation. They then established a fundamental representation of normal n-ary operators in terms of $n + 1$-ary relations. This was based on a bijective correspondence between normal *completely* additive n-ary operators f on a powerset Boolean algebra $\mathcal{P}(S)$ and $n + 1$-ary relations $R_f \subseteq S^{n+1}$. Here

$$R_f(x_0, \ldots, x_{n-1}, y) \quad \text{iff} \quad y \in f(\{x_0\}, \ldots, \{x_{n-1}\}).$$

Under this bijection an arbitrary $R \subseteq S^{n+1}$ corresponds to the n-ary operator f_R on $\mathcal{P}(S)$, where

$$y \in f_R(X_0, \ldots, X_{n-1}) \quad \text{iff} \quad R(x_0, \ldots, x_{n-1}, y) \text{ for some elements } x_i \in X_i.$$

Thus any relational structure $\mathfrak{S} = (S, R_i : i \in I)$ whatsoever gives rise to the complete atomic BAO

$$\mathsf{Cm}\,\mathfrak{S} = (\mathcal{P}(S), f_{R_i} : i \in I)$$

of all subsets of S with the completely additive normal operators f_{R_i}. Conversely, any complete and atomic BAO whose operators are normal and completely additive was shown to be isomorphic to $\mathsf{Cm}\,\mathfrak{S}$ for some structure $\mathfrak{S}$ [1951, theorem 3.9]. This representation is relevant to an understanding of the *incompleteness* phenomenon to be discussed later in section 6.1. When applied to the perfect extension $\mathfrak{A}^\sigma$ of a BAO $\mathfrak{A}$, it can be seen as defining a relational structure on the Stone space of $\mathfrak{A}$. This is now known as the *canonical structure* of $\mathfrak{A}$, denoted $\mathbb{C}\mathsf{st}\,\mathfrak{A}$, and its role will be explained further in section 6.5. The above property (iii) expresses the fact that in $\mathbb{C}\mathsf{st}\,\mathfrak{A}$, if R is the relation corresponding to some n-ary operator f_i^σ, then for each point y the set

$$\{\langle x_0, \ldots, x_{n-1} \rangle : R(x_0, \ldots, x_{n-1}, y)\}$$

is closed in the n-fold product of the Stone space topology.

$\mathsf{Cm}\,\mathfrak{S}$ is *the complex algebra of* $\mathfrak{S}$, and any subalgebra of $\mathsf{Cm}\,\mathfrak{S}$ is *a complex algebra*. This terminology derives from an old usage of the word "complex" introduced into group theory by Frobenius in the (pre-set-theoretic) 1880's to mean a collection of elements in a group. The *binary* product

$$HK = \{hk : h \in H \text{ and } k \in K\}$$

of subsets (complexes) H, K of a group G is precisely the operator f_R on $\mathcal{P}(G)$ corresponding to the *ternary* graph $R = \{(h, k, hk) : h, k \in G\}$ of the group operation.

Combining the Extension Theorem with the representation of a complete atomic algebra (like $\mathfrak{A}^\sigma$) as one of the form $\mathsf{Cm}\,\mathfrak{S}$, Jónsson and Tarski established that

> *every BAO with normal operators is isomorphic to a subalgebra of the complex algebra of a relational structure.*

The case $n = 1$ of this analysis of operators is highly germane to modal logic: the algebraic semantics discussed so far has been based on interpreting $\diamond$ as an operator on a Boolean algebra, and a normal one in the case of S4 and S5. Jónsson and Tarski observed that basic properties of a binary relation $R \subseteq S^2$ correspond to simple equational properties of the operator f_R. Thus R is reflexive iff the BAO $(\mathcal{P}(S), f_R)$ satisfies $x \leq fx$, and transitive iff it satisfies $ffx \leq x$. Hence $\mathsf{Cm}\,(S, R)$ is a closure algebra iff R is reflexive and transitive, i.e. a *quasi-ordering*. Since these conditions $x \leq fx$ and $ffx \leq x$ are preserved by perfect extensions, it followed [1951, Theorem 3.14] that

> *every closure algebra is isomorphic to a subalgebra of the complex algebra of a quasi-ordered set.*

This result, along with the Extension Theorem and the representation of a normal BAO as a complex algebra, were all stated in the abstract [1948].

A number of other properties of R were discussed in [1951], including *symmetry*. This was shown to be characterised by *self-conjugacy* of f_R, meaning that $\mathsf{Cm}\,(S, R)$ satisfies the condition $f(x) \cdot y = 0$ iff $x \cdot f(y) = 0$, which can be expressed equationally, for example by $f0 = 0$ and $fx \cdot y \leq f(x \cdot fy)$. The characterisation was used to give a representation of certain two-dimensional cylindric algebras as complex algebras over a pair of equivalence relations. Self-conjugacy of an operator is also equivalent to the equation $x \cdot f - fx = 0$, corresponding to the Brouwerian modal axiom $p \to \Box \Diamond p$. In closure algebras this is equivalent to every closed element being open: a self-conjugate closure algebra is the same thing as a monadic algebra.

As already mentioned, this study of BAO's was later overlooked. [Dummett and Lemmon, 1959] makes extensive use of complex algebras over quasi-orderings in studying extensions of S4, but makes no mention of the Jónsson–Tarski article, taking its lead instead from the McKinsey–Tarski papers and a construction in [Birkhoff, 1948] that gives a correspondence between partial orderings (i.e. anti-symmetric quasi-orderings) and closure operations of certain topologies on a set. The same omission occurs in [Lemmon, 1966b], which re-proves the representation of a unary operator on a Boolean algebra as a complex algebra over a binary relation, although it does extend the result by allowing the operator to be non-normal (see section 5.1).

3.4 Could Tarski Have Invented Kripke Semantics?

A question like this can only remain a matter of speculation. But it is not just idle speculation, given that Tarski had worked on modal logic during the same period, and given his pioneering role in the development of model theory, including the formalisation of the notions of truth and satisfaction in relational structures.

The Jónsson–Tarski work on closure algebras applies directly to the McKinsey–Tarski results on modal logic to show that S4 is characterised by the class of complex algebras of quasi-orderings. It can also be applied to show that S5 is characterised by the class of complex algebras of equivalence relations. Now the complex algebra of an equivalence relation R is a subdirect product of the complex algebras of the equivalence classes of R, each of which is a set on which R is universal. Moreover, the complex algebra of a universal relation is a *simple* monadic algebra. These observations could have been used to give a more accessible approach to the structural analysis of S5-algebras that appears in [Scroggs, 1951].

But the Jónsson–Tarski paper makes no mention of modal logic at all. Jónsson [1993] has explained that their theory evolved from Tarski's research on the algebra

of binary relations, beginning with the finite axiom system in [Tarski, 1941] which was designed to formalise the calculus of binary relations that had been developed in the nineteenth century by De Morgan, Peirce and Schröder. The primitive notions of that paper were those of Boolean algebra together with the binary operation $R_1; R_2$ of relational composition, the unary operation $R^{\smile}$ of inversion, and the distinguished constant 1' for the identity relation. Tarski asked whether any model of his axiom was representable as an algebra of actual binary relations. He later gave an equational definition of a *relation algebra* as an abstract BAO $(\mathfrak{B}, ; ,^{\smile}, 1')$ that forms an involuted monoid under $; ,^{\smile}, 1'$ and satifies the condition $x^{\smile}; -(x; y) \leq -y$. Concrete examples include the set $\mathcal{P}(S \times S)$ of all binary relations on a set S and, more generally, the set $\mathcal{P}(E)$ of subrelations of an equivalence relation E on S. Any algebra isomorphic to a subalgebra of the normal BAO $(\mathcal{P}(E), ; ,^{\smile}, 1')$ is called *representable*, and Tarski's representation question became the problem of whether every abstract relation algebra is representable in this sense.[20]

Late in 1946 Tarski communicated to Jónsson a proof that every relation algebra is embeddable in a complete and atomic one. That construction became the prototype for the Jónsson–Tarski Extension Theorem for BAO's (see [Jónsson, 1993, §1.2]). The second part of their joint work [1952] is entirely devoted to relation algebras and their representations.

It appears then that in developing his ideas on BAO's Tarski was coming from a different direction: modal logic was not on the agenda. According to [Copeland, 1996b, p. 13], Tarski told Kripke in 1962 that he was unable to see a connection with what Kripke was then doing.

4   RELATIONAL SEMANTICS

Leibniz had a good deal to say about possible worlds, including that the actual world is the best of all of them. Apparently he never *literally* described necessary truths as being "true in all possible worlds", but he did say of them that

> Not only will they hold as long as the world exists, but also they would have held if God had created the world according to a different plan.

He defined a truth as being necessary when its opposite implies a contradiction, and also said that there are as many worlds as there are things that can be conceived without contradiction (see [Mates, 1986, pp. 72–73, 106–107]).

This way of speaking has provided the motivation and intuitive explanation for a mathematical semantics of modality using relational structures that are now often called *Kripke models*. A formula is assigned a truth-value relative to each point of a model, and these points are thought of as being possible worlds or states of affairs.

---

[20]This was answered negatively by Lyndon [1950]. Work of Tarski, Monk and Jónsson eventually showed that the representable relation algebras form an equational class that is not finitely axiomatisable, with any equational definition of it requiring infinitely many variables.

An account will now be given of the contribution of Saul Kripke, followed by a survey of some of its "anticipations".

4.1   Kripke's Relatively Possible Worlds

Kripke's first paper [1959a] on modal logic gave a semantics for a quantificational version of S5 that included propositional variables as the case $n = 0$ of n-ary predicate variables. A *complete assignment* for a formula α in a non-empty set D was defined to be any function that assigns an element of D to each free individual variable in α, a subset of D^n to each n-ary predicate variable occurring in α, and a truth-value ($\top$ or $\bot$) to each propositional variable of α. A *model* of α in D is a pair (G, K), where K is a set of complete assignments that all agree on their treatment of the free individual variables of α, and G is an element of K. Each member H of K assigns a truth value to each subformula of α, by induction on the rules of formation for formulas. The truth-functional connectives and the quantifiers $\forall$, $\exists$ behave as in standard predicate logic, and the key clause for modality is that

> H assigns $\top$ to $\Box\beta$ iff every member of K assigns $\top$ to β.

A formula α is *true*[21] in a model (G, K) over D iff it is assigned $\top$ by G; *valid* over D iff true in all of its models in D; and *universally valid* iff valid in all non-empty sets D.

An axiomatisation of the class of universally valid formulas was given, with the completeness proof employing the method of *semantic tableaux* introduced in [Beth, 1955]. It was then observed that for purely propositional logic this could be turned into a truth table semantics. A complete assignment becomes just an assignment of truth values to the variables in α, i.e. a row of a truth table, and a model (G, K) is just a classical truth table with some (but not all) of the rows omitted and G some designated row. Formula $\Box\beta$ is assigned $\top$ in every row if β is assigned $\top$ in every row of the table; otherwise it is assigned $\bot$ in every row. The resulting notion of "S5-tautology" precisely characterises the theorems of propositional S5, a result that Kripke had in fact obtained first, before, as he explained in [1959a, fn. 4],

> aquaintance with Beth's paper led me to generalize the truth tables to semantic tableaux and a completeness theorem.

Kripke's informal motivation for these models was that the assignment G represents the "real" or "actual" world, and the other members of K represent worlds that are "conceivable but not actual". Thus $\Box\beta$ is "evaluated as true when and only when β holds in all conceivable worlds". The lack of any further structure on K reflects the assumption that "any combination of possible worlds may be associated with the real world".

The abstract [Kripke, 1959b] announced the availability of "appropriate model theory" and completeness theorems for a raft of modal systems, including S2–S5,

---

[21]Actually "*valid* in a model" was used here, but changed to "*true*" in [Kripke, 1963a].

the Feys–von Wright system T (or M), Lemmon's E-systems, systems with the Brouwerian axiom, deontic systems, and others. Various extensions to quantificational logic with identity were described, and it was stated that "the methods for S4 yields a semantical apparatus for Heyting's system which simplifies that of Beth". The details of this programme appeared in the papers [1963a; 1963b; 1965a; 1965b].

The normal propositional logics S4, S5, T and B are the main focus of [Kripke, 1963a], which defines a *normal model structure* as a triple (G, K, R) with $G \in K$ and R a *reflexive* binary relation on K. A *model* for a propositional formula α on this structure is a function $\Phi(p, H)$ taking values in $\{\top, \bot\}$, with p ranging over variables in α and H ranging over K. This is extended to assign a truth value $\Phi(\beta, H)$ to each subformula β of α and each $H \in K$, with

$$\Phi(\Box\beta, H) = \top \quad \text{iff} \quad \Phi(\beta, H') = \top \text{ for all } H' \in K \text{ such that } HRH'.$$

α is *true* in the model if $\Phi(\alpha, G) = \top$.

In addition to the introduction of the relation R, the other crucial conceptual advance here is that the set K of "possible worlds" is no longer a collection of value assignments, but is permitted to be an arbitrary set. This allows that there can be different worlds that assign the same truth values to atomic formulas. As to the relation R, Kripke's intuitive explanation is as follows [1963a, p. 70]:

> we read "$H_1 R H_2$" as H_2 is "possible relative to H_1", "possible in H_1" or "related to H_1"; that is to say, every proposition true in H_2 is to be possible in H_1. Thus the "absolute" notion of possible world in [1959a] (where every world was possible relative to every other) gives way to relative notion, of one world being possible relative to another. It is clear that every world H is possible relative to itself; for this simply says that every proposition *true* in H is *possible* in H. In accordance with this modified view of "possible worlds" we evaluate a formula A as *necessary* in a world H_1 if it is *true* in every world possible relative to H_1. ...Dually, A is possible in H_1 iff there exists H_2, possible relative to H_1, in which A is *true*.

Semantic tableaux methods are again used to prove completeness theorems: a formula is true in all models iff it is a theorem of T; true in all transitive models iff it is an S4-theorem, true in all symmetric models iff a B-theorem, and true in all transitive and symmetric models iff an S5-theorem. The arguments also give decision procedures, and show that attention can be restricted to models that are *connected* in the sense that each $H \in K$ has GR^*H, where R^* is the *ancestral* or reflexive-transitive closure of R. Kripke notes that

> in a connected model in which R is an equivalence relation, *any* two worlds are related. This accounts for the adequacy, for S5, of the model theory of [1959a].

An illustration of the tractability of the new model theory is given by a new proof of the deduction rule in S4 that if $\Box\alpha \lor \Box\beta$ is deducible then so is one of α and

β. If neither α nor β is derivable then each has a falsifying S4-model. Take the disjoint union of these two models and add a new "real" world that is R-related to everything. The result is an S4-model falsifying $\Box\alpha \vee \Box\beta$. This argument is much easier to follow than the McKinsey–Tarski construction involving well-connected algebras described in section 3.2., and it adapts readily to other systems.

Other topics discussed include the presentation of models in "tree-like" form, and the association with each model structure of a matrix, essentially the modal algebra of all functions $\rho : K \to \{\top, \bot\}$, which are called *propositions*, with the ones having $\rho(G) = \top$ being designated. A model can then be viewed as a device for associating a proposition $H \mapsto \Phi(p, H)$ to each propositional variable p. The final section of the paper raises the possibility of defining new systems by imposing various requirements on R, and concludes that

> [i]f we were to drop the condition that R be reflexive, this would be equivalent to abandoning the modal axiom $\Box A \to A$. In this way we could obtain systems of the type required for deontic logic.

Non-normal logics are the subject of [Kripke, 1965b], which focuses mainly on Lewis's S2 and S3 and the corresponding systems E2 and E3 of [Lemmon, 1957]. The E-systems have no theorems of the form $\Box\alpha$, and this suggests to Kripke the idea of allowing worlds in which any formula beginning with $\Box$ is false, and hence any beginning with $\Diamond$, even $\Diamond(p \wedge \neg p)$, is true. A model structure now becomes a quadruple (G, K, R, N) with N a subset of K, to be thought of as a set of *normal* worlds, and R a binary relation on K as before, but now required to be reflexive on N only. The semantic clause for $\Box$ in a model on such a structure is modified by stipulating that

> $\Phi(\Box\beta, H) = \top$ iff H is normal, i.e. $H \in N$, and $\Phi(\beta, H') = \top$ for all $H' \in K$ such that HRH';

and hence

> $\Phi(\Diamond\beta, H) = \top$ iff H is non-normal or else $\Phi(\beta, H') = \top$ for some $H' \in K$ such that HRH'.

This has the desired effect of ensuring $\Phi(\Box\beta, H) = \bot$ and $\Phi(\Diamond\beta, H) = \top$ whenever H is non-normal. Thus in a non-normal world, even a contradiction is *possible*.

These models characterise E2, and the ones in which R is transitive characterise E3. Requiring that the "real" world G belongs to N gives models that characterise S2 and S3 in each case.[22] A number of other systems are discussed and applications given, including a proof of a long-standing conjecture that the Feys–von Wright system has no finite axiomatisation with detachment as its sole rule of inference.

Kripke's semantics for quantificational modal logic is presented in his [1963b]. A model structure now has the added feature of a function assigning a set $\psi(H)$

---

[22]A semantics for S1 was devised in 1969 by Max Cresswell, modifying Kripke's S2-models to allow some formulas $\Diamond\beta$ to be false in a non-normal world under certain restrictions, defined with the help of a *neighbourhood* relation $R' \subseteq K \times \mathcal{P}(K)$. See [Cresswell, 1972; 1995].

to each $H \in K$. Intuitively, $\psi(H)$ is the set of all individuals existing in H, and it provides the range of values for a variable x when a formula beginning with $\forall x$ is evaluated at H. A model now assigns to each n-ary predicate letter and each $H \in K$ an n-ary relation on the set $\bigcup\{\psi(H') : H' \in K\}$ of individuals that exist in any world. Axioms are given for quantificational versions of the basic modal logics and it is stated that the completeness theorems of [1963a] can be extended to them. An indication of how that would work can be obtained from Kripke's [1965b], which gives a tableaux completeness proof for his semantics for Heyting's intuitionistic predicate calculus.

4.2   So Who Invented Relational Models?

Kripke's abstract [1959b] notes that "for systems based on S4, S5, and M, similar work has been done independently and at an earlier date by K. J. J. Hintikka". This acknowledgement is repeated in [1963a, fn. 2] where he draws attention to prior work by a number of researchers, including Bayart, Jónsson and Tarski, and Kanger, explaining that his own work was done independently of all of them. He states that the modelling of [Kanger, 1957b] "though more complex, is similar to that in the present paper", and also records that he discovered the Jónsson–Tarski paper when his own was almost finished.

Key ideas surrounding relational interpretations of modality had occurred to several people. In the next few sections we survey some of this background, before expressing a view about the relative significance of Kripke's work.

As mathematics progresses, notions that were obscure and perplexing become clear and straightforward, sometimes even achieving the status of "obvious". Then hindsight can make us all wise after the event. But we are separated from the past by our knowledge of the present, which may draw us into "seeing" more than was really there at the time. This should be borne in mind in reading what follows.

4.3   Carnap and Bayart on S5

A *state-description* is defined by Rudolf Carnap in [1946; 1947] to be set of sentences which consists of exactly one of α and $\neg\alpha$ for each *atomic* α. State-descriptions are said to "represent Leibniz's possible worlds or Wittgenstein's possible states of affairs". A sentence is called *L-true* if it holds in every state-description, this being "an explicatum for what Leibniz called necessary truth and Kant analytic truth" [1947, p. 8].

Of course it needs to be explained what it is to hold in a state-description. An atomic sentence holds in a state description iff it belongs to it, the conditions for the connectives $\neg$, $\wedge$, and $\vee$ are as expected, and the criterion for Carnap's necessity connective N is that

> $N\alpha$ holds in every state-description if α holds in every state-description; otherwise, $N\alpha$ holds in no state-description

[1946, D9-5i], [1947, 41-1]. His list of L-truths ([1946, p. 42], [1947, p. 186]) includes the axioms for S5, and he also notes the similarity between N and $\forall$, and between $\Diamond$ and $\exists$ under this semantics. The 1946 paper observes that there is a procedure for deciding L-truth that is "*theoretically* effective": if a sentence α has n atomic components then there are 2^n state-descriptions that have to be considered in evaluating it, and therefore 2^{2^n} possibilities for the *range* of α, which is the set of state-descriptions in which α holds. We can examine all possibilities to see if the range includes all state-descriptions. Carnap defines a version of S5 which he calls MPC and proves that it is complete with respect to his semantics, by a reduction of formulas to a normal form[23] which also gives a decision procedure that is

> practicable, i.e. sufficently short for modal sentences of ordinary length.

He attributes the completeness result to a paper of Mordchaj Wajsberg from 1933. Footnote 8 of [1946] gives a description of Wajsberg's system and also contains the information that Carnap constructed MPC independently in 1940 and later found that it was equivalent to Lewis's S5.

A contribution to possible worlds model theory that has been largely overlooked is the work of the Belgian logician A. Bayart, whose papers of [1958] and [1959] gave a semantics for a version of second order quantificational S5, and a complete axiomatisation of it using a Gentzen-style sequent calculus. The models used allow a restricted range of interpretation of predicate variables. This idea had been introduced in [Henkin, 1950] to give a completeness result for non-modal higher order logic, and Bayart commented [1959, p. 100] that he had just adapted Henkin's theorem to S5.[24] The other source of motivation he gives [1958, p. 28] is Leibniz's definition of necessity as truth in all possible worlds,[25] and his bibliography cites the items [Carnap, 1946; 1947].

In Bayart's theory a *universe* U is defined to be a disjoint pair A, B of sets, with members of A called *individuals* and members of B called *worlds* ("mondes"). An *n-place intensional predicate* is a function of $n + 1$ arguments, taking the values "true" or "false", having a world as its first argument, and having individuals as the remaining arguments when $n \neq 0$. A *value system* relative to U is a function S assigning a member of A to each individual variable, and an n-place intensional predicate to each n-place predicate variable. The notion of a formula being true or false for the universe U, the world M and the value system S — or more briefly for UMS — is defined in the expected way for the non-modal connectives and quantifiers, including quantifiers binding predicate variables. For modalized formulas Lp and Mp it is declared that

> Lp is true for UMS iff for every world M' of U, p is true for $UM'S$;

---

[23]Called *modal conjunctive normal form* in [Hughes and Cresswell, 1968, p. 116], where a variant of the proof is given.

[24]"En réalité notre exposé n'est qu'une adaptation du théorème de Henkin à la logique modale S5."

[25]"...en nous inspirant de la définition Leibnizienne du nécessaire, comme étant ce qui est vrai dans tous les mondes possibles."

Mp is true for UMS iff for some world M' of U, p is true for $UM'S$.

A formula is *valid* in the universe U if it is true for UMS for every world M and value system S of U.

Bayart used the notation $\ddot{a}, I, \ddot{e}$ for a Gentzen sequent, with $\ddot{a}$ (the antecedent) and $\ddot{e}$ (the consequent) being finite sequences of formulas, and I a separating symbol. The sequent is true in UMS if some member of $\ddot{a}$ is false or else some member of $\ddot{e}$ is true. He adopted the axiom schema $\ddot{p}, I, \ddot{p}$ and a system of twenty-five deduction rules, showing in [1958] that all deducible sequents are valid in all universes. There are four modal rules, allowing the introduction of the modalities L and M into antecedents and consequents:

$$\frac{p, \ddot{a}, I, \ddot{e}}{Lp, \ddot{a}, I, \ddot{e}} \qquad \frac{p, \ddot{a}, I, \ddot{e}}{Mp, \ddot{a}, I, \ddot{e}} \qquad \frac{\ddot{a}, I, \ddot{e}, p}{\ddot{a}, I, \ddot{e}, Lp} \qquad \frac{\ddot{a}, I, \ddot{e}, p}{\ddot{a}, I, \ddot{e}, Mp}.$$

The last two rules are subject to the restriction that any formula appearing in $\ddot{a}$ or $\ddot{e}$ must be "couverte", meaning that it is formed from formulas of the types Lq and Mq using only the non-modal connectives and quantifiers. Such a formula has the same truth value in UMS and $UM'S$ for all worlds M, M'.

The [1959] paper proved the completeness of this sequent system for validity in certain *quasi-universes* obtained by allowing predicate variables to take values in a restricted class of intensional predicates. From this it was shown that the first order fragment of the system is complete for validity in all universes. The method used was subsequently generalised in [Cresswell, 1967] to obtain a completeness theorem for the relational semantics of a first order version of the modal logic T (see section 5.1).

It is worth recording Bayart's explanation of why the set of worlds of a universe $U = A, B$ is essential to this theory. He considered the possibility of dispensing with B, requiring a value system S to interpret an n-place predicate variable as an *extensional* predicate (i.e. a truth-valued function on A^n), and modelling the necessity modality by declaring that

Lp is true of US iff p is true of US' for every value system S'.

He noted that this interpretation fails to validate the formula

$$\exists y\, L(bx \vee \neg by)$$

(where b is a unary predicate variable), a formula that is valid according to the above semantics. His explanation of the flaw in this alternative approach is that it gives Lp the same meaning as the universal closure of p (i.e. $\forall v_1 \cdots \forall v_n p$, where $v_1, \ldots, v_n$ are the free variables of p), and confuses necessity with validity.

4.4   Meredith, Prior and Geach

Arthur Prior [1967, p. 42] wrote that

> In some notes made in 1956, C. A. Meredith related modal logic to what he
> called the 'property calculus'.

This material was made available by Prior as a one-page departmental mimeo-graph [Meredith, 1956] which was published much later in the collection [Copeland, 1996a]. Its basic idea was to express modal formulas in the first-order language of a binary predicate symbol U, beginning with the following definitions, in which L and M are connectives for necessity and possibility (but the other notation is that of this paper rather than the original Polish):

$$(\neg p)a = \neg(pa)$$
$$(p \to q)a = (pa) \to (qa)$$
$$(Lp)a = \forall b(Uab \to pb)$$
$$(Mp)a = (\neg L \neg p)a = \exists b(Uab \wedge pb).$$

Possible axioms for U are then listed:

1. $Uab \vee Uba$
2. $Uab \to (Ubc \to Uac)$
3. $Uab \to (Ucb \to Uac)$
4. Uaa
5. $Uab \to Uba,$

and it is noted that "1 gives 4"; "3, 4 give 5"; and "3, 5 give 2". The notes are written in this telegraphic style with no interpretation of the symbolism, but presumably "pa" may be read "a has property p".

It is stated that quantification theory alone allows the derivation of

$$\big(L(p \to q) \to (Lp \to Lq)\big)a,$$

and then formal deductions are given of $(Lp \to p)a$ using 4; of $(Lp \to LLp)a$ using 2; of $(MLp \to Lp)a$ using 2 and 5; and of $\forall apa$ from $(Lp)a$ using 1 and 5. The conclusion is as follows:

> Thus 1, or 4, gives T; 1, 2 or 4, 2 gives S4; 1, 3 or 4, 3 gives S5; and 1, 3 (but
> not 4, 3) gives the equivalence of the above $(Lp)a$ with the usual S5 $(Lp)a$,
> i.e. $\forall apa$.

Prior's article "Possible Worlds" [1962a, p. 37] gives a fuller exposition of this U-calculus, saying "This whole symbolism I owe to C. A. Meredith". He applies an interpretation of the predicate U, suggested to him by P. T. Geach in 1960,[26] as a relation of *accessibility*. Here is Prior's account of that interpretation.

> Suppose we define a 'possible' state of affairs or world as one which can be
> reached from the world we are actually in. What is meant by reaching or

---

[26]This date is given in [Prior, 1962b, p. 140], where the acknowledgement of Meredith is repeated once more.

travelling to one world from another need not here be amplified; we might
reach one world from another merely in thought, or we might reach it more
concretely in some dimension-jumping vehicle dreamed up by science-fiction
(the case originally put by Geach), or we might reach it simply by the passage
of time (one important sense of 'possible state of affairs' is 'possible outcome
of the present state of affairs'). What I want to amplify here is the idea
(the core of Geach's suggestion) that we may obtain different modal systems,
different versions of the logic of necessity and possibility, by making different
assumptions about 'world-jumping'.

Prior was the founder of tense logic (also known as temporal logic). He wanted to
analyse the arguments of the Stoic logician Diodorus Chronos, who had defined a
proposition to be *possible* if it either is true or will be true. Prior conceived the
idea of using a logical system with temporal operators analogous to those of modal
logic, and thus introduced the connectives

F it will be the case that
P it has been the case that
G it will always be the case that
H it has always been the case that.

Here F and P are "diamond" type modalities, with duals G and H respectively.
In the paper "The Syntax of Time-Distinctions" [Prior, 1958] a propositional logic
called the PF-*calculus* is defined.[27] It is a normal logic with respect to G and H,
has the axioms $Gp \rightarrow Fp$, $FFp \rightarrow Fp$ and $Fp \rightarrow FFp$, as well as an "interaction"
axiom $p \rightarrow GPp$ and a *Rule of Analogy* allowing that from any theorem another
may be deduced by replacing F by P and vice versa.

   This system is then interpreted into what Prior calls the l-*calculus*, a first-order
language whose variables x, y, z range over *dates*, and which has a binary symbol
l taking dates as arguments, with the expression lxy being read "x is later than
y".[28] Variables p, q, r stand for propositions considered as functions of dates, with
the expression px being read "p at x". The following interpretations are given
of propositional formulas, using an arbitrarily chosen date variable z to represent
"the date at which the proposition under consideration is uttered".

Fp $\exists x(lxz \wedge px)$
Pp $\exists x(lzx \wedge px)$
Gp $\forall x(lxz \rightarrow px)$
Hp $\forall x(lzx \rightarrow px)$.

Prior observes that the interpretations of some theorems of the PF-calculus are
provable in the l-calculus just from the usual axioms and rules for quantificational
logic. This applies to any PF-theorem derivable from the basis for normal logics
together with the interaction axiom $p \rightarrow GPp$ and the rule of Analogy. He then

---

[27]The contents of this paper are reviewed on [Prior, 1967, pp. 34–41].
[28]Prior notes that the structure of the calculus would be unchanged if l were read "is earlier
than".

states that the interpretation of $Gp \to Fp$ requires for its proof the axiom $\exists x \, lxz$ ("infinite extent of the future"), and that $FFp \to Fp$ depends similarly on *transitivity*: $lxy \to (lyz \to lxz)$, while $Fp \to FFp$ depends on the density condition $lxz \to \exists y(lxy \wedge lyz)$.

The modality M of possibility is given a temporal reading by defining Mp to be an abbreviation for $p \vee Fp \vee Pp$, i.e. "p is true at some time, past present or future". This makes the dual Lp equivalent to $p \wedge Gp \wedge Fp$, "at all times, p". Prior notes that to derive the S5-principle $M\neg Mp \to \neg Mp$, which is "clearly a law" under this interpretation of M, requires *trichotomy*: $x = y \vee lxy \vee lyx$. His explorations here are quite tentative. For instance he defines *asymmetry*: $lxy \to \neg lyx$, but makes no use of it, and he fails to note that the S4-principle $MMp \to Mp$ also depends on trichotomy and not just transitivity.

Why did Prior give such unequivocal credit to Meredith for the 1956 U-calculus? The puzzle about this is that his paper on the l-calculus, although published in 1958, was presented much earlier, on 27 August 1954, as his Presidential Address to the New Zealand Philosophy Congress at the Victoria University of Wellington. Perhaps he was crediting Meredith with the extension of the symbolism to *modal* logic as he understood it, i.e. the logic of necessity and possibility, as distinct from tense logic. The l-calculus was intended to describe a very specific situation: an ordered system of dates or moments in time that forms an "infinite and continuous linear series" [1958, p. 115]. In the absence of any corresponding interpretation of the U-predicate, the purely formal application of the symbolism by Meredith may have been seen by Prior as a significant advance.

Prior made much use of l and U calculi in his papers and books on tense logic. He did not however pursue their implicit relational model theory, and would not have thought it philosophically worthwhile to do so. Although he described the l-calculus as "a device of considerable metalogical utility" [1958, p. 115], he went on to deny that the interpretation of the PF-calculus within the l-calculus has any metaphysical significance as an

> explanation of what we mean by "is", "has been" and "will be".

On the contrary he proposed that what was needed was an interpretation in the reverse direction [1958, p. 116]:

> the l-calculus should be exhibited as a logical construction out of the PF-calculus.

This proposal became a major programme for Prior. He used formulas like $p \wedge \neg Pp \wedge \neg Fp$ which can be true at only one point of the linear series of moments, or *instants*. If $M(p \wedge \neg Pp \wedge \neg Fp)$ is true at some time, the variable p must itself be true at exactly one instant and may be identified with that instant. Then the formula $L(p \to \alpha)$ expresses that "it is the case at p that α", and so if p and q are both such instance-variables, $L(p \to Pq)$ asserts that it is true at p that it has been q, i.e. p is later than q, and q is earlier than p.

Systems having variables identified with unique instants or worlds are developed most fully in the book of [Prior and Fine, 1977, p. 37], where Prior gives an emphatic statement of his metaphysical propensity:

> ... I find myself quite unable to take 'instants' seriously as individual entities; I cannot *understand* 'instants', and the earlier-than relation that is supposed to hold between them, except as logical constructions out of tensed facts. Tense logic is for me, if I may use the phrase, *metaphysically fundamental*, and not just an artificially torn-off fragment of the first-order theory of the earlier-than relation.

4.5 Kanger

A semantics is given by Stig Kanger in [1957b] for a version of modal predicate logic whose atomic formulas are propositional variables and expressions of the form $(x_1, \ldots, x_n) \, \varepsilon \, y$, where $n \geq 1$ and the x_i and y are individual variables or constants. The language included a list of modal connectives M_1, M_2, ...

A notion of a *system* is introduced as a pair (r, V) where r is a *frame* and V a *primary valuation*. Here r is a certain kind of sequence of non-empty sets whose elements provide values of individual symbols of various types. V is a binary operation that assigns a truth value $V(r, p)$, belonging to $\{0, 1\}$, to each propositional variable p and frame r, as well as interpreting individual symbols and the symbol ε in each frame in a manner that need not concern us. Then a "secondary" truth valuation $T(r, V, \alpha)$ is inductively specified, allowing each formula α to be defined to be *true in system* (r, V) iff $T(r, V, \alpha) = 1$. For this purpose each modality M_i is assumed to be associated with a class R_i of *quadruples* (r', V', r, V), and it is declared that

$$T(r, V, M_i\alpha) = 1 \text{ iff } \quad T(r', V', \alpha) = 1 \text{ for each } r' \text{ and } V' \text{ such that } R_i(r', V', r, V)$$

(so M_i is a "box" type of modality).

Kanger states the following *soundness* results. The theorems of the Feys–von Wright system T are valid (i.e. true in all systems) iff $R_i(r, V, r, V)$ always holds. S4 is validated iff $R_i(r, V, r, V)$ always holds and so does the condition

$$R_i(r, V, r', V') \text{ and } R_i(r'', V'', r, V) \text{ implies } R_i(r'', V'', r', V').$$

S5 is validated iff the S4 conditions hold along with

$$R_i(r, V, r', V') \text{ and } R_i(r'', V'', r', V') \text{ implies } R_i(r'', V'', r, V).$$

Proofs of these assertions are not provided. (In fact it is readily seen that the given conditions on R_i imply validity for the corresponding logics in each case, but the converses are dubious.) A result is proved that equates the existence of an R_i fulfilling the above definition of $T(r, V, M_i\alpha)$ to the preservation of certain inference rules involving M_i. Kanger says of this that

[s]imilar results in the field of Boolean algebras with operators may be found in [Jónsson and Tarski, 1951].

Completeness theorems are not proved, or even stated, for this modal semantics. But there is a completeness proof for the non-modal fragment of the language which has a remarkable aspect. Kanger wishes to have the symbol ε interpreted as the genuine set membership relation, and he applies the (much-overused) adjective *normal* to a primary valuation V which does give this interpretation to ε in every frame. Since his language allows atomic formulas like $x\,\varepsilon\,x$, normal systems must have *non-well-founded* sets. He introduces a new set-theoretical principle to ensure that enough such sets exist to give the completeness theorem with respect to normal structures.[29]

Different definitions of R allow the modelling of different notions of necessity. Kanger [1957a, p. 35] defines *set-theoretical necessity* to be the modality given by requiring

$$R_i(r', V', r, V) \text{ iff } V' \text{ is normal with respect to } \varepsilon.$$

This means that M_i gets the reading "in all normal systems". *Analytic necessity* is modelled by the R_i having

$$R_i(r', V', r, V) \text{ iff } V' = V,$$

and *logical necessity* arises when $R_i(r', V', r, V)$ always holds. Thus "logically necessary" means "true in all systems", which is reminiscent of the modelling of the S5 necessity connective by Carnap and Bayart (section 4.3).

There is no doubt much scope for defining other modalities in this way, and Kanger offers one other brief suggestion:

> We may, for instance, define 'geometrical necessity' in the way we defined set-theoretical necessity except that (roughly speaking) V' shall be normal also with respect to the theoretical constants of geometry.

The paper [Kanger, 1957a] addresses difficulties raised by Quine (in [1947] and other writings) about the possibility of satisfactorily interpreting quantificational modal logic. One such obstacle concerns the principle of *substitutivity of equals*, formalised by the schema

$$x \approx y \to (\alpha \to \alpha')$$

where α' is any formula differing from α only in having free occurrences of y in some places where α has free occurrences of x. Taking α to be the valid $\Box(x \approx x)$, this allows derivation of

$$x \approx y \to \Box(x \approx y),$$

which is arguably invalid. For example, it is an astronomical fact that the Morning Star and the Evening Star are the same object (Venus), but this equality is not a *necessary* truth.

---

[29]This principle is discussed further in [Aczel, 1988, pp. 28–31 and 108].

Kanger pointed out that his new semantics for quantification and modality made it possible to "recognize and explain the error in the Morning Star paradox": the principle of substitutivity of equals is not valid without restriction, but only in the weaker form

$$\Box(x \approx y) \rightarrow (\alpha \rightarrow \alpha').$$

Jaakko Hintikka [1969] later expressed the opinion that this discussion by Kanger of the Morning Star paradox will

> remain a historical landmark as the first philosophical application of an explicit semantical theory of quantified modal logic.

4.6 Montague

Kanger's quaternary relation R_i might equally well be viewed as a binary relation $(r', V')\, R_i\, (r, V)$ between systems. Such a notion appears in a paper by Richard Montague [1960] which was originally presented to a philosophy conference at the University of California, Los Angeles, in May of 1955. Montague did not initially plan to publish the paper because "it contains no results of any great technical interest", but eventually changed his mind after the appearance of Kanger's and Kripke's ideas.

The aim of the paper is to interpret logical and physical necessity, and the deontic modality "it is obligatory that", and to relate these to the use of quantifiers. Tarski's model theory for first-order languages is employed for this purpose: a model is taken to be a structure $\mathcal{M} = (D, R, f)$ where D is a domain of individuals, R a function fixing an interpretation of individual constants and finitary predicates in D in the now-familiar way, and f is an assignment of values in D to individual variables. Montague uses these models to provide a semantics for formulas that are constructible from atomic first-order formulas by using the propositional connectives and $\Box$, but not quantifiers.[30] His approach is to take a relation X between such models, and then inductively define

$\mathcal{M}$ satisfies $\Box\alpha$ iff for every model $\mathcal{M}'$ such that $\mathcal{M} X \mathcal{M}'$, $\mathcal{M}'$ satisfies α.

His first example shows that the Tarskian semantics for $\forall$ fits this definition. Taking X to be the relation Q_x specified by

$\mathcal{M} Q_x \mathcal{M}'$ iff $D = D'$, $R = R'$ and f and f' agree except on x

gives $\Box$ the interpretation "for all x". Thus quantification could be handled by associating a modality with each variable, and Montague suggests that this should dispel Quine's uneasiness about combining modality with quantification.

The relation

$\mathcal{M} L \mathcal{M}'$ iff $D = D'$ and $f = f'$

---

[30]Montague uses several symbols for various kinds of modality, but $\Box$ will suffice here.

gives $\Box\alpha$ the interpretation "it is logically necessary that α", meaning that α holds no matter what its individual constants and predicates denote.

To interpret physical necessity, Montague uses the idea that a statement is physically necessary if it is deducible from some set of physical laws specified in advance. This is formalised by fixing a set K of first-order $\Box$-free sentences and specifying a relation P by

$$\mathcal{M}\,P\,\mathcal{M}' \text{ iff } D = D', \, f = f' \text{ and } \mathcal{M}' \text{ is a model of } K.$$

Similarly, "it is obligatory that α" is taken to mean that α is deducible from some set of ethical laws specified in advance. This is formalised by fixing a class I of *ideal models*, those in which the constants and predicates mean what they ought to according to these laws. Montague suggests as an example that I could be

> the class of models which, in Tarski's sense, satisfy the ten commandments formulated as declarative, rather than imperative, sentences.

The deontic modality then corresponds to the model-relation E such that

$$\mathcal{M}\,E\,\mathcal{M}' \text{ iff } D = D', \, f = f' \text{ and } \mathcal{M}' \text{ belongs to } I.$$

If a model-relation X fulfills the conditions

> for all $\mathcal{M}$ there exists $\mathcal{M}'$ with $\mathcal{M}\,X\,\mathcal{M}'$,
>
> $\mathcal{M}\,X\,\mathcal{M}'$ and $\mathcal{M}'\,X\,\mathcal{M}''$ implies $\mathcal{M}\,X\,\mathcal{M}''$,
>
> $\mathcal{M}\,X\,\mathcal{M}'$ and $\mathcal{M}\,X\,\mathcal{M}''$ implies $\mathcal{M}'\,X\,\mathcal{M}''$,

(the last two mirror Kanger's conditions) then every S5-theorem is valid, i.e. satisfied by every model. Montague states that the converse is true, and that there is a decision method for the class of formulas valid in this sense.

4.7   Hintikka

If $\mathcal{M}$ is a model for predicate logic, of the kind used by Montague, let $\mu_{\mathcal{M}}$ be the set of all formulas that it satisfies. In Jaakko Hintikka's approach to semantics, such models $\mathcal{M}$ are in effect replaced by the sets $\mu_{\mathcal{M}}$. These sets can be characterised by their syntactic closure properties, obtained by replacing "$\mathcal{M}$ satisfies α" by "$\alpha \in \mu_{\mathcal{M}}$" in the clauses of the inductive definition of satisfaction of formulas. A *model set* is defined as a set μ of formulas that has certain closure properties, such as

> if α is atomic then not both $\alpha \in \mu$ and $\neg\alpha \in \mu$,
>
> if $\alpha \wedge \beta \in \mu$, then $\alpha \in \mu$ and $\beta \in \mu$,
>
> if $\alpha \vee \beta \in \mu$, then $\alpha \in \mu$ or $\beta \in \mu$,
>
> if $\exists x\alpha \in \mu$, then $\alpha(y/x) \in \mu$ for some variable y,

that are sufficient to guarantee that μ can be extended to a *maximal* model set which has *all* such closure properties corresponding to the conditions for satisfaction for the truth-functional connectives and the quantifiers.[31]

Hintikka's article [1957] gives a definition of satisfaction for formulas of quantified deontic logic using model sets whose conditions

> may be thought of as expressing properties of the set of all statements that are true under some particular state of affairs.

He notes [1957, p. 10] that his treatment derives from a

> new general theory of modal logics I have developed.

This general modelling of modalities was published in [1961], where he views a maximal model set as the set of all formulas that hold in some state-description in the sense of Carnap, and says that

> a model set is the formal counterpart to a partial description of a possible state of affairs (of a 'possible world'). (It is, however, large enough a description to make sure that the state of affairs in question is really possible.)

The point of the last sentence is that for non-modal quantificational logic, every model set is included in $\mu_{\mathcal{M}}$ for some actual model $\mathcal{M}$. Hence a set of non-modal formulas is satisfiable in the Tarskian sense if it is included in some model set.

The 1957 article deals with a system that has quantifiable variables ranging over individual acts, and dual modalities for *obligation* and *permission*, with formulas $O\alpha$ and $P\alpha$ being read "α is obligatory" and "α is permissible", respectively. The paper makes very interesting historical reading, especially on pages 11 and 12 where one can almost see the notion of a binary relation between model sets quickening in the author's mind as he grapples with the question of what we mean by saying that α is permitted. His answer is that

> we are saying that one could have done α without violating one's obligations. In other words, we are saying that a state of affairs different from the actual one is consistently thinkable, *viz.* a state of affairs in which α is done but in which all the obligations are nevertheless fulfilled.

Thus if the actual state is (partially) represented by a model set μ, then to represent this different and consistently thinkable state we need

> another set μ^* related to μ in a certain way. This relation will be expressed by saying that μ^* is *copermissible with* μ.

Hintikka is thus led to formulate the following rules.

> If $P\alpha \in \mu$, then there a set μ^* copermissible with μ such that $\alpha \in \mu^*$.

> If $O\alpha \in \mu$ and if μ^* is copermissible with μ, then $\alpha \in \mu^*$.

---

[31]In fact it is assumed that formulas are in a certain normal form, but we can overlook the technicalities here.

The second rule addresses the requirement that all actual obligations be fulfilled in the state in which a permissible act is done. Then there are two more rules:

> If $O\alpha \in \mu^*$ and if μ^* is copermissible with some other set μ, then $\alpha \in \mu^*$.
>
> If $O\alpha \in \mu$ and if μ^* is copermissible with μ, then $O\alpha \in \mu^*$.

Motivation for third rule is as follows.

> But not only one must be thought of in μ^* as fulfilling the obligations one has now. Sometimes one is permitted to do something only at the cost of new obligations. These must be thought of as being fulfilled in μ^* in order to be sure that all the obligations one has really are compatible with α's being done.

The fourth rule is justified because

> there seems to be no reason why the actually existing obligations should not also hold in the alternative state of affairs contemplated in μ^*. What is thought of as obligatory in μ must hence also be obligatory in μ^*.

Hintikka is well aware that the relation between μ and μ^* cannot be functional: there may be different acts that are each permissible in μ but cannot or must not be performed together, hence must be done in different states copermissible with μ. Also, μ^* may have its own formulas of the form $P\alpha$, requiring further model sets μ^{**} copermissible with μ^*, and so on. The upshot is that a set λ of formulas is defined to be *satisfiable* iff it is included in some model set which itself belongs to a *collection* of model sets that carries a binary relation (called the relation of copermission) obeying the closure rules for P and O.[32] A formula α is *valid* if $\{\neg\alpha\}$ is not satisfiable in this sense.

   This approach gives a method for demonstrating satisfiability and validity, by starting with a set λ and attempting to build a suitable collection of model sets by repeatedly applying all the closure rules. New sets are added to the collection when the rule for P is applied. The other rules enlarge existing sets. If at some point a violation of the rule of consistency is produced, in the form of a contradictory pair $\alpha, \neg\alpha$ in some set, then the original λ is not satisfiable.

   Hintikka gives a striking illustration of the effectiveness of this technique for analysing the subtleties of denotic logic. He demonstrates the invalidity of the principle

$$O\alpha \wedge (\alpha \to O\beta) \to O\beta,$$

which Prior had thought was a "quite plain truth", by observing that its negation is satisfied in the simple collection consisting of the two model sets

$$\{O\alpha,\ \neg\alpha \vee O\beta,\ P\neg\beta,\ \neg\alpha\} \qquad \{O\alpha,\ \neg\beta,\ \alpha\}.$$

---

[32] Note that the second rule is a consequence of the third and fourth.

However the principle can be turned into a valid one by making it obligatory:

$$O[O\alpha \wedge (\alpha \to O\beta) \to O\beta].$$

Any attempt to build a satisfying structure for the negation of this formula leads to violation of consistency. Several other applications like this are given, analysing complex principles involving the interchange of quantifiers and deontic modalities.

With the advantage of hindsight we can see that the notion of a collection of model sets with closure rules is reminiscent of the notion of a collection of semantic tableaux used in Kripke's completeness proofs. Hintikka did not however take up an axiomatic development of his system.

The paper [1961] deals with the necessity (N) and possibility (M) modalities, and here the description of satisfiability is essentially the same, but more crisply presented. A *model system* is defined a pair (Ω, R) with R being a binary relation of "alternativeness" on Ω, and Ω being a collection of model sets that satisfies the following conditions.

> If $M\alpha \in \mu \in \Omega$, then there is in Ω at least one alternative ν to μ such that $\alpha \in \nu$.
>
> If $N\alpha \in \mu \in \Omega$, and if $\nu \in \Omega$ is an alternative to μ, then $\alpha \in \nu$.
>
> If $N\alpha \in \mu \in \Omega$, then $\alpha \in \Omega$.

The first two of these are the same as the first two rules for P and O. The third reflects the requirement that any necessary truth be actually true. Hintikka's description of the new alternativeness relation is that $\mu R \nu$ when ν is a partial description of

> some other state of affairs that could have been realised instead of μ.

A set λ of formulas is *satisfiable* (as before) iff there is such a model system with $\lambda \subseteq \mu$ for some $\mu \in \Omega$, and a formula α is *valid* if $\{\neg\alpha\}$ is not satisfiable. Hintikka states that the valid formulas are precisely the theorems of the logic T. Restricting to transitive model systems gives a characterisation of the theorems of S4, while the symmetric systems determine B and the ones that are both transitive and symmetric determine S5. These assertions apply to the propositional version of the logics. To prove them would require showing in each case that a deductively consistent formula is a member of some model set that belongs to a model system of the appropriate kind, but again the issue of axioms and proof theory is not taken up. The paper is mainly devoted to a discussion of the problem of combining modalities with quantifiers, and proposes various modifications on the closure properties of Ω depending on whether it is required that whatever exists in a particular state of affairs should do so necessarily.

4.8   The Place of Kripke

The earlier efforts to develop the seminal ideas of Kripke semantics have inevitably raised questions of priority. In fact, as the above material is intended to show,

the idea of using a binary relation to model modality occurred independently to a number of people, and for different reasons, with Hintikka being the first to explain it in terms of conceivable alternatives to a given state of affairs. Kanger was the first to recognise the relevance of [Jónsson and Tarski, 1951] to modal logic,[33] and the first to apply this kind of semantical theory to the resolution of philosophical questions about existence and identity.

But it is only in Kripke's writings that we see such seminal ideas developed into an attractive model theory of sufficent power to fully resolve the long-standing issue of a satisfactory semantics for modality and of sufficient generality to advance the field further. A fundamental point (mentioned in section 4.1) is that he was the first to propose, and make effective use of, *arbitrary* set-theoretic structures as models. The methods of Hintikka, Kanger and Montague are all variations on the theme of a binary relation between models of the non-modal fragment of the predicate languages they use. Also, they did not present complete axiomatisations of their semantics. Kripke was the first to do this, and by allowing R to be any relation on any set K, he opened the door to all kinds of model constructions, which were rapidly provided by himself and then others. (His models for non-normal logics appear to lack any historical antecedents.) It is due to his innovation that we now have a model *theory* for intensional logics.

As already noted in section 4.2, Kripke developed his ideas independently. His analysis of S5 was inititiated in 1956 when he was still at high-school (he turned 16 years old on November 13th of that year). From the paper [Prior, 1956] he learned of the axioms for S5, and began to think of modelling that system by truth tables with missing rows (see section 4.1). Early in 1957 E. W. Beth sent him his papers on the method of semantic tableaux, which provided Kripke with a technique for proving completeness theorems. By 1958 Kripke had worked out his relational semantics for modal and intuitionistic systems, as announced in his abstract [1959b] which was received by the editors on 25 August 1958. It was through exploring different conditions connecting tableaux in order to model the different subsystems of S5 that Kripke came to the idea of using a binary relation between worlds as the basis of a model theory.

Kripke had been introduced to Beth by Haskell B. Curry, who wrote to Beth on 24 January 1957 that

> I have recently been in communication with a young man in Omaha Nebraska, named Saul Kripke. ... This young man is a mere boy of 16 years; yet he has read and mastered my Notre Dame Lectures and writes me letters which would do credit to many a professional logician. I have suggested to him that he write you for preprints of your papers which I have already mentioned. These of course will be very difficult for him, but he appears to be a person of extraordinary brilliance, and I have no doubt something will come of it.[34]

---

[33]As Føllesdal [1994] emphasis.

[34]Quoted from [de Jongh and van Ulsen, 1998–1999, pp. 290–291].

The Notre Dame Lectures of [Curry, 1950] presented a number of deductive systems of modal logic, including one equivalent to Lewis's S4 for which a cut elimination theorem was demonstrated in [Curry, 1952]. Other such sources that were influential for Kripke included the McKinsey–Tarski papers and the paper of Lemmon [1957] which showed how to axiomatize the Lewis systems in the style of Gödel.

In late 1958 Kripke entered Harvard University as an undergraduate, and encountered a philosophical environment that was hostile to modal logic. He was advised to abandon the subject and concentrate on majoring in mathematics. This caused the evident delay in publication of his work until the appearance of the major articles of 1963 and 1965.

Looking back over the intervening decades we see the strong influence of Kripke's ideas on many areas of mathematical logic, ranging across the foundations of constructive logic and set theory, substructural logics (including relevance logic, linear logic), provability logic, the Kripke-Joyal semantics in topos theory and numerous logics of transition systems in theoretical computer science.

A *proposition* is defined in [Kripke, 1963a] to be a function from worlds to truth values, while in [1963b] an n-ary predicate letter is modelled as a function from worlds to n-ary relations. Those definitions formed a cornerstone of Montague's approach to intensional logic,[35] and stimulated the substantial development of formal semantics for natural languages in the theories of Montague [1974], Cresswell [1973], Barwise [1989] and others. Kripke's models, and his intuitive descriptions of them, also stimulated many philosophical and formal investigations of the nature of possible worlds, and the questions of existence and identity that they generate (see [Loux, 1979]).

5   THE POST-KRIPKEAN BOOM OF THE SIXTIES

The 1960's was an extraordinary time for the introduction of new model theories. At the beginning of the decade Abraham Robinson created nonstandard analysis by constructing models of the higher-order theory of the real numbers. Then Paul Cohen's invention of forcing revolutionized the study of models of set theory, and freed up the log-jam of questions that had been building since the time of Cantor. Kripke related forcing to his models of Heyting's predicate calculus, and Dana Scott and Robert Solovay re-formulated it as the technique of Boolean-valued models. Scott then replaced "Boolean-valued" by "Heyting-valued" and extended the topological interpretation from intuitionistic predicate logic to intuitionistic real analysis. F. William Lawvere's search for categorical axioms for set theory and the foundations of mathematics and his collaboration with Miles Tierney on axiomatic sheaf theory culminated at the end of the decade in the development of elementary topos theory. This encompassed, in various ways, both classical and intuitionistic higher order logic and set theory, including the models

---

[35]As acknowledged in several places, e.g. [Montague, 1970, fn. 5].

of Kripke, Cohen, Scott, and Solovay, as well as incorporating the sheaf theory of the Grothendieck school of algebraic geometry. Scott's construction of models for the untyped lambda calculus in 1969 was to open up the discipline of denotational semantics for programming languages, as well as stimulating new investigations in lattice theory and topology, and further links with categorical and intuitionistic logic.

The introduction of Kripke models had a revolutionary impact on modal logic itself. Binary relations are much easier to visualise, construct, and manipulate than operators on Boolean algebras. They fall into many naturally definable classes that can be used to define corresponding logics. Here then were the tools that would enable an exhaustive investigation of the subject, and some important new ideas were developed during this period.

5.1   The Lemmon and Scott Collaboration

Pioneers in this investigation were John Lemmon and Dana Scott, who conducted an extensive collaboration. They planned to write a book called *Intensional Logic*, for which Lemmon had drafted some inital chapters when he died in 1966. Scott then made this material available in a mimeographed form [Lemmon and Scott, 1966] which was circulated informally for a number of years, becoming known as the "Lemmon Notes". Eventually it was edited by Scott's student Krister Segerberg, and published as [Lemmon, 1977]. Scott also investigated broad issues of intensional logic (individuals and concepts, possible worlds and indices, intensional relations and operators etc.) in discussion with Montague, Kaplan and others. Some of his ideas were presented in [Scott, 1970]. His considerable influence on the subject has been disseminated through the publications of Lemmon and Segerberg, and is also reported in [Prior, 1967] in relation to tense logic, and in a number of Montague's papers.

The relationship between modal algebras and model structures was first systematically explored in Lemmon's two part article [1966a; 1966b]. Here a model structure has the form $\mathfrak{S} = (K, R, Q)$, with Q playing the role of the set of non-normal ("queer") worlds.[36] Notably absent is Kripke's *real* world $G \in K$. Instead a formula α is said to be *valid* in $\mathfrak{S}$ if in all models on $\mathfrak{S}$, α is *true* (i.e. assigned the value $\top$) at all points of K.

Associated with $\mathfrak{S}$ is the modal algebra $\mathfrak{S}^+$ comprising the powerset Boolean algebra $\mathcal{P}(K)$ with the additive operator

$$f(X) = \{x \in K : x \in Q \text{ or } \exists y \in X(xRy)\}$$

to interpret $\Diamond$. Note that $f(\emptyset) = Q$, so f is a *normal* operator iff K has only normal members. Lemmon proved the result that a formula is valid in $\mathfrak{S}$ iff it is satisfied in the algebra $\mathfrak{S}^+$ with just the element $1 (= K)$ designated. This follows from the natural correspondence between models Φ on $\mathfrak{S}$ and assignments

---

[36]At the time this work was done [Kripke, 1965b] had not appeared, but Lemmon had learned about non-normal worlds in conversation with Kripke.

to propositional variables in $\mathfrak{S}^+$, under which a variable p is assigned the set $\{x : \varPhi(p, x) = \top\} \in \mathfrak{S}^+$. The result itself is an elaboration of the construction in [Kripke, 1963a] of the matrix of propositions associated with any model structure. It remains true for S2-like systems if validity in $\mathfrak{S}$ is confined to truth at normal worlds, and also all elements of $\mathfrak{S}^+$ that include $K - Q$ are designated.

Any finite modal algebra $\mathfrak{A} = (\mathfrak{B}, f)$ is readily shown to be isomorphic to one of the form $\mathfrak{S}^+$, with $\mathfrak{S}$ based on the set of atoms of $\mathfrak{B}$. Combining that observation with McKinsey's finite algebra constructions enabled Lemmon to deduce the completeness of a number of modal logics with respect to validity in their (finite) model structures. For an arbitrary $\mathfrak{A}$ he gave a representation theorem, "due in essentials to Dana Scott", that embeds $\mathfrak{A}$ as a subalgebra of some $\mathfrak{S}^+$. This was done by an extension of Stone's representation of Boolean algebras, basing $\mathfrak{S}$ on the set K of all ultrafilters of $\mathfrak{B}$, with uRt iff $\{fx : x \in t\} \subseteq u$ for all ultrafilters u, t, while $Q = \{x \in K : f0 \in x\}$. Each $x \in \mathfrak{A}$ is represented in $\mathfrak{S}^+$ by the set $\{u \in K : x \in u\}$ of ultrafilters containing x, as in Stone's theory.

In the Lemmon Notes there is a model-theoretic analogue of this representation of modal algebras that has played a pivotal role ever since. Out of any normal logic $\varLambda$ is constructed a model

$$\mathcal{M}^\varLambda = (K^\varLambda, R^\varLambda, \varPhi^\varLambda)$$

in which $K^\varLambda$ is the set of all *maximally $\varLambda$-consistent* sets of formulas, with

$$uR^\varLambda t \quad \text{iff} \quad \{\Diamond\alpha : \alpha \in t\} \subseteq u \quad \text{iff} \quad \{\alpha : \Box\alpha \in u\} \subseteq t,$$

and $\varPhi^\varLambda(p, u) = \top$ iff $p \in u$. The key property of this construction is that an arbitrary formula α is true in $\mathcal{M}^\varLambda$ at u iff $\alpha \in u$. This implies that $\mathcal{M}^\varLambda$ is a *model of* α, i.e. α is true at all points of $\mathcal{M}^\varLambda$, iff α is an $\varLambda$-theorem. Thus $\mathcal{M}^\varLambda$ is a single characteristic model for $\varLambda$, now commonly called the *canonical* $\varLambda$-model. Moreover, the properties of this model are intimately connected with the proof-theory of $\varLambda$. For example, if $(\Box\alpha \to \alpha)$ is an $\varLambda$-theorem for all α, then it follows directly from properties of maximally consistent sets that $R^\varLambda$ is reflexive. This gives a technique for proving that various logics are characterised by suitable conditions on models, a technique that is explored extensively in [Lemmon and Scott, 1966].

If Scott's representation of modal algebras is applied to the Lindenbaum algebra of $\varLambda$, the result is a model structure isomorphic to $(K^\varLambda, R^\varLambda)$. The construction can also be viewed as an adaptation of the method of completeness proof introduced in [Henkin, 1949], and first used for modal logic in [Bayart, 1958] (see section 4.3). There were others who independently applied this approach to the relational semantics for modal logic, including David Makinson [1966] and Max Cresswell [1967], their work being completed in 1965 in both cases. Makinson dealt with propositional systems, while Cresswell's appears to be the first Henkin-style construction of relational models of *quantificational* modal logic. David Kaplan outlined a proof of this kind in his review [1966] of [Kripke, 1963a], explaining that

the idea of adapting Henkin's technique to modal systems had been suggested to him by Dana Scott.

Another construction of lasting importance from the Lemmon Notes is a technique for proving the finite model property by forming *quotients* of the model $\mathcal{M}^L$. To calculate the truth-value of a formula α at points in $\mathcal{M}^A$ we need only know the truth-values of the finitely many subformulas of α. We can regard two members of $\mathcal{M}^A$ as *equivalent* if they assign the same truth-values to all subformulas of α. If there are n such subformulas, then there will be at most 2^n resulting equivalence classes of elements of $\mathcal{M}^A$, even though $\mathcal{M}^A$ itself is uncountably large. Identifying equivalent elements allows $\mathcal{M}^A$ to be collapsed to a *finite* quotient model which will falsify α if $\mathcal{M}^A$ does. This process, which has become known as *filtration*,[37] was first developed in a more set-theoretic way in [Lemmon, 1966b, p. 209] as an alternative to McKinsey's finite algebra construction. In its model-theoretic form it has proven important for completeness proofs as well as for proofs of the finite model property. Some eighteen modal logics were shown to be decidable by this method in [Lemmon and Scott, 1966].

5.2 Bull's Tense Algebra

A singular contribution from the 1960's is the algebraic study by Robert Bull, a student of Arthur Prior,[38] of logics characterised by linearly ordered structures. Prior had observed that the Diodorean temporal reading of $\Box\alpha$ as "α is and always will be true" leads, on intuitive grounds, to a logic that includes S4 but not S5. In his 1956 John Locke Lectures at Oxford on *Time and Modality* (published as [Prior, 1957]) he attempted to give a mathematical precision to this reading by interpreting formulas as sets of sequences of truth values. In effect he was dealing with the complex closure algebra $\mathsf{Cm}(\omega, \leq)$, where $\omega = \{0, 1, 2, \ldots\}$ is the set of natural numbers viewed as a sequence of moments of time. The question became one of identifying the logic that is characterised by this algebra, or equivalently by the model structure $(\omega, \leq)$. Prior called this logic D.[39]

In 1957 Lemmon observed that D includes the formula

$$\Box(\Box p \rightarrow \Box q) \vee \Box(\Box q \rightarrow \Box p),$$

which arises from the intuitionistically *in*valid formula $(p \rightarrow q) \vee (q \rightarrow p)$ by applying the translation of [McKinsey and Tarski, 1948]. Lemmon's formula is therefore not an S4-theorem, and when added as an axiom to S4 produces a system called S4.3. In 1958 Michael Dummett showed that the formula

$$\Box(\Box(p \rightarrow \Box p) \rightarrow \Box p) \rightarrow (\Diamond\Box p \rightarrow \Box p)$$

---

[37]This term was first used in [Segerberg, 1968a], where "canonical model" was also introduced.

[38]Initially at Christchurch, New Zealand, and then at Manchester, England. Bull was one of two graduate students from New Zealand who studied with Prior at Manchester at the beginning of the 1960's. The other was Max Cresswell, who later became the supervisor of the present author.

[39]The letter D later became a label for the system $\mathrm{K}+(\Box p \rightarrow \Diamond p)$, or equivalently $\mathrm{K}+\Diamond\top$, because of its connection with Deontic logic.

also belongs to D, and then Prior [1962b] pointed out that this is due to the *discreteness* of the ordering $\leq$ on ω: if time were a continuous ordering then Dummett's formula would not be valid, but Lemmon's would. In fact the property used by Prior to invalidate Dummett's formula was *density* (between any two moments there is a third) rather than continuity in the sense of Dedekind (no "gaps").

Kripke showed in 1963 that D is exactly the normal logic obtained by adding Dummett's formula as an axiom to S4.3. His proof, using semantic tableaux, is unpublished. Dummett conjectured to Bull that taking time as "continuous" would yield a characterisation of S4.3.[40] Bull proved this in his paper [1965] which, in addition to giving an algebraic proof of Kripke's completeness theorem for D, showed that S4.3 is characterised by the complex algebra of the ordering $(\mathbb{R}^+, \leq)$ of the positive real numbers. He noted that $\mathbb{R}^+$ could be replaced here by the positive rationals, or any linearly ordered set with a subset of order type ω^2. In particular this shows that propositional modal formulas are incapable of expressing the distinction between dense and continuous time under the relational semantics.

Bull made effective use of Birkhoff's fundamental decomposition [Birkhoff, 1944] of an abstract algebra into a subdirect product of subdirectly irreducible algebras. Birkhoff had observed that subdirectly irreducible closure algebras are *well-connected* in the sense of [McKinsey and Tarski, 1944] (see section 3.2). Applying this to Lindenbaum algebras shows that every normal extension of S4 is characterised by well-connected closure algebras, and in the case of extensions of S4.3 the closed ($\mathsf{C}x = x$) elements of a well-connected algebra are linearly ordered. Bull used this fact, together with the strategy of McKinsey's finite algebra construction, to build intricate embeddings of finite S4.3-algebras into $\mathsf{Cm}(\mathbb{R}^+, \leq)$ or $\mathsf{Cm}(\omega, \leq)$. He later refined this technique to establish in [Bull, 1966] one of the more celebrated meta-theorems of modal logic:

every normal extension of S4.3 has the finite model property.

Proofs of this result using relational models were subsequently devised by Kit Fine [1971] and Håkan Franzén (see [Segerberg, 1973]). Fine gave a penetrating analysis of finite S4.3 models to establish that there are exactly $\aleph_0$ normal extension of S4.3, all of which are finitely axiomatisable and hence decidable. Segerberg [1975] proved that in fact every logic extending S4.3 is normal.

The indistinguishability of rational and real time is overcome by passing to the more powerful language of Prior's PF-calculus for tense logic (section 4.4). A model structure for this language would in principle have the form (K, R_P, R_F), with R_P and R_F being binary relations on K interpreting the modalities P and F. But for modelling tense logic, with its *interaction* principles $p \rightarrow GPp$ and $p \rightarrow HFp$, the relations R_P and R_F should be mutually inverse. Thus we continue to use structures (K, R) with the understanding that what we really intend is (K, R^{-1}, R). For linearly ordered structures, the ability of the two modalities

---

[40]See [Prior, 1967, ch. II] as well as [Bull, 1965] for this historical background.

to capture properties "in each direction" of the ordering produces formulas that express the Dedekind continuity of $\mathbb{R}$, a fact that was first realised by Montague and his student Nino Cocchiarella.[41]

Bull applied his algebraic methodology in the [1968] paper to give complete axiomatisations of the tense logics characterised by each of the strictly linearly ordered structures $(\mathbb{Z}, <)$, $(\mathbb{Q}, <)$ and $(\mathbb{R}, <)$. In addition to a common set of axioms for linear orderings without first or last element, for integer time $\mathbb{Z}$ he used the special axiom

$$\Box(Gp \rightarrow p) \rightarrow \Box Gp \vee \Box \neg Gp,$$

where $\Box$ is the S5-modality defined by $\Box \alpha = \alpha \wedge G\alpha \wedge H\alpha$. For rational time $\mathbb{Q}$ this was replaced by the *density* axiom $Fp \rightarrow FFp$. The axiomatisation of real time required the density axiom as well as

$$\Box(Gp \rightarrow PGp) \rightarrow \Box Gp \vee \Box \neg Gp.$$

(The reader may find it instructive to verify that validity of this last formula in any model on $(\mathbb{R}, <)$ depends on the fact that there are no unfilled Dedekind cuts in the real line.) Bull also established that the tense logics of rational and real time have the finite model property, but that the logic of integer time does not.[42]

This is not quite the end of the story about Diodorean modality. Prior made an interesting observation in [Prior, 1967, p. 203] about the (non-linear) temporal ordering of locations in relativistic spacetime. In the Minkowskian spacetime of *special* relativity theory, this ordering is *directed*: for any two locations x, y there is a third that is in the future of both x and y. This is because any two future light-cones eventually intersect (but not so in *general* relativity, where the effect of gravitation can prevent light-cones overlapping). Directedness causes the Diodorean interpretation of $\Box$ to validate the formula $\Diamond \Box p \rightarrow \Box \Diamond p$, which is itself equivalent in the field of S4 to the formula $\Box \neg \Box p \vee \Box \Diamond \Box p$ that arises by the McKinsey–Tarski translation of the intuitionistically invalid $\neg p \vee \neg \neg p$. Adding $\Diamond \Box p \rightarrow \Box \Diamond p$ to S4 gives the logic S4.2. Both S4.2 and S4.3 were introduced in [Dummett and Lemmon, 1959], and shown to have the finite model property in [Bull, 1964].

In [Goldblatt, 1980] a completeness proof is given to show that S4.2 is exactly the Diodorean logic of n-dimensional Minkowski spacetime for all $n \geq 2$, as well as being the logic of the product structure $(\mathbb{R}, \leq) \times (\mathbb{R}, \leq)$.[43] But the problem of axiomatising the PF-calculi characterised by these spacetimes remains open.

5.3   Segerberg's Essay

Krister Segerberg's dissertation, *An Essay in Classical Modal Logic* [1971], provided a comprehensive semantic analysis of whole families of modal logics, as well

---

[41]See [Prior, 1967, pp. 57, 72].

[42]An error in the proof for rational time is corrected in [Bull, 1969].

[43]The latter result was obtained independently by V. B. Shehtman [1983].

as developing important new concepts, some of which had been announced in his papers of [1968a] and [1970]. These works established some notational and terminological conventions that have been lasting. For instance the term $frame$[44] was used in place of $model\ structure$, and the Lemmon–Scott satisfaction notation $\models_x^{\mathcal{M}} \alpha$ was used throughout in place of Kripke's $\Phi(\alpha, x) = \top$, where $\mathcal{M} = (\mathfrak{S}, \Phi)$. Later authors have tended to reduce the use of superscripts and write $\mathcal{M} \models_x \alpha$ instead of $\models_x^{\mathcal{M}} \alpha$. $\mathcal{M} \models \alpha$ then means that α is $true\ in$ $\mathcal{M}$, i.e. true at all points of $\mathcal{M}$, and $\mathfrak{S} \models \alpha$ means that α is valid in the frame $\mathfrak{S}$.

The weakest system discussed in the $Essay$ is E, the smallest logic that is closed under the rule $from\ \alpha \leftrightarrow \beta\ infer\ \Box\alpha \leftrightarrow \Box\beta$. An algebraic semantics for this logic would employ algebras $\mathfrak{A} = (\mathfrak{B}, f)$ having f as a unary function on $\mathfrak{B}$ satisfying no particular conditions. The corresponding "relational" models use $neighbourhood$ $semantics$, the idea of which is attributed to Montague [1968] and Scott [1970]. Segerberg presents this by the device of a $neighbourhood\ frame$ $\mathfrak{S} = (K, N)$, where N, the $neighbourhood\ system$, is a function assigning to each $x \in K$ a collection N_x of subsets of K, called $neighbourhoods\ of\ x$.[45] Writing $\mathcal{M}(\alpha)$ for the "truth set" $\{y \in K : \mathcal{M} \models_y \alpha\}$ interpreting α in $\mathcal{M}$, the satisfaction clause for $\Box$ in a model $\mathcal{M}$ on such a frame $\mathfrak{S}$ is

$$\mathcal{M} \models_x \Box\alpha \quad \text{iff} \quad \mathcal{M}(\alpha) \in N_x.$$

A topology on K has a naturally associated neighbourhood system in which $X \in N_x$ iff x is interior to X, i.e. $x \in U \subseteq X$ for some open set U. In this case $\mathcal{M}(\Box\alpha)$ is the topological interior of $\mathcal{M}(\alpha)$, and the result is an S4-model. But different logics can be characterised by validity in frames with weaker conditions imposed on their neighbourhoods. A relational frame (K, R) is equivalent to the neighbourhood frame (K, N) having $U \in N_x$ iff $\{y : xRy\} \subseteq U$.

Any neighbourhood frame (K, N) has an associated algebra $(\mathcal{P}(K), f^N)$, where the operation f^N, interpreting $\Box$ on the powerset algebra $\mathcal{P}(K)$, is given by

$$f^N(X) = \{x \in K : X \in N_x\}.$$

Inversely, any function $f : \mathcal{P}(K) \to \mathcal{P}(K)$ induces the neighbourhood system N^f on K, where

$$X \in N_x^f \quad \text{iff} \quad x \in f(X).$$

Thus, whereas Jónsson and Tarski's analysis shows that relational semantics corresponds to $completely\ additive\ and\ normal$ operators on powerset algebras (see section 3.3), neighbourhood systems can be used to represent arbitrary operations on such algebras. The relationship between neighbourhood frames and modal algebras has been systematically investigated by Kosta Došen [1989].

Filtration (see section 5.1) was used extensively by Segerberg to prove completeness theorems. This technique can be effective in dealing with logics whose

---

[44]This term was suggested to Segerberg by Scott.
[45]Some authors use a relation $R \subseteq K \times \mathcal{P}(K)$ in place of N, where xRU iff $U \in N_x$.

canonical model does not satisfy some desired property, and comes into its own when seeking to axiomatise logics defined by some condition on *finite* frames. For example, Segerberg showed [1971, p. 68] that the normal logic K4W,[46] with axioms

$$4: \quad \Box p \to \Box\Box p$$
$$\text{W}: \quad \Box(\Box p \to p) \to \Box p,$$

is characterised by the class of finite frames (K, R) in which R is transitive and irreflexive, i.e. a strict ordering. (This logic later proved important in studies of the provability interpretation of modality. See section 7.5.) The basic method was to obtain a falsifying model for a given non-theorem by filtration of the canonical model, and then to "deform" this into a model of the desired kind without affecting the truth value of the formula concerned. This involved an analysis of the way a transitive relations presents itself as an ordered set of connected components, called *clusters*. The method was applied in the *Essay* and the [1970] paper to axiomatise a whole range of logics, including those characterised by the classes of finite partial orderings, finite linear orderings (both irreflexive and reflexive), and the modal and tense logics of the structures (K, R) where K is any of ω, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, while R is any of $<$, $>$, $\leq$, and $\geq$.

The logic characterised by the class of all finite partial orderings is particularly significant. Segerberg proved [1971, p. 101] that it is S4Grz, the normal logic axiomatised by adding to S4 the axiom

$$\text{Grz}: \quad \Box(\Box(p \to \Box p) \to p) \to p.$$

He named this for Andrzej Grzegorczyk whose paper [1967] added a further insight to the relationship between intuitionistic and modal logic. Grzegorczyk showed that the formula

$$[((p \dashv\vdash \Box q) \dashv\vdash \Box q) \wedge ((\neg p \dashv\vdash \Box q) \dashv\vdash \Box q)] \dashv\vdash \Box q$$

is not a theorem of S4 (nor indeed of S5), and when added to S4 gives a system into which the intuitionistic logic IPC can be translated by the Gödel–McKinsey–Tarski procedures. The translation of a propositional formula is an S4-theorem iff it is a theorem of Grzegorczyk's stronger logic, which is deductively equivalent to S4Grz.

Segerberg initiated the use of truth-preserving maps between relational models and frames in [1968a]. Given models $\mathcal{M}$ and $\mathcal{M}'$ on frames $\mathfrak{S} = (K, R)$ and $\mathfrak{S}' = (K', R')$ respectively, a function φ from K *onto* K' was called a *pseudo-epimorphism* from $\mathcal{M}$ to $\mathcal{M}'$ if

(i) xRy implies $\varphi(x)R'\varphi(y)$,

(ii) $\varphi(x)R'\varphi(y)$ implies $\exists z \in K(xRz \ \& \ \varphi(z) = \varphi(y))$, and

---

[46]K4W could be called KW, since the axiom 4: $\Box p \to \Box\Box p$ is deducible from W, as was shown independently by several people, including de Jongh, Kripke and Sambin.

(iii) $\mathcal{M} \models_x p$ iff $\mathcal{M}' \models_{\varphi(x)} p$.

For such a function every formula α has $\mathcal{M} \models_x \alpha$ iff $\mathcal{M}' \models_{\varphi(x)} \alpha$, so if $\mathcal{M}$ is a model of α, then $\mathcal{M}'$ will be also. From this it can be shown that if α is valid in $\mathfrak{S}$, then the existence of a function from K onto K' satisfying (i) and (ii) implies that α is valid in $\mathfrak{S}'$ as well.[47]

The name "pseudo-epimorphism" was shortened to "p-morphism" by Segerberg in [1970; 1971] and this uninformative term has been very widely adopted, even for functions that are not surjective but, in place of (ii), satisfy

(ii′) $\varphi(x)R'w$ implies $\exists z \in K(xRz \ \& \ \varphi(z) = w)$.

The notion was generalised by Johan van Benthem [1976a] to that of a "p-relation" between models, which is itself intimately related to the concept of a *bisimulation relation* that has been fundamental to the study of computational processes (see section 7.2).

There is another explanation of why functions of this type are natural and important in the modal context. Any function $\varphi : K \to K'$ induces the function $\varphi^+ : \mathcal{P}(K') \to \mathcal{P}(K)$ in the reverse direction, taking each subset X of K' to its inverse image $\{x \in K : \varphi(x) \in X\}$. This φ^+ is a Boolean algebra homomorphism. The conditions (i) and (ii′) are precisely what is required for it to preserve the operators f_R and $f_{R'}$, and hence be a homomorphism between the *modal* algebras $\mathsf{Cm}(K', R')$ and $\mathsf{Cm}(K, R)$. If φ is surjective, then φ^+ is injective and so makes $\mathsf{Cm}\,\mathfrak{S}'$ isomorphic to a subalgebra of $\mathsf{Cm}\,\mathfrak{S}$. Hence all modal-algebraic equations satisfied by $\mathsf{Cm}\,\mathfrak{S}$ will be satisfied by $\mathsf{Cm}\,\mathfrak{S}'$. But a propositional modal formula α can be viewed as a *term* in the language of the algebra $\mathsf{Cm}\,\mathfrak{S}$, with α being valid in the frame $\mathfrak{S}$ precisely when the algebraic equation "$\alpha \approx 1$" is satisfied by $\mathsf{Cm}\,\mathfrak{S}$. This gives another perspective on why validity is preserved by surjective p-morphisms.

Of equal importance is the validity-preserving notion of *subframe*. This originated in Kripke's definition in [1963a] of a model structure (G, K, R) as being *connected* when $K = \{H : GR^*H\}$, where R^* is the reflexive-transitive closure of R. Lemmon adapted this in his [1966b] to the notion of the *connected model structure* $\mathfrak{S}_x$ *generated from* $\mathfrak{S}$ *by an element* x, which is the substructure of $\mathfrak{S}$ based on $\{y : xR^*y\}$. He observed that a formula falsified by $\mathsf{Cm}\,\mathfrak{S}$ must be falsified by $\mathsf{Cm}\,\mathfrak{S}_x$ for some x. Segerberg showed in [1971, p. 36] that a model $\mathcal{M}$ on $\mathfrak{S}$ can be restricted to a model $\mathcal{M}_x$ on $\mathfrak{S}_x$ (the submodel of $\mathcal{M}$ *generated by* x) in such a way that in general $\mathcal{M}_x \models_y \alpha$ iff $\mathcal{M} \models_y \alpha$. From this it follows that any formula valid in $\mathfrak{S}$ will be valid in $\mathfrak{S}_x$, and conversely a formula valid in $\mathfrak{S}_x$ for all x in $\mathfrak{S}$ will be valid in $\mathfrak{S}$ itself (as essentially observed by Lemmon). This notion of *point-generated* substructure turned out to be the relational analogue of the notion of *subdirectly irreducible* algebra. Indeed the algebra $\mathsf{Cm}\,\mathfrak{S}$ is subdirectly

---

[47]A surjection between partial orderings that satisfies (i) and (ii) was defined to be *strongly isotone* in [de Jongh and Troelstra, 1966], where the notion was used to demonstrate connections between partial orderings and certain algebraic models for intuitionistic propositional logic.

irreducible iff $\mathfrak{S}$ is equal to $\mathfrak{S}_x$ for some x, a fact that was first demonstrated by Wim Blok [1978b, p. 12], [1980, Lemma 4.1].

A frame $\mathfrak{S}$ is a *subframe* of frame $\mathfrak{S}'$ if it is a substructure of $\mathfrak{S}'$ that is closed under R', i.e. if $x \in K$, then $\{y \in K' : xR'y\} \subseteq K$ (some authors call this a "generated" subframe even though there is no longer any generator involved). Then the inclusion function $\varphi : K \hookrightarrow K'$ is a p-morphism inducing φ^+ as a *surjective* homomorphism from $\mathsf{Cm}\,\mathfrak{S}'$ to $\mathsf{Cm}\,\mathfrak{S}$. Since equations are preserved by surjective homomorphisms, modal-validity is preserved in passing from $\mathfrak{S}'$ to the subframe $\mathfrak{S}$.

The disjoint union $\coprod_J \mathfrak{S}_j$ of a collection $\{\mathfrak{S}_j : j \in J\}$ of frames also preserves validity. The construction was first applied to modal model theory in [Goldblatt, 1974] and [Fine, 1975b]. $\coprod_J \mathfrak{S}_j$ is simply the union of a collection of pairwise disjoint copies of the $\mathfrak{S}_j$'s. Each $\mathfrak{S}_j$ is isomorphic to a subframe of $\coprod_J \mathfrak{S}_j$, and so the above properties of subframes guarantee that a formula is valid in $\coprod_J \mathfrak{S}_j$ iff it is valid in every $\mathfrak{S}_j$.

These observations about morphisms, subframes and disjoint unions form the basis of a theory of *duality* between frames and modal algebras that is discussed in section 6.5.

6   METATHEORY OF THE SEVENTIES AND BEYOND

The semantic analysis of particular logics eventually gave way to investigations of the nature of the relational semantics itself: the strengths and limitations of its techniques, and its relationship to other formalisms, particularly first-order and monadic second-order predicate logic. Some of the questions raised have yet to be answered.

Throughout chapter 6 the term "logic" will always mean a *normal* logic.

6.1   *Incompleteness*

A logic Λ is *sound* with respect to a class $\mathcal{C}$ of frames if every member of $\mathcal{C}$ is a Λ-*frame*, i.e. validates all Λ-theorems. By definition Λ is sound with respect to the class $Fr(\Lambda)$ of all Λ-frames. In the converse direction, Λ is *complete* with respect to $\mathcal{C}$ if any formula that is valid in all members of $\mathcal{C}$ is a Λ-theorem. For example, every normal logic is complete with respect to $\mathcal{C} = \{\mathfrak{S}^\Lambda\}$, where $\mathfrak{S}^\Lambda = (K^\Lambda, R^\Lambda)$ is the *canonical frame* of Λ as defined in section 5.1. For if a formula is valid in $\mathfrak{S}^\Lambda$, then it is true in the canonical model $\mathcal{M}^\Lambda$ on $\mathfrak{S}^\Lambda$, and so is a Λ-theorem. Whether or not Λ is sound with respect to $\mathfrak{S}^\Lambda$ is an important issue that will be discussed in section 6.6.

A logic Λ is *characterised* by a class $\mathcal{C}$ if it is both sound and complete with respect to $\mathcal{C}$. Λ is *complete* per se if it is complete with respect to some class $\mathcal{C}$ of Λ-*frames*, in which case it is characterised by that $\mathcal{C}$, as well as by the class $Fr(\Lambda)$ of all Λ-frames. It is important to recognise that a given logic may be characterised by many different classes. For example, S4 is characterised by each of the class of

all quasi-orderings, the class of finite quasi-orderings, and the class of all partial-orderings (but not the *finite* partial-orderings, which characterise S4Grz as we saw in section 5.3).

Lemmon was sufficiently taken with the power of Kripke semantics to conjecture that every normal logic is characterised by some class of relational frames [Lemmon, 1977, p. 74]. It turned out that this was as far from the truth as it could be. Wim Blok showed that, in a manner which will be explained below, "most" logics Λ are not characterised by any class of frames, and hence are *incomplete* in the sense that there exist formulas that are valid in all Λ-frames but are not Λ-theorems.

The first example of an incomplete logic was devised by Steven Thomason [1972b], and is a readily described tense logic in Prior's PF-language. In addition to a set of postulates for linearly-ordered frames it has the axioms

$$Gp \rightarrow Fp$$
$$Pp \rightarrow P(p \land \neg Pp)$$
$$GFp \rightarrow FGp.$$

The first of these is valid in a frame (K, R) only if the "endless time" condition $\forall x \exists y (xRy)$ is satisfied. The second axiom is equivalent to $H(Hp \rightarrow p) \rightarrow Hp$, which is Segerberg's axiom W for the past modality H. Its validity entails that R is irreflexive. Thus if x_0 is a point in any frame validating the first two axioms, $\{y : x_0 Ry\}$ is an irreflexive linear ordering with no last element. Interpreting p as a set such that both it and its complement are unbounded in $\{y : x_0 Ry\}$ then gives a model on the frame that falsifies the third axiom at x_0. In this model the truth-value of p alternates forever over time.

Thus Thomason's logic is not valid on any frame whatsoever! In other words it is indistinguishable in terms of frame-validity from the *inconsistent* logic in which all formulas are theorems. But it is not itself inconsistent, because it is satisfied by the algebra which consists of all the finite and cofinite subsets of the structure $(\omega, <)$. In this algebra the interpretation of each formula is constrained to cease changing with time.

It proved more difficult to devise incomplete $\Box$-logics, i.e. propositional logics in a language with just one modality $\Box$. Unlike tense logic, any consistent normal $\Box$-logic is validated by some frame, and in fact by some one-element frame. There are two such structures: $\mathfrak{S}_\circ$ is the one consisting of a single reflexive point, while $\mathfrak{S}_\bullet$ consists of a single irreflexive point. $\mathfrak{S}_\circ$ characterises the normal logic $\Lambda_\circ = \mathrm{K} + (\Box p \leftrightarrow p)$ and $\mathfrak{S}_\bullet$ characterises $\Lambda_\bullet = \mathrm{K} + \Box\bot$, both of which are *maximal* logics in the sense of having no proper consistent extensions. Makinson [1971] proved that every consistent normal $\Box$-logic is either valid in $\mathfrak{S}_\circ$ or valid in $\mathfrak{S}_\bullet$ and so is a sublogic of one of $\Lambda_\circ$ and $\Lambda_\bullet$.

The first incomplete $\Box$-logics were found by Thomason [1974a] and Kit Fine [1974], who independently constructed some rather complicated examples. Later van Benthem [1978; 1979] found some simpler ones. The simplest unearthed to

date is the normal logic with axiom

$$\Box(\Box p \leftrightarrow p) \rightarrow \Box p.$$

Lon Berk showed that any frame validating this formula also validates Segerberg's axiom W, while Roberto Magari showed that W is not a theorem of the logic. Proofs of these results are presented in [Boolos and Sambin, 1985].

The *degree of incompleteness* of a logic Λ was defined by Fine [1974] as the number of logics that are valid in exactly the same frames that Λ is. For any class $\mathcal{C}$, the set $\Lambda_{\mathcal{C}} = \{\alpha : \mathcal{C} \models \alpha\}$ of all formulas validated by $\mathcal{C}$ is, by definition, characterised by $\mathcal{C}$. If some other logic Λ is valid in all members of $\mathcal{C}$ and no other frames, then Λ must be a proper sublogic of $\Lambda_{\mathcal{C}}$, with both having degree of incompleteness ≥ 2. The logic K has degree 1: it is the only logic valid in all frames whatsoever. Any Λ that has degree 1 must be complete, since it must be equal to $\Lambda_{\mathcal{C}}$ where $\mathcal{C}$ is the class of all Λ-frames. Fine asked which cardinals can occur as the degree of incompleteness of some logic, and whether there are any logics other than K that are "intrinsically complete" in the sense of having degree 1.

Those questions were resolved in a remarkable way by Blok, who proved that any logic Λ containing the axiom $\Box p \rightarrow p$ must have degree of incompleteness $2^{\aleph_0}$, so that there are uncountably many different logics which are indistinguishable from Λ by the Kripke relational semantics. The same applies whenever Λ contains the axiom $\Box^n p \leftrightarrow \Box^{n+1} p$ for some natural number n. As just one illustration of this situation, consider the case of $\Lambda_\circ$ itself. The only connected $\Lambda_\circ$-frame is the one-element reflexive frame $\mathfrak{S}_\circ$ (and any other $\Lambda_\circ$-frame is just a disjoint union of copies of $\mathfrak{S}_\circ$). But there are uncountably many other (incomplete) logics whose only connected validating frame is also $\mathfrak{S}_\circ$.

These results were obtained in 1979–1977, and published in [Blok, 1980]. The report [Blok, 1978b] then gave the following complete answer to Fine's two questions: every normal logic is either of degree 1 or of degree $2^{\aleph_0}$, and there are $2^{\aleph_0}$ logics of degree 1. The degree 1 logics all have the finite model property. Moreover Blok provided a semantic characterisation of these degree 1 logics, using the notion of a *splitting logic*. This is a logic Λ_s for which there is some other logic Λ_s' such that every logic Λ has either $\Lambda_s \subseteq \Lambda$ or $\Lambda \subseteq \Lambda_s'$, but not both. Thus the collection of all normal logics is split into the two disjoint collections $\{\Lambda : \Lambda_s \subseteq \Lambda\}$ and $\{\Lambda : \Lambda \subseteq \Lambda_s'\}$. A simple example is given by putting $\Lambda_s = \mathrm{K} + \Diamond\top$ and $\Lambda_s' = \Lambda_\bullet = \mathrm{K} + \Box\bot$. If $\Lambda \not\subseteq \Lambda_\bullet$, then by the maximality of $\Lambda_\bullet$, $\Box\bot$ cannot be consistently added to Λ, hence its negation $\Diamond\top$ is a Λ-theorem, showing $\mathrm{K} + \Diamond\top \subseteq \Lambda$.

Let $\Lambda/\mathfrak{S}$ be the intersection of all logics that are not validated by frame $\mathfrak{S}$. Then a logic is a splitting logic iff it is equal to the logic $\Lambda/\mathfrak{S}$ for some finite frame $\mathfrak{S}$ that is generated from a point and has $\mathfrak{S} \models \Box^n\bot$ for some n. The last condition holds for a finite $\mathfrak{S}$ iff $\mathfrak{S}$ is *circuit-free*, i.e. it includes no sequence of the form $x_1 R x_2 \cdots R x_k R x_1$ for any k. If $\Lambda_s = \Lambda/\mathfrak{S}$ is a splitting logic, then the corresponding Λ_s' is the logic $\{\alpha : \mathfrak{S} \models \alpha\}$ characterised by $\mathfrak{S}$.

Every splitting logic is of degree 1, and is finitely axiomatisable. A logic Λ is

of degree 1 if and only if it is a *join of splitting logics*, i.e. is equal to the least logic that includes the splitting logics $\Lambda/\mathfrak{S}$ for all $\mathfrak{S}$ in some collection $\mathcal{C}$ of finite generated circuit-free frames. This is the same as requiring that Λ be the least logic not validated by any member of $\mathcal{C}$.

Blok used algebraic methods, studying *varieties*, or equationally defined classes, of modal algebras rather than normal logics directly. He applied some powerful new techniques, including the splitting notion that had been developed in lattice theory by Ralph McKenzie [1972], and an important lemma of Jónsson [1967] characterising subdirectly irreducible algebras in congruence distributive varieties.

Blok's resolution of the issue of incompleteness for Kripke semantics was announced in his abstract [1978a], but his report [Blok, 1978b] giving the detailed proofs was not published. Model-theoretic accounts of the results may be found in [Chagrov and Zakharyaschev, 1997, ch. 10] and [Kracht, 1999, ch. 7].

The issue of the adequacy of neighbourhood semantics (see section 5.3) was investigated in a series of papers by Martin Gerson [1975a; 1975b; 1976], who showed that the two logics of [Thomason, 1974a] and [Fine, 1974], which are not characterised by their relational frames, are also incomplete with respect to their neighbourhood frames. He then gave examples of normal logics that are complete under the neighbourhood semantics but not complete for any class of relational frames. These possibilities can also be revealingly expressed in terms of algebraic semantics, beginning with the observation that complete and atomic Boolean algebras are, up to isomorphism, the same thing as powerset algebras. As we observed in section 5.3, relational frames correspond to completely additive and normal operators on powerset algebras, while neighbourhood frames represent arbitrary operations on such algebras. Thus a logic that is incomplete for the relational semantics is one that is not characterised by those of its complete and atomic algebras whose operators are completely additive and normal; while a logic that is incomplete for the neighbourhood semantics is one that is not characterised by complete and atomic algebras at all.

6.2   *Decidability and Complexity*

The finite model property does not give a universal method for proving the decidability of modal logics. Although every finitely axiomatisable logic with the finite model property is decidable, the converse is not true. This was shown by Dov Gabbay, building on some earlier work of Makinson [1969] which had exhibited the first example of a normal logic that lacked the finite model property. Makinson's example is a proper sublogic of S4, but all of its finite algebras satisfy S4 as well.

Gabbay's paper [1972] extended Makinson's idea to produce finitely axiomatisable modal and tense logics that lacked the finite model property, but could still be shown to be decidable by appealing to a powerful result of Michael Rabin [1969]. This concerns the decidability of monadic second-order theories of successor functions, and has many applications. For each ordinal n with $2 \leq n \leq \omega$, consider

the structure

$$\mathfrak{S}_n = (T_n, \{s_m : m < n\}, \leq, \preccurlyeq),$$

where T_n is the n-ary branching tree of all finite sequences of elements of the set $[n] = \{m \in \omega : m < n\}$, s_m is the successor function $x \mapsto xm$ on the tree, $\leq$ is the "initial segment" ordering of sequences, and $\preccurlyeq$ is their lexicographical ordering induced by the natural ordering $<$ on $[n]$. Rabin proved that the monadic second-order theory SnS of the structure $\mathfrak{S}_n$ is decidable. To do this he developed a theory of finite-state automata that process infinite labelled trees, and established the decidability of the *emptiness problem* of whether any given automaton accepts at least one tree. The decidability of SnS was then reduced to this emptyness problem. It was later shown that the decision problem for SnS is intractable: Albert Meyer [1975] proved that no algorithm for deciding if a sentence is in SnS can run in elementary time, i.e. time bounded by some fixed number of compositions of exponential functions.

Gabbay developed a method of coding Kripke models into the structure $\mathfrak{S}_\omega$ and thereby reducing the decidability problem for certain logics to Rabin's decidability results for $S\omega S$. The technique is explained in Part 5 of the book [Gabbay, 1976], where it is used to establish decidability results for many modal systems.

Gabbay's method was later used by Cresswell [1984] in adapting an incomplete logic from [van Benthem, 1979] to construct a decidable modal logic that is finitely axiomatisable but incomplete with respect to Kripke frames (and hence lacks the finite model property). Cresswell's example is a proper sublogic of the logic characterised by the class of finite strict linear orderings, but the two logics are validated by exactly the same frames.

For any logic Λ, the problem of deciding if a given formula is Λ-provable is the same as the Λ-*validity problem* of deciding if a given formula is true in all models $\mathcal{M}$ such that $\mathcal{M} \models \Lambda$. The Λ-*satisfiability problem* of whether a given formula is true at some point of some Λ-model is equivalent to the validity problem in the sense that α is Λ-satisfiable iff its negation $\neg\alpha$ is not Λ-valid. Thus a deterministic algorithm that solved the validity problem could be used to solve the satisfiability problem, and vice versa. But if nondeterministic algorithms are considered, the two problems may differ as to their computational complexity. The classic example of this concerns the set of non-modal propositional formulas. Satisfiability of any of these can be tested in nondeterministic polynomial time. But the same is not known for validity: to test the validity of a formula with n variables appears to require examination of all 2^n truth-value assignments to these variables.

To discuss this further, recall that NPTIME, or more briefly NP, is (informally) the class of all problems that are solvable by a nondeterministic algorithm whose running time for any execution is bounded above by some polynomial function of the length of the input. Co-NP is the class of problems whose complement is in NP. The Λ-satisfiability problem is in NP iff the Λ-validity problem is in co-NP. The satisfiability of non-modal formulas is NP-*hard*, meaning that any problem in NP has a polynomial-time reduction to this problem [Cook, 1971]. The Λ-satisfiability problem for any consistent modal logic Λ is therefore also NP-hard.

Since non-modal satisfiability itself belongs to NP, it is said to be an NP-*complete* problem.

PSPACE is the class of problems solvable by a deterministic algorithm using an amount of space that is polynomially bounded by the length of the input. PSPACE includes NPTIME and is closed under complementation. It is also known that any *non*deterministic polynomially space-bounded algorithm is equivalent to a deterministic one [Savitch, 1970]. Thus

$$\text{NP} \subseteq \text{PSPACE} = \text{co-PSPACE} = \text{NPSPACE}.$$

It is not known if the stated inclusion is proper, but it is widely believed that PSPACE-complete problems are not in NP.

Richard Ladner [1977] applied these concepts to determine computational complexities of some of the basic normal modal logics. He showed that the satisfiability problem for each of the logics K, T, and S4 is in PSPACE, by optimising the space requirements of the decision procedures from [Kripke, 1963a]. Hence the provability problems for these logics is in PSPACE as well. He proved further that any problem in PSPACE has a polynomial time reduction[48] to the provability problem of any normal sublogic of S4. Thus provability for any of these logics is PSPACE-hard, and for K, T, and S4 it is PSPACE-complete. The method used was to reduce to Λ-provability a known PSPACE-complete problem, namely the validity of quantified non-modal propositional formulas.

The logic S5 is more tractable than the sublogics of S4. Ladner showed that S5-satisfiability is in NP, and therefore is NP-complete. The key to this result is that S5 has the *poly-size model property*: poly-size model property any non-theorem is falsifiable in a model whose size is a polynomial in the size of the formula. Edith Spaan [1993] extended this to prove that every one of the ($\aleph_0$ many) extensions of the logic S4.3 has the poly-size model property and has an NP-complete satisfiability problem. On the other hand Joseph Halpern and Yoram Moses [1985; 1992] showed that satisfiability for any logic having at least two S5-modalities is PSPACE-hard.

As to *un*decidability, there must be undecidable logics because there are uncountably many logics altogether but only countably many algorithms. In [Thomason, 1975d] an undecidable modal logic is exhibited that is finitely axiomatisable, and so cannot have the finite model property. This was produced by encoding a presentation of a recursive function with undecidable range into a model of a logic with a large number of temporal modalities, and then reducing this to a logic with one modality by methods that are described below in section 6.4.

The question of how undecidable a logic can be was answered by Alasdair Urquhart [1981] who showed that for any set X of natural numbers there exists a normal modal logic Λ_X such that the decision problem for X is reducible to

---

[48] Actually he showed that these reductions are in "log-space": they have a space requirement bounded by a logarithmic function of the length of the input. This implies a polynomial time-bound. Ladner originally proved the reduction result for T and for S4, and subsequently used an argument of S. K. Thomason to extend it to all normal sublogics of S4.

that of Λ_X. Urquhart used this to construct a logic with the finite model property that has a decidable set of axioms but is undecidable. Spaan [1993] showed that there are (uncountably many) undecidable logics that have the poly-size model property.

Undecidability of quantificational modal logic was considered by Kripke [1962] in an early application of his model theory from [1959a]. Whereas the first-order calculus of monadic predicates is decidable, the modal monadic calculus turns out to be undecidable. Kripke showed that the decision problem for provability of non-modal first-order formulas in a binary predicate R, which is known to be undecidable, is reducible to that of modal formulas in two monadic predicates P and Q, by replacing $R(x, y)$ by $\Diamond(P(x) \wedge Q(y))$. This applies to any modal system which is a sublogic of the quantificational version of S5 of [Kripke, 1959a] and which obeys certain general rules satisfied by all then known systems and "probably by the vast majority of those that will be proposed in the future".

6.3   First-Order Definability

Validity of a modal formula α in a relational frame $\mathfrak{S} = (K, R)$ is an intrinsically *second-order* concept. α is valid when true at all points in all models on $\mathfrak{S}$. Since a model interprets each propositional variable p in α as a subset of K, this amounts to treating p as a set variable, or a *monadic predicate* variable. Meredith's U-calculus associates with α a formula $(\alpha)x$ in the first-order language of $\mathfrak{S}$, with x as its sole free individual variable. If the propositional variables of α are $p_1, \ldots, p_k$, then regarding these as set variables we have that α is valid in $\mathfrak{S}$ iff $\mathfrak{S}$ is a model of the sentence

$$\forall p_1 \cdots \forall p_k \forall x \, (\alpha)x$$

of the monadic second-order language of a binary predicate, i.e. the second-order language in which all the second-order variables are monadic. This is a simple kind of second-order sentence, technically known as Π_1^1, with all its second-order quantifiers being universal and at the front.

Some modal formulas express properties that are well-recognised as being second-order in nature. For example, Segerberg's axiom W is valid in $\mathfrak{S}$ iff R^{-1} is transitive and *well-founded* (see [Boolos, 1979, p. 82]). However, a substantial reason for the great success of the relational semantics is that many logics were shown to be to be characterised by frames satisfying simple *first-order* conditions on R, like reflexivity, transitivity, linearity etc. To consider this phenomenon, recall that a class of relational frames is called *elementary* if it is definable in first-order logic, i.e. if it is the class of all models of some set of sentences in the first-order language of a binary predicate R. A *basic elementary* class is one that is defined by a single first-order sentence.[49] A modal logic is (*basic*) *elementary* if it is characterised by some (basic) elementary class of frames.

---

[49]Some authors use "Δ-elementary" in place of "elementary", and "elementary" in place of "basic elementary".

The Lemmon Notes provided many examples of basic elementary logics, and formulated a conjecture about the situation, which will now be briefly described. First we say that a modal formula is *positive* if it can be built from propositional variables using only the connectives $\wedge$, $\vee$, $\Diamond$, and $\Box$. If β is any positive formula with variables $p_1, \ldots, p_k$ and $m = (m_1, \ldots, m_k)$ and $n = (n_1, \ldots, n_k)$ are any k-tuples of natural numbers, consider the formula

$$\beta_n^m : \quad \Diamond^{m_1}\Box^{n_1}p_1 \wedge \cdots \wedge \Diamond^{m_k}\Box^{n_k}p_k \rightarrow \beta.$$

Associated with β_n^m is a certain first-order condition $R\beta_n^m$ on binary relations, which can be read off from the formation of β_n^m itself. The conjecture was that the normal logic axiomatised by adding β_n^m to K is characterised by the basic elementary class of frames satisfying $R\beta_n^m$ (see [Lemmon, 1977, p. 78]). This was confirmed independently by the present author and Henrik Sahlqvist in 1973 [Goldblatt, 1974; 1975b; Sahlqvist, 1975], but Sahlqvist generalised the result considerably to consider any formula of the type $\Box^n(\alpha \rightarrow \beta)$ where $n \geq 0$, β is positive, and α is constructed from propositional variables and/or their negations using only the connectives $\wedge$, $\vee$, $\Diamond$, $\Box$ in such a way that no positive occurrence of a variable is in a subformula that has $\wedge$, $\vee$, or $\Diamond$ within the scope of a $\Box$. He proved that the class of frames validating such a formula is definable by an explicit first-order sentence, and that this basic elementary class characterises the normal logic axiomatised by adding the formula to K. The result has been extensively analysed and extended to "polymodal" logics and to equational classes of BAO's in general: see [Sambin and Vaccaro, 1989; Jónsson, 1994; de Rijke and Venema, 1995; Givant and Venema, 1999].

The simplest formula not covered by Sahlqvist's scheme is

$$\text{M}: \quad \Box\Diamond p \rightarrow \Diamond\Box p,$$

commonly known as the *McKinsey axiom*.[50] This is the $\Box$-version of the formula $GFp \rightarrow FGp$ that figures as an axiom in Thomason's incomplete tense logic. In the Lemmon Notes a proof was given that the normal logic S4+M is characterised by the elementary class of all quasi-ordered frames satisfying the condition

$$\forall x \exists y (xRy \wedge \forall z(yRz \rightarrow y = z)).$$

Segerberg [1968a] then showed that this logic has the finite model property and is characterised by the finite quasi-orders satisfying this condition. But the status of the logic K+M remained unresolved.

It turned out that the class of all frames validating the McKinsey axiom is not elementary, let alone basic elementary. This was proved in [Goldblatt, 1974, §17], which showed further that no elementary class can characterise the logic K+M, and indeed any class that does characterise this logic must fail to be closed

---

[50]This is something of a misnomer. The system S4+$\Box\Diamond p \wedge \Box\Diamond q \rightarrow \Diamond(p \wedge q)$ was investigated by McKinsey [1945], who called it S4.1. Sobociński [1964] showed that it is the same as S4+($\Box\Diamond p \rightarrow \Diamond\Box p$), and renamed it K1, since it is not a subsystem of S4.2.

under ultraproducts. Van Benthem [1975] gave a Löwenheim-Skolem argument to show that the class of all frames validating M is not even closed under elementary equivalence.[51] On the other hand Fine [1975a] proved that the logic K+M is in some respects quite well-behaved: it has the finite model property, so is decidable and is characterised by its (finite) validating frames.

From such examples the question naturally arises of when the collection $Fr(\alpha) = \{\mathfrak{S} : \mathfrak{S} \models \alpha\}$ of all frames validating the formula α is an elementary class. To answer this, note first that the complement of $Fr(\alpha)$ is always closed under ultraproducts. That can be shown directly, or by observing that the complement of $Fr(\alpha)$ is defined by an existential second-order sentence

$$\exists p_1 \cdots \exists p_k \exists x \, \neg(\alpha)x$$

of the kind (Σ^1_1) that is always preserved by ultraproducts.[52] From this it follows by the Keisler-Shelah characterisation of elementary classes[53] that $Fr(\alpha)$ *is elementary iff it is basic elementary iff it is closed under ultraproducts* [Goldblatt, 1974; 1975a]. But then van Benthem discovered a striking strengthening of the result:

$Fr(\alpha)$ *is basic elementary iff it is closed under elementary equivalence.*

This means that any class of the form $Fr(\alpha)$ is quite special: if it is closed under ultra*powers* then it must be closed under ultraproducts. VanBenthem's proof was an interesting model-theoretic compactness argument,[54] but in his published version [van Benthem, 1976b] he used instead a subsequent argument of the present author, namely that there is an injective p-morphism

$$\left(\textstyle\prod_J \mathfrak{S}_j\right)/F \longrightarrow \left(\textstyle\coprod_J \mathfrak{S}_j\right)^J/F$$

of any ultraproduct of frames $\mathfrak{S}_j$ into the associated ultrapower of their disjoint union $\coprod_J \mathfrak{S}_j$, and this maps the ultraproduct isomorphically onto a *subframe* of the ultrapower. Since $Fr(\alpha)$ is invariably closed under disjoint unions, subframes and isomorphism, the desired result follows immediately from this embedding. But the argument also works for the class $Fr(\Lambda)$ of all frames validating a *set* Λ of formulas, to show that

$Fr(\Lambda)$ *is elementary iff it is closed under elementary equivalence.*

The study of the definability of modal formulas in predicate logic was dubbed *Correspondence Theory* by van Benthem [1976a], who gave further expositions of this theory in his works of [1983] and [1984].

---

[51]Two structures are elementarily equivalent when they satisfy the same first-order sentences.
[52][Chang and Keisler, 1973, Corollary 4.1.14].
[53][Chang and Keisler, 1973, Corollary 6.1.16].
[54]A discussion of van Benthem's original proof is presented in [Goldblatt, 1999].

6.4 Thomason's Second-Order Reduction

A deep investigation of the expressive power of modal semantics was made by Thomason in a series of papers [1974b; 1975b; 1975c; 1975d] reporting work, carried out in 1973, that constitutes a tour de force of model-theoretical analysis in combination with coding techniques of the kind used in recursion theory. This confirmed his belief, expressed earlier in [1972a], that

> propositional modal logic (with the usual relational semantics) must be understood as a rather strong fragment of (classical) second-order predicate logic.

A "logic" is taken to consist of a symbolic language together with a semantic interpretation specifying when a formula is *valid* in a *structure*. **M** is the logic given by the language of propositional modal logic with the semantics based on frames (K, R) as structures, while **T** is the propositional tense logic of Prior's PF-language with structures (K, R^{-1}, R). Each logic determines a *logical consequence* relation $\Gamma \models \alpha$ between sets of formulas Γ and formulas α, meaning that α is valid in every structure in which all members of Γ are valid. Thomason proved in [1972a] that the Compactness Theorem fails in **M** for this relation: there is a case of an α which is a logical consequence of some set Γ but not of any finite subset of Γ. In the paper [1975b] he showed that there is a **T**-formula γ whose set $\{\alpha : \gamma \models \alpha\}$ of logical consequences is not effectively enumerable, and has a high degree of undecidability—technically what is known as a *complete Π_1^1 set*. Moreover γ is *categorical* in the sense that all its connected validating structures are isomorphic. In addition, for $0 \leq m < \omega + \omega$ there is a categorical formula γ_m whose unique validating structure has size $\beth_m$, where $\beth_0 = \aleph_0$, $\beth_{m+1} = 2^{\beth_m}$, and $\beth_\omega = \lim\{\beth_m : m < \omega\}$. The formula γ describes a structure which encodes presentations of certain recursive functions that define a complete Π_1^1 set. The formulas γ_m describe structures that encode copies of the iterated powersets ω, $\mathcal{P}(\omega), \mathcal{P}(\mathcal{P}(\omega)), \ldots$. The proofs of these facts are reminiscent of the arithmetisation procedures and expressibility results involved in Gödel's incompleteness theorems, and graphically illustrate the expressive power of **T**. The facts themselves are quite contrary to the situation in first-order logic, where the logical consequences of a given sentence are effectively enumerable, and no sentence with an infinite model is categorical.

A logic L_1 is said to be *reducible* to a logic L_2 if there exists an L_2-formula δ and an effective transformation ψ of L_1-formulas to L_2-formulas such that for every collection $\Gamma \cup \{\alpha\}$ of L_1-formulas,

$$\Gamma \models \alpha \quad \text{iff} \quad \{\delta\} \cup \{\psi(\gamma) : \gamma \in \Gamma\} \models \psi(\alpha).$$

This definition captures the idea that L_1 can be regarded as a fragment of the logic L_2, and is motivated by a notion of interpretation of one first-order theory in another that appears in [Shoenfield, 1967]. Here δ may be thought of as describing a certain structure, with $\psi(\gamma)$ asserting that γ is valid in that structure. In [Thomason, 1974b] it is shown that tense logic **T** is reducible to modal logic **M**. The

formula δ used for this has the property that for any **T**-structure $\mathfrak{S} = (K, R^{-1}, R)$ there is an **M**-structure $\mathfrak{S}'$ that contains within it definable copies of (K, R) and (K, R^{-1}) in such a way that "P" statements about $\mathfrak{S}$ can be interpreted as "$\diamondsuit$" statements about $\mathfrak{S}'$. Applying this reduction to the results about **T** from [1975b], Thomason concludes that there is an **M**-formula whose set of logical consequences is a complete Π_1^1 set.

The full monadic second-order theory **S** of a binary predicate is shown to be reducible to **M** in [Thomason, 1975c]. For this purpose the logic $\mathbf{T}_n$ of n temporal orderings is introduced. It has n pairs of modalities $P_1, F_1, \ldots, P_n, F_n$, and structures having n binary relations and their inverses to interpret these connectives. It is shown that for $n > 1$, $\mathbf{T}_n$ is reducible to $\mathbf{T}_{n-1}$. Since reducibility is a transitive relation, it follows that each $\mathbf{T}_n$ is reducible to **T** $(= \mathbf{T}_1)$, and hence reducible to **M**. This is then applied to prove the reducibility of **S**. The argument involves defining a $\mathbf{T}_{15}$-formula δ with the property that for each frame $\mathfrak{S} = (K, R)$ there is a model of δ with 15 temporal orderings that includes within it definable copies of $\mathfrak{S}$; the powerset $\mathcal{P}(K)$; the membership relation from K to $\mathcal{P}(K)$; the set of all (codes for) **S**-formulas, the set of all assignments in K and $\mathcal{P}(K)$ to the individual and set variables of **S**; and the satisfaction relation between **S**-formulas and assignments in $\mathfrak{S}$ as a second-order model. This leads to a reduction of **S** to $\mathbf{T}_{15}$, which can then be combined with the reduction of $\mathbf{T}_{15}$ to **M** to give the desired result. Thomason concludes that

> the logical consequence relation of propositional modal logic (with the Kripke relational semantics) is as complex as it could possibly be.

6.5   Duality and the Calculus of Class Operations

The keystone constructions in the general theory of algebras are homomorphic images, subalgebras, and direct products. The famous Variety Theorem due to Garrett Birkhoff [1935] states that a class of abstract algebras is a *variety*, i.e. is definable by equations, iff it is closed under these three constructions. The standard convention in this subject is to use the letters H, S and P for the operations that assign to each class of algebras its closure under homomorphic images, subalgebras, and direct products, respectively. Thus Birkhoff's theorem states that a class $\mathcal{A}$ of algebras is a variety if and only if $\mathsf{H}\mathcal{A} \subseteq \mathcal{A}$ and $\mathsf{S}\mathcal{A} \subseteq \mathcal{A}$ and $\mathsf{P}\mathcal{A} \subseteq \mathcal{A}$. A refinement due to Tarski [1946; 1955a] is that for each class $\mathcal{A}$ of algebras, $\mathsf{HSP}\mathcal{A}$ is the smallest variety that includes $\mathcal{A}$. Hence $\mathsf{HSP}\mathcal{A}$ is known as the variety *generated* by $\mathcal{A}$.

The corresponding constructions for relational modal semantics are subframes, p-morphic images, and disjoint unions. As explained in section 5.3, a p-morphism $\varphi : \mathfrak{S} \to \mathfrak{S}'$ induces an algebraic homomorphism $\varphi^+ : \mathsf{Cm}\,\mathfrak{S}' \to \mathsf{Cm}\,\mathfrak{S}$, allowing us to show that if $\mathfrak{S}$ is (isomorphic to) a subframe of $\mathfrak{S}'$ then $\mathsf{Cm}\,\mathfrak{S}$ is a homomorphic image of $\mathsf{Cm}\,\mathfrak{S}'$, and if $\mathfrak{S}'$ is a p-morphic image of $\mathfrak{S}$ then $\mathsf{Cm}\,\mathfrak{S}'$ is (isomorphic to) a subalgebra of $\mathsf{Cm}\,\mathfrak{S}$. Disjoint unions of structures correspond naturally to

direct products of algebras via an isomorphism

$$\mathsf{Cm}\coprod_J \mathfrak{S}_j \cong \prod_J \mathsf{Cm}\,\mathfrak{S}_j \tag{1}$$

between the complex algebra of a disjoint union and the direct product of the complex algebras of its factors.

The assignments $\mathfrak{S} \mapsto \mathsf{Cm}\,\mathfrak{S}$ and $\varphi \mapsto \varphi^+$ form a contravariant functor from the category **Frm** of frames and p-morphisms to the category **Malg** of normal modal algebras and homomorphisms. In the reverse direction there is a construction that assigns to each normal BAO $\mathfrak{A}$ a certain relational structure $\mathbb{C}\mathsf{st}\,\mathfrak{A}$, called the *canonical structure* of $\mathfrak{A}$, whose points are the ultrafilters of $\mathfrak{A}$. The complex algebra $\mathsf{Em}\,\mathfrak{A} = \mathsf{Cm}\,\mathbb{C}\mathsf{st}\,\mathfrak{A}$ of this structure is the *canonical embedding algebra* of $\mathfrak{A}$, and is isomorphic to the perfect extension $\mathfrak{A}^\sigma$, as described in section 3.3. The Jónsson–Tarski representation of $\mathfrak{A}$ amounts to the fact that there is an injective homomorphism $\mathfrak{A} \rightarrowtail \mathsf{Em}\,\mathfrak{A}$.

When applied to modal algebras, the assignment $\mathfrak{A} \mapsto \mathbb{C}\mathsf{st}\,\mathfrak{A}$ gives rise to a contravariant functor from **Malg** to **Frm** that takes each homomorphism $\theta : \mathfrak{A} \to \mathfrak{A}'$ to a p-morphism $\mathbb{C}\mathsf{st}\,\mathfrak{A}' \to \mathbb{C}\mathsf{st}\,\mathfrak{A}$ which maps each ultrafilter of $\mathfrak{A}'$ to its θ-inverse image in $\mathfrak{A}$. These functors provide a *duality* between frames and modal algebras. It is not however a dual *equivalence*, because we do not in general have $\mathfrak{S}$ isomorphic to $\mathbb{C}\mathsf{st}\,\mathsf{Cm}\,\mathfrak{S}$, or $\mathfrak{A}$ isomorphic to $\mathsf{Cm}\,\mathbb{C}\mathsf{st}\,\mathfrak{A}$: the assignment $\mathfrak{S} \mapsto \mathsf{Cm}\,\mathfrak{S}$ increases cardinality, as does $\mathfrak{A} \mapsto \mathbb{C}\mathsf{st}\,\mathfrak{A}$ for infinite $\mathfrak{A}$.

The category **Frm** is dually equivalent to the category of *complete and atomic* modal algebras with $\sum$-preserving homomorphisms [Thomason, 1975a]. To obtain a category of structures equivalent to **Malg** it is necessary to modify the notion of "frame". A first attempt at this was made by Makinson [1970] who defined a *relational model* as a structure $(K, R, \mathcal{H})$, where $\mathcal{H}$ is a collection of truth-valuations $\varPhi$ on (K, R) in Kripke's sense that satisfies certain closure properties. That did not produce a full equivalence between algebras and models. A language independent-approach was taken by Thomason [1972b] who defined a "first-order semantics" using structures $\mathfrak{S} = (K, R, P)$, where P is a collection of subsets of K that forms a subalgebra of the full complex algebra $\mathsf{Cm}\,(K, R)$. This subalgebra is taken in place of $\mathsf{Cm}\,(K, R)$ as the algebra assigned to $\mathfrak{S}$. Validity in $\mathfrak{S}$ is defined as truth in all models $\mathcal{M} = (\mathfrak{S}, \varPhi)$ on $\mathfrak{S}$ satisfying the constraint that the set $\mathcal{M}(p) = \{x : \varPhi(p, x) = \top\}$ belongs to P for all variables p.

By imposing suitable restrictions on P, essentially set-theoretic versions of the conditions (i)–(iii) of section 3.3 that defined the Jónsson-Tarski perfect extensions, a notion of "descriptive" frame (K, R, P) is arrived at. This theory was developed in [Goldblatt, 1974], where the descriptive frames were shown to form a category dually equivalent to **Malg**. A topological approach to duality for closure algebras and quasi-orderings was independently investigated by Leo Èsakia [1974].

Connections between relational structures and algebras can be conveniently expressed in the "calculus" of class operations. We use the symbols $\mathbb{S}$, $\mathbb{H}$, and $\mathbb{U}\mathsf{d}$ for the operations of closing a class of structures under subframes, p-morphic images, and disjoint unions, respectively. $\mathbb{P}\mathsf{u}$ and $\mathbb{P}\mathsf{w}$ are used for closure under

ultraproducts and ultrapowers, while

$$\mathsf{Cm}\,\mathcal{C} = \{\mathfrak{A} : \mathfrak{A} \cong \mathsf{Cm}\,\mathfrak{S} \text{ for some } \mathfrak{S} \in \mathcal{C}\}$$

is the class of all (isomorphic copies of) complex algebras of structures in the class $\mathcal{C}$. Then the isomorphism (1) above implies that $\mathsf{Cm}\,\mathbb{Ud}\,\mathcal{C} = \mathsf{P}\,\mathsf{Cm}\,\mathcal{C}$ for any class $\mathcal{C}$ of frames. Similarly, the representation

$$\left(\textstyle\prod_J \mathfrak{S}_j\right)/F \longrightarrow \left(\textstyle\coprod_J \mathfrak{S}_j\right)^J/F$$

from section 6.3 of an ultraproduct of frames as a subframe of an ultrapower of a disjoint union yields the conclusion that in general

$$\mathbb{Pu}\,\mathcal{C} \subseteq \mathbb{SPw}\,\mathbb{Ud}\,\mathcal{C}.$$

There are numerous properties that can be express in this way using class operations, for example

$$\mathbb{S}\,\mathbb{H}\,\mathcal{C} \subseteq \mathbb{H}\,\mathbb{S}\,\mathcal{C}, \quad \mathsf{S}\,\mathsf{Cm}\,\mathbb{H}\,\mathcal{C} = \mathsf{S}\,\mathsf{Cm}\,\mathcal{C}, \quad \mathbb{S}\,\mathbb{Ud}\,\mathcal{C} = \mathbb{Ud}\,\mathbb{S}\,\mathcal{C}, \quad \mathbb{Pu}\,\mathbb{S}\,\mathbb{H}\,\mathcal{C} \subseteq \mathbb{H}\,\mathbb{S}\,\mathbb{Pu}\,\mathcal{C}.$$

An inventory of such facts may be found in [Goldblatt, 1995; 2000].

Dual to the formation of the algebra $\mathsf{Em}\,\mathfrak{A} = \mathsf{Cm}\,\mathbb{Cst}\,\mathfrak{A}$ is the association with any structure $\mathfrak{S}$ of its *canonical extension* $\mathbb{Ex}\,\mathfrak{S} = \mathbb{Cst}\,\mathsf{Cm}\,\mathfrak{S}$, a structure whose points are the ultrafilters on the underlying set of $\mathfrak{S}$ (hence $\mathbb{Ex}\,\mathfrak{S}$ is sometimes called the *ultrafilter extension* of $\mathfrak{S}$). There is a p-morphism

$$\mathfrak{S}^J/F \twoheadrightarrow \mathbb{Ex}\,\mathfrak{S}$$

from a suitably chosen ultrapower of any given frame $\mathfrak{S}$ onto $\mathbb{Ex}\,\mathfrak{S}$, yielding the observation that in general

$$\mathbb{Ex}\,\mathcal{C} \subseteq \mathbb{H}\,\mathbb{Pw}\,\mathcal{C}. \tag{2}$$

The proof of this requires the choice of a sufficiently saturated ultrapower of $\mathfrak{S}$ [Goldblatt, 1989, §3.6] and is motivated by a model construction of [Fine, 1975b] that is discussed further in the next section.

Duality can be used to bring methods of universal algebra to bear on relational semantics. A notable example is the problem of characterising classes of the form $Fr(\Lambda)$, the class of all frames validating a set Λ of modal formulas. The question of when $Fr(\Lambda)$ is elementary was discussed in section 6.3. It is natural to ask, conversely, for conditions under which a given elementary class of frames is equal to the class $Fr(\Lambda)$ for some Λ. The following answer was given in [Goldblatt and Thomason, 1975], where the $\mathbb{Ex}$ construction was first introduced (see also [Goldblatt, 1993, 1.20.6], [Goldblatt, 1989, 3.7.6(2)]).

> If $\mathcal{C}$ is an elementary class of frames, then $\mathcal{C}$ is equal to $Fr(\Lambda)$ for some set Λ of modal formulas if, and only if,
>
>   1. $\mathcal{C}$ is closed under disjoint unions, p-morphic images and subframes; and

>   *2. the complement of $\mathcal{C}$ is closed under canonical extensions, i.e.*
> $\mathbb{Ex}\mathfrak{S} \in \mathcal{C}$ *implies* $\mathfrak{S} \in \mathcal{C}$.

The proof applies the Birkhoff–Tarski analysis of varieties to the variety generated by $\mathsf{Cm}\mathcal{C}$, and uses the construction for (2) above to show that if $\mathcal{C}$ is elementary and closed under p-morphic images then it is closed under canonical extensions.

Duality theory has been developed for arbitrary relational structures and BAO's by using suitable generalisations of p-morphisms and subframes, called "bounded" morphisms and "inner" substructures (Goldblatt [1989; 1995]). This provides algebraic and relational semantics for *polymodal* languages having n-ary connectives which generate formulas $\Box(\alpha_1, \ldots, \alpha_n)$ for $n > 1$. Most of the ideas and results we have discussed about completeness, canonicity, elementarity, class operations etc. carry over to this broader context and apply to cylindric algebras, relation algebras and other kinds of BAO's in addition to modal algebras. This reveals that, mathematically, much of modal semantics is just the case $n = 1$ of a broader structural theory of finitary operators on lattices. A survey of this general theory is given in [Goldblatt, 2000].

If Λ is a normal logic, then the class $V(\Lambda)$ of modal algebras that satisfy all Λ-theorems is a variety. Algebraic constructions in $V(\Lambda)$ provide tools for studying metalogical questions about Λ, such as whether it fulfills analogues of the Beth Definability Theorem and the Craig Interpolation Theorem. This is related to amalgamation properties of algebras in $V(\Lambda)$, as has been shown by Larisa Maksimova, whose article of [1992] gives an account of the subject and further references to the literature.

6.6   Canonicity

A logic Λ is called *canonical* if it is valid in its canonical frame $\mathfrak{S}^\Lambda$, in which case it is characterised by this frame, and so is complete. Almost all proofs that a particular logic is *elementary* have consisted of a demonstration that $\mathfrak{S}^\Lambda$ satisfies some first-order conditions that imply validity of Λ. Such a proof establishes also that Λ is canonical, a conclusion that is inescapable in view of the following profound results of Kit Fine [1975b].

(i)   If the class $Fr(\Lambda)$ of all Λ-frames is closed under elementary equivalence and characterises Λ (i.e. Λ is complete), then Λ is canonical.

(ii)   If Λ is elementary (i.e. characterised by *some* elementary class), then Λ is canonical.[55]

In fact something much stronger was proved. We have been using a language for propositional modal logic that is based on a countably infinite set of variables, but

---

[55]At the time, (i) was not recognised as a consequence of (ii). However, as explained at the end of section 6.3, it was later discovered that closure of $Fr(\Lambda)$ under elementary equivalence implies the ostensibly stronger assertion that $Fr(\Lambda)$ is elementary. So (ii) does imply (i).

we could consider larger languages by assuming we have available a variable p_ξ for each ordinal ξ. Then for a given ordinal η we can generate the set $Form(\eta)$ of modal formulas having variables from the set $\{p_\xi : \xi < \eta\}$. A logic Λ as originally conceived is a subset of $Form(\omega)$, but it has a manifestation $\Lambda_\eta \subseteq Form(\eta)$ for each η, obtained by closing Λ under uniform substitution in $Form(\eta)$ when $\omega < \eta$, and by putting $\Lambda_\eta = \Lambda \cap Form(\eta)$ when $\eta < \omega$. Then we can define a canonical frame $\mathfrak{S}_\eta^\Lambda$ for each η, based on the maximally Λ_η-consistent subsets of $Form(\eta)$. $\mathfrak{S}_\eta^\Lambda$ is of cardinality $2^{\mathrm{card}\eta}$. If it validates Λ_η, we say that Λ is η-canonical.

Fine proved that under each of the hypotheses given in (i) and (ii), Λ is η-canonical for all ordinals η. He also gave an example of a logic that is η-canonical for all η, and is elementary, but for which $Fr(\Lambda)$ is not closed under elementary equivalence. Thus the converse of (i) is false.

The idea of the proof of (i) was to use disjoint unions to obtain a single model $\mathcal{M}$ that characterised Λ_η and was based on a Λ_η-frame, then to view $\mathcal{M}$ as a first-order model and take a *saturated* elementary extension of it that could be mapped onto the canonical frame $\mathfrak{S}_\eta^\Lambda$ by a p-morphism. This was the first application of saturated models to modal logic, and it motivated the construction for result (2) of the previous section. The proof of (ii) combined it with an additional ultraproduct construction.

Canonicity of a logic Λ is intimately connected with the question of whether satisfaction of Λ is preserved by perfect extensions $\mathsf{Em}\,\mathfrak{A} = \mathsf{Cm}\,\mathbb{C}\mathsf{st}\,\mathfrak{A}$ of algebras or canonical extensions $\mathbb{E}\mathsf{x}\,\mathfrak{S} = \mathbb{C}\mathsf{st}\,\mathsf{Cm}\,\mathfrak{S}$ of frames. VanBenthem [1980] refined the proof of Fine's result (ii) above to show that

> if a logic Λ is elementary, then the class $Fr(\Lambda)$ of all Λ-frames is closed under canonical extensions, i.e. $\mathfrak{S} \models \Lambda$ implies $\mathbb{E}\mathsf{x}\,\mathfrak{S} \models \Lambda$.

Another way to describe this conclusion is to say that if $Alg(\Lambda)$ is the variety (equational class) of all modal algebras satisfying Λ, then in general $\mathsf{Cm}\,\mathfrak{S} \in Alg(\Lambda)$ implies $\mathsf{Cm}\,\mathbb{E}\mathsf{x}\,\mathfrak{S} \in Alg(\Lambda)$. But $\mathsf{Cm}\,\mathbb{E}\mathsf{x}\,\mathfrak{S} = \mathsf{Em}\,\mathsf{Cm}\,\mathfrak{S}$, so the conclusion says that $Alg(\Lambda)$ contains the canonical embedding algebras of all its full complex algebras. This can then be strengthened, by applying duality theory, to show that $Alg(\Lambda)$ contains the algebra $\mathsf{Em}\,\mathfrak{A}$ for any of its members $\mathfrak{A}$ [Goldblatt, 1989, Theorem 3.5.5]. Actually, to conclude that $Alg(\Lambda)$ is closed under canonical embedding algebras it is enough to know that Λ is valid in the canonical frame $\mathfrak{S}_\kappa^\Lambda$ for all infinite cardinals κ. This follows by duality from the fact that $\mathfrak{S}_\kappa^\Lambda$ is isomorphic to the canonical structure $\mathbb{C}\mathsf{st}\,\mathfrak{A}_\kappa$, where $\mathfrak{A}_\kappa$ is the free algebra in $Alg(\Lambda)$ on κ-many generators, together with the fact that each member of $Alg(\Lambda)$ is a homomorphic image of some such free algebra.

Ultimately this analysis can be generalised to any kind of Boolean algebra with operators, to yield the following result:

> if $\mathcal{C}$ is any class of relational structures of the same type that is closed under ultraproducts, then the variety of BAO's generated by the class of algebras $\mathsf{Cm}\,\mathcal{C}$ is closed under canonical embedding algebras.

This theorem was first formulated in [Goldblatt, 1989, Theorem 3.6.7], with a proof that used the important result of [Jónsson, 1967] on subdirectly irreducible algebras in congruence-distributive varieties and an obscure diagonal construction on ultraproducts. An entirely different argument was given in [Goldblatt, 1991b] and analysed further in [Goldblatt, 1995]. It used the fact (2) from the previous section, i.e. $\mathbb{E}\mathsf{x}\mathcal{C} \subseteq \mathbb{H}\mathbb{P}\mathsf{w}\mathcal{C}$, and another formula,

$$\mathbb{C}\mathsf{st}\,\mathsf{HSPCm}\mathcal{C} \subseteq \mathbb{S}\mathbb{H}\mathbb{U}\mathsf{d}\mathbb{P}\mathsf{u}\mathcal{C},$$

which shows how the canonical structures of algebras from the variety generated by $\mathsf{Cm}\mathcal{C}$ can themselves be built from members of $\mathcal{C}$. When $\mathcal{C}$ is closed under ultraproducts, so that $\mathbb{P}\mathsf{u}\mathcal{C} = \mathcal{C}$, this takes the form

$$\mathfrak{A} \in \mathsf{HSPCm}\mathcal{C} \quad \text{implies} \quad \mathbb{C}\mathsf{st}\,\mathfrak{A} \in \mathbb{S}\mathbb{H}\mathbb{U}\mathsf{d}\mathcal{C},$$

showing how canonical structures mediate between the dual operations on algebras and structures. This result in turn depends on another fundamental fact,

$$\mathbb{P}\mathsf{u}\,\mathbb{U}\mathsf{b}\mathcal{C} \subseteq \mathbb{U}\mathsf{b}\mathbb{P}\mathsf{u}\mathcal{C},$$

which states that the ultraproduct operation commutes with *bounded* unions. A structure $\mathfrak{S}$ is the bounded union of a collection $\{\mathfrak{S}_j : j \in J\}$ if the $\mathfrak{S}_j$'s are all inner substructures (subframes) of $\mathfrak{S}$ and their union is $\mathfrak{S}$ itself. This notion is dual to that of *subdirect* product, and indeed in the situation just described there is a subdirect product representation

$$\mathsf{Cm}\,\mathfrak{S} \rightarrowtail \prod_J \mathsf{Cm}\,\mathfrak{S}_j$$

of $\mathsf{Cm}\,\mathfrak{S}$ induced by the surjections $\mathsf{Cm}\,\mathfrak{S} \twoheadrightarrow \mathsf{Cm}\,\mathfrak{S}_j$ [Goldblatt, 2000, §4.5].

The first example of non-canonicity in the modal context occurs in [Kripke, 1967], where it is stated that Dummett's Diodorean axiom

$$\Box(\Box(p \to \Box p) \to \Box p) \to (\Diamond\Box p \to \Box p)$$

is not preserved by the Jónsson–Tarski representation of modal algebras. The McKinsey axiom $\Box\Diamond p \to \Diamond\Box p$ was shown not to be canonical in [Goldblatt, 1991a].

The formulas of Sahlqvist(see 6.3) define logics Λ for which the class $Fr(\Lambda)$ is elementary and includes all the canonical frames $\mathfrak{S}_\eta^\Lambda$. These formulas have been generalized by Maarten de Rijke and Yde Venema [1995], who defined *Sahlqvist equations* for any type of BAO and showed that the structures $\mathfrak{S}$ whose complex algebras $\mathsf{Cm}\,\mathfrak{S}$ satisfy such an equation form a basic elementary class. Jónsson [1994] has refined the techniques of [Jónsson and Tarski, 1951] to develop an elegant algebraic proof that varieties of BAO's defined by Sahlqvist equations are closed under canonical embedding algebras.

Fine's theorem (ii) was strengthened by the present author to show that if Λ is characterised by some elementary class then it is valid, not just in any canonical frame $\mathfrak{S}_\eta^\Lambda$, but also in any frame that is elementarily equivalent to a canonical

frame. In fact an even stronger generalization of (ii) can be obtained by restricting attention to *quasi-modal* sentences. These are first-order sentences of the syntactic form $\forall v \varphi$, with φ being constructed from amongst atomic formulas and the constants $\perp$ (False) and $\top$ (True) using at most $\wedge$ (conjunction), $\vee$ (disjunction), and the *bounded* universal and existential quantifiers forms $\forall z(yRz \rightarrow \psi)$ and $\exists z(yRz \wedge \psi)$ with $y \neq z$. The relevance of quasi-modal sentences, and the reason for the name, is that they are precisely those first-order sentences whose satisfaction is preserved by the basic modal-validity preserving operations of $\mathbb{S}$, $\mathbb{H}$, and $\mathbb{U}\mathsf{d}$ [van Benthem, 1983; Goldblatt, 1989]. By the *quasi-modal theory* of a structure $\mathfrak{S}$ we mean the set of all quasi-modal first-order sentences that are true in $\mathfrak{S}$.

It transpires that there is no quasi-modally-expressible property that can differentiate the canonical frames of a logic Λ: the structures $\mathfrak{S}^{\Lambda}_{\eta}$ have exactly the same quasi-modal first-order theory for all η. We will denote this unique quasi-modal theory of the canonical Λ-frames by Ψ^{Λ}. Moreover, if Λ is not canonical, then it always has a largest canonical proper sublogic Λ^c and a largest elementary sublogic Λ^e (with $\Lambda^e \subseteq \Lambda^c$), and the quasi-modal theories Ψ^{Λ^e} and Ψ^{Λ^c} of these other logics are identical to Ψ^{Λ}. These results are all proven in [Goldblatt, 2001a]. The strengthening of Fine's result is as follows [Goldblatt, 1993, 11.4.2]:

> If a modal logic Λ is characterized by some elementary class of frames, then it is characterized by the elementary class of all models of the quasi-modal first-order theory Ψ^{Λ} (which includes all the canonical frames of Λ).

Fine asked if the converse of his (ii) was true: if a logic is canonical, must it be characterised by an elementary class? The algebraic version of this question asks whether a variety of BAO's that is closed under canonical embedding algebras must be generated by the complex algebras of some elementary class of relational structures. This remained a perplexing open problem for three decades, during which time a positive answer was found for all of the canonically closed varieties of modal algebras, cylindric algebras and relation algebras that had been investigated. Eventually however it was discovered that the converse of (ii) fails in general, and does so as badly as it could. This is shown by Goldblatt, Hodkinson and Venema [2004; 2003], exhibiting $2^{\aleph_0}$ different canonical logics that are not characterised by any elementary class. These examples all have the finite model property. They include logics of every degree of unsolvability, and in particular undecidable logics with decidable sets of axioms. Some of the examples are based on ideas from the proof of the non-canonicity of the McKinsey axiom, while others use constructions from the theory of graph colouring, and are related to the modal logic KMT studied by George Hughes [1990]. The validating frames for KMT can be described as those directed graphs satisfying the non-elementary condition that the set $\{y : xRy\}$ of children of any node x has no finite colouring. The logic has an infinite sequence of axioms whose n-th member rules out colourings that use n colours. But KMT is also characterised by the *elementary* class of graphs whose edge relation R satisfies

$\forall x \exists y (x R y R y)$, meaning that every node has a reflexive child. The canonical KMT-frame satisfies this condition.

Some of the logics that violate the converse of (ii) also have axioms that impose reflexive points on canonical frames. But now a canonical frame is essentially the disjoint union of a family of directed graphs, and it is only the *infinite* members of the family that are required to have a reflexive point to ensure canonicity. This is a non-elementary requirement. The proof that the logics are never elementarily characterised involves a famous piece of graph theory of Paul Erdős [1959], who showed that for each integer n there is a finite graph $\mathcal{G}_n$ whose chromatic number and girth are both greater than n, the girth being the length of the shortest cycle in the graph and the chromatic number being the smallest number of colours needed to colour it. The essence of the application is that if a certain logic Λ were characterised by an elementary class $\mathcal{C}$, and infinitely many of the $\mathcal{G}_n$'s validated Λ, then by a compactness argument it would follow that $\mathcal{C}$ contained an infinite graph that had no cycles of odd length. But such a graph can be coloured using only *two* colours, a property that invalidates one of the axioms defining Λ. Hence the existence of $\mathcal{C}$ is impossible.

7   SOME MATHEMATICAL MODALITIES

The seed of relational semantics sown in the 1950's has grown into a tree with many branches. The most notable new dimension of activity beyond that already described has been the application of relational modal semantics to a range of formalisms of computational and mathematical interest. This final section will briefly survey some studies of this kind, providing a sketch of the key ideas and a guide to the literature.

7.1   *Dynamic Logic of Programs*

Dynamic logic was invented by Vaughan Pratt, who described its origins in [1980a] as follows.

> In the spring of 1974 I was teaching a class on the semantics and axiomatics of programming languages. At the suggestion of one of the students, R. Moore, I considered applying modal logic to a formal treatment of a construct due to C. A. R. Hoare, "$p\{a\}q$", which expresses the notion that if p holds before executing program a, then q holds afterwards. Although I was skeptical at first, a weekend with Hughes and Cresswell convinced me that a most harmonious union between modal logic and programs was possible. The union promised to be of interest to computer scientists because of the power and mathematical elegance of the treatment. It also seemed likely to interest modal logicians because it made a well-motivated and potentially very fruitful connection between modal logic and Tarski's calculus of binary relations.[56]

---

[56]The "weekend" reference is of course to the classic text of [Hughes and Cresswell, 1968].

Pratt's idea was to assign a box-modality $[\pi]$ to each program π, with the formula $[\pi]\alpha$ being read "after π, α". Then Hoare's construct[57] $p\{\pi\}q$ can be defined as $p \to [\pi]q$, but more complex assertions about program correctness and termination can be formalised by combining $[\pi]$ with other connectives, including modalities for other programs. The connective $[\pi]$ is interpreted, not as an accessibility relation between possible worlds, but as a transition relation R_π between "possible execution states", with $xR_\pi y$ when there is an execution of π that starts in state x and terminates in state y. The dual modality $\langle\pi\rangle\alpha$, definable as $\neg[\pi]\neg\alpha$, asserts that there is an execution of π that terminates with α true. In particular, $\langle\pi\rangle\top$ asserts that there exists a terminating execution of program π.

Pratt's first paper [1976] describes a predicate language with modalities for a class of programs generated from basic assignments and tests by a number of operations, including *alternation* $\pi \cup \pi'$ and *composition* $\pi; \pi'$. The interpreting relations for programs satisfy appropriate conditions, including $R_{\pi\cup\pi'} = R_\pi \cup R_{\pi'}$ and $R_{\pi;\pi'} = R_\pi \circ R_{\pi'}$. A complete axiomatisation was presented for the language of these "loop-free" programs, and then the class of *regular* programs was defined by adding the *iteration* construct π^*, with interpretation $R_{\pi^*} =$ reflexive transitive closure of R_π. The universal quantifier $\forall x$ was identified with a modality $[x \leftarrow \mathrm{RANDOM}]$ corresponding to a random assignment to the variable x.

The purely propositional fragment of this language was isolated by Michael Fisher and Richard Ladner [1977; 1979] who defined the system PDL of *propositional dynamic logic of regular programs*. Its programs are generated from some set of atomic commands by the operations of alternation, composition and iteration. A Kripke model for PDL assigns a binary relation to each atomic program, and then interprets complex programs by the above conditions on $R_{\pi\cup\pi'}$, $R_{\pi;\pi'}$ and R_{π^*}. Fischer and Ladner proved that this semantically defined logic has the finite model property by a version of the filtration construction. That method produces a falsifying model for a given non-theorem α whose size is exponential in the length of α. The result was used to establish an upper bound of *nondeterministic exponential time* for the complexity of the satisfiability problem: there is a nondeterministic algorithm for deciding PDL-satisfiability that runs in a time bounded above by an exponential function c^n of the length n of the formula concerned (for some constant c). They also gave a lower bound of *deterministic* exponential time for the complexity of this problem: there is a constant $d > 1$ such that no deterministic algorithm can decide the satisfiability question for all formulas in time less than d^n. The technique used was to construct a PDL-formula that encodes the computations of a certain kind of Turing machine that was known to require exponential running time. The gap between these upper and lower bounds was closed by Pratt [1980b], who used Hintikka's model sets and tableaux methods to give a deterministic exponential time algorithm for deciding satisfiability/validity in PDL.

---

[57][Hoare, 1969].

A finite axiomatisation of PDL was proposed in [Segerberg, 1977], the most notable feature being the *induction* axiom

$$p \to ([\pi^*](p \to [\pi]p) \to [\pi^*]p).$$

The first proof of completeness for PDL was published by Rohit Parikh [1978a], with other proofs being attributed to Gabbay, Segerberg [1982] and Pratt.[58] The first extensive study of quantificational dynamic logic was made in David Harel's 1978 dissertation under Pratt's supervision, published as [Harel, 1979].

Many variants of dynamic logic have been studied by varying the modelling, the set of formulas, and the set of programs having associated modalities. *Deterministic* programs are modelled by requiring R_π to be a functional relation. Program predicates may be used to express computational behaviour of particular programs, such as **loop**(π), meaning that some execution of π fails to terminate, and **repeat**(π), meaning that π can be repeatedly executed infinitely many times. PDL programs can be viewed as regular sets of sequences of basic commands, but allowing *context-free* sets of sequences as programs results in a stronger logic that is Π_1^1-complete and hence highly undecidable. This was shown by Harel, Pnueli and Stavi [1983].

Dynamic *algebras* were introduced by Dexter Kozen and Pratt in 1979 and their structure and representations investigated in a number of papers.[59] They comprise a "Kleene algebra" that abstracts the algebra of regular expressions and acts as a collection of operators on a Boolean algebra. Concrete models are provided by the complex algebras of Kripke models for PDL. But the relationship between the operators interpreting π and π^* in the algebra of a Kripke model is not equationally expressible, and there are dynamic algebras that belong to the equational class generated by the algebras of Kripke models but are not themselves representable in such models.

Process logic was introduced in [Pratt, 1979] by interpreting a program, not as a relation between states, but as the set of possible state-sequences that can be generated by executing the program. In addition to "after", he proposed the following modalities

throughout π, α:	α holds at every state of any sequence generated in executing π.
during π, α:	every π-computation has α true at some point.
π *preserves* α:	in every π-computation, once α becomes true it remains so thereafter.

Parikh [1978b] developed a decidable system of second-order process logic that subsumed Pratt's, and allowed quantification over states and state-sequences. Then Nishimura [1980] combined PDL with some temporal connectives to devise a system extending Parikh's. All of these were subsumed by the powerful system of

---

[58]More background on the beginnings of dynamic logic is provided in [Goldblatt, 1986].

[59]See [Kozen and Tiuryn, 1990] for references.

process logic of Harel, Kozen and Parikh [1982] which was shown to be decidable by reduction to the second-order decidability results of [Rabin, 1969].

The article [Harel, 1984] surveys the first decade of dynamic logic, and there is a further review in [Kozen and Tiuryn, 1990].

7.2   Hennessy–Milner Logic

Matthew Hennessy and Robin Milner [1980; 1985] applied modal logic to *process algebra* in a manner that is reminiscent of the Kripke modelling of PDL. They used a modal language to express assertions about transitions between processes in such a way that two processes prove to be "observationally equivalent" just when they satisfy the same modal properties.

A *process* is viewed as an agent that interacts with its environment by performing observable actions which cause it to change its state. Processes are identified with their states, so an observation changes a process into a new process. The notation $\langle p, p' \rangle \in R_i$ means that process p can become p' by performing, or participating in, the observation i. Thus R_i is a binary relation on a given set P of processes, and we envisage a collection $\{R_i : i \in I\}$ of such observation relations corresponding to a set I of "types of observation". A particular pair $\langle p, p' \rangle \in R_i$ represents a single observation, and is also viewed as an "experiment" performed by the observer on process p. (In subsequent literature the notation $p \xrightarrow{i} p'$ became standard in place of $\langle p, p' \rangle \in R_i$.)

The Hennessy–Milner modal language has no propositional variables, but constructs formulas from the constant $\top$ by the truth-functional connectives and the modalities $\langle i \rangle$ for $i \in I$. The box modality $[i]$ is defined to be $\neg \langle i \rangle \neg$. The relation $p \models \alpha$, meaning "process p satisfies formula α", is defined inductively, with

$$p \models \langle i \rangle \alpha \quad \text{iff} \quad \text{for some } i\text{-experiment } \langle p, p' \rangle, \ p' \models \alpha.$$

Two processes are regarded as equivalent if there is no observable action that either can perform to distinguish them. Informally this means that to each observable action that one can perform there is an action that the other can perform which leads to an equivalent outcome, so each process can "simulate" the other. Spelling this out,

p is equivalent to q if, and only if,

1. for every result p' of an experiment on p, there is an equivalent result q' of an experiment on q; and

2. for every result q' of an experiment on q, there is an equivalent result p' of an experiment on p

[Milner, 1980, p. 41]. As a definition of equivalence this appears to be circular, since the word "equivalence" occurs on both sides of the "if and only if". To formalise the idea, a sequence of equivalence relations $\sim_n$ for $n \geq 0$ is defined on P. For each relation $S \subseteq P \times P$, define a relation $E(S)$ by putting $\langle p, q \rangle \in E(S)$ if for every $i \in I$,

1. $\langle p, p' \rangle \in R_i$ implies, for some q', $\langle q, q' \rangle \in R_i$ and $\langle p', q' \rangle \in S$;   and

2. $\langle q, q' \rangle \in R_i$ implies, for some p', $\langle p, p' \rangle \in R_i$ and $\langle p', q' \rangle \in S$.

Put $p \sim_0 q$ for all $p, q \in P$, and inductively $p \sim_{n+1} q$ if $\langle p, q \rangle \in E(\sim_n)$. Then p and q are defined to be *observationally equivalent*, written $p \sim q$, if $p \sim_n q$ for every n.

Now a relation $R \subseteq P \times P$ is *image-finite* if the set $\{p' : \langle p, p' \rangle \in R\}$ is finite for each $p \in P$. Hennessy and Milner gave a logical characterisation of observational equivalence by showing that if each R_i is image-finite, two processes are equivalent iff they satisfy the same formulas:

$$p \sim q \quad \text{iff} \quad \text{for all formulas } \alpha, \ p \models \alpha \text{ iff } q \models \alpha. \tag{$*$}$$

Note that the operator E on relations is *monotonic*: $R \subseteq S$ implies $E(R) \subseteq E(S)$. This property implies, by induction, that $\sim_{n+1} \subseteq \sim_n$, and so iteration of E generates a *decreasing* chain of relations

$$\sim_0 \supseteq \sim_1 \supseteq \sim_2 \supseteq \cdots \supseteq \sim_n \supseteq \cdots \cdots$$

Let $\sim_\omega = \bigcap \{\sim_n : n \geq 0\}$ be the intersection of the chain. Then in the image-finite case, $\sim_\omega$ is the *largest fixed point* of the operator E, i.e. putting $S = \sim_\omega$ gives the largest solution to the equation $S = E(S)$ (see [Hennessy and Milner, 1985, Theorem 2.1]). In that case $\langle p, q \rangle \in S$ iff $\langle p, q \rangle \in E(S)$, legitimizing the circular definition of equivalence.

The monotonicity of E alone is enough to guarantee that E has a largest fixed point (see section 7.4), but in the absence of image-finiteness this fixed point need not be the relation $\sim_\omega$. It may be a proper subrelation of $\sim_\omega$ that can only be reached by iterating E transfinitely often. Consequently this largest fixed point has become the general definition of the observational-equivalence relation $\sim$, and it is only in the image-finite case that $\sim$ is identified with $\sim_\omega$.

This analysis indicates that standard induction on natural numbers n (applied to the relations $\sim_n$) may not be effective as a method for proving equivalence of processes. Instead, as was first realised by David Park,[60] a new kind of proof rule is called for, based on the notion of a *bisimulation*. This is a relation $S \subseteq P \times P$ satisfying $S \subseteq E(S)$, i.e. $\langle p, q \rangle \in S$ implies (1) and (2) hold. The union of any collection of bisimulations is a bisimulation, and so there is a largest bisimulation— the union of all of them–which turns out to be the same as the largest fixed point of E. In other words, the observational relation $\sim$ is the largest bisimulation on any structure $(P, \{R_i : i \in I\})$. It is an equivalence relation in the mathematical sense (reflexive, symmetric and transitive) and is known as *bisimulation equivalence* or *bisimilarity* [Milner, 1989]. It

> admits an elegant proof technique; to show $p \sim q$, it is necessary and sufficient to find some bisimulation containing the pair $\langle p, q \rangle$

---

[60]Information from Robin Milner, personal communication.

[Milner, 1983, p. 283]. In the general setting, when $\sim$ is not equal to $\sim_\omega$, the same modal-logical characterisation of bisimilarity as $(*)$ above can be obtained by expanding the class of formulas to allow formation of the conjunction $\bigwedge_{j \in J} \alpha_j$ for any set $\{\alpha_j : j \in J\}$ (possibly infinite) of formulas.

The term "bisimulation" was first used in [Park, 1981] for a relation of mutual simulation between states of two automata, with motivation from an earlier notion of simulation of programs from [Milner, 1971]. Park showed that if two deterministic automata are related by a bisimulation, then they accept the same set of inputs. The concept and its use was systematically developed in [Milner, 1983]. It is closely related to the notion of "p-relation" of van Benthem [1976a] mentioned in section 5.3. Segerberg's p-morphisms are essentially bisimulations (between Kripke models) that are total and functional.

Process algebra is now a substantial field, with many concepts and constructions for building processes, and many important variations on the notion of observational equivalence or bisimilarity (see [Bergstra et al., 2001]). For any given family of transition systems, i.e. systems of observation relations, we can seek to devise modalities that generate formulas giving a logical characterisation of the bisimilarity relations for those systems in the manner of $(*)$. This programme has been carried out for many cases. Logics for more recently developed theories of "mobile" and "message-passing" processes are discussed in [Milner et al., 1993] and [Hennessy and Liu, 1995]. They provide modalities that formalise complex structural assertions, for example the formula $\langle c!x \rangle \alpha$ expressing "it is possible to output some value v on channel c and thereby evolve to a state in which $\alpha[v/x]$ is true".

Axiomatisations of various modal process logics may be found, inter alia, in [Stirling, 1987] and [Larsen, 1990]. Other work on modal aspects of process algebra is collected in [Ponse et al., 1995].

7.3 Temporal Logic for Concurrency

In 1977 Amir Pnueli, motivated by a reading of [Rescher and Urquhart, 1971],[61] proposed to use temporal logic to formalising reasoning about the behaviour of *concurrent* programs involving a number of processors acting in parallel and sharing a memory environment, so that each can alter the values of variables used by the others (see Pnueli [1977; 1981]). This is particularly relevant to the specification and analysis of *reactive* programs, like operating systems and systems for airline reservation or process control, that repeatedly interact with their environment and are not expected to terminate. As such a program runs, each success state is obtained by one processor being chosen to execute one instruction. Thus from an initial state x_0, many different sequences $x_0, x_1, \ldots$ of states may be generated depending on which processors get chosen to act at each step.

Pnueli observed that temporal modalities could be used to formulate computationally significant properties of execution sequences, such as *fair scheduling* (no processor is delayed forever), freedom from *deadlock* (when none can act), and

---

[61]See [Hasle and Øhrstrøm, 2004, p. 222].

many others. He used Prior's future-tense modality G (and its dual F), but with the Diodorean reading of "at all future states including the present", as well as a connective X with the reading "at the next state". The latter had first been introduced to tense logic for discrete time by Dana Scott (see [Prior, 1967, p. 66]). Programs do not appear in the syntax in this approach. Instead, temporal formulas describe properties of a particular execution sequence of a single (concurrent) program.

The paper of Gabbay, Pnueli, Shelah and Stavi [1980] added a binary connective U to this formalism, with $\alpha\,U\beta$ meaning "α *until* β", i.e. "β will be true, and α will be true at all times *until* β is". This connective and its past-tense version α *since* β had been studied by Hans Kamp [1968] who showed that they form an *expressively complete* set of connectives in the sense that for models in which time is a complete linear ordering, all tense-logical connectives can be defined in terms of them. Gabbay *et al.* adapted this to show that U by itself plays a similar role for the future-tense logic of state sequences. They gave an axiomatisation for this extended logic, which they called DUX, and proved that it is decidable. By way of illustration of the expressive completeness of U, they noted that $F\alpha$ can be defined as $\top\,U\alpha$, and then $G\alpha$ as $\neg F\neg\alpha$, while $\mathsf{X}\alpha$ can be defined as $\bot\,U\alpha$. DUX is now more commonly known as PLTL (propositional linear temporal logic).

Since there are many different execution sequences with a given starting state any particular sequence is just one "branch" or "path" of the "tree" of all possible future states. Considering the tree as a whole gives rise to some interesting new modalities that can formalise reasoning about future behaviour. This line was pursued by Ben-Ari, Pnueli and Manna [Ben-Ari *et al.*, 1983], defining a system $\mathcal{UB}$ (the *unified* system of *branching* time), which combined G and X with the symbols $\forall$, $\exists$ for quantification over paths to produce the following modal forms:

$$\forall G\alpha: \quad \text{along all future paths, } \alpha \text{ is true at all states.}$$
$$\exists G\alpha: \quad \text{along some path, } \alpha \text{ is true at all states.}$$
$$\forall \mathsf{X}\alpha: \quad \text{along all paths, } \alpha \text{ is true at the next state.}$$

Dual modalities were defined by writing $\exists F$ for $\neg\forall G\neg$, $\forall F$ for $\neg\exists G\neg$, and $\exists\mathsf{X}$ for $\neg\forall\mathsf{X}\neg$. The logic $\mathcal{UB}$ was shown to be finitely axiomatisable and have the finite model property, using semantic tableaux methods. It was also stated that, in contrast to PLTL, no temporal language for branching time with a finite number of modalities could be expressively complete, this theorem being credited to Gabbay.

The *until* connective U was added to $\mathcal{UB}$ by Edmund Clarke and Allen Emerson [1981] to define the system CTL of Computation Tree Logic, which was axiomatised and shown to have the finite model property by Emerson and Joseph Halpern [1982; 1985]. CTL has the limitation that the path quantifiers $\forall$ and $\exists$ are tied to a single linear-time state quantifier (modality) as in the forms $\forall G$, $\exists F$, or a single instance of U as in $\exists(\alpha\,U\beta)$ etc. It does not allow a combination like $\exists GF\alpha$, expressing "there is a path along which α is true infinitely often", a property of relevance to fair scheduling conditions. Emerson and Halpern [1983; 1986] devised a new system CTL* that allows such formations. It distinguishes

between *state formulas*, which are true or false at each state, and *path formulas*, which are true or false of each path. The path formulas include the state formulas and both categories are closed under the truth-functional connectives. If α, β are path formulas then $\alpha U \beta$, $G\alpha$ and $\mathsf{X}\alpha$ are path formulas, while $\forall\alpha$ and $\exists\alpha$ are state formulas. $\forall\alpha$ (respectively $\exists\alpha$) is true at state s iff α is true of all (respectively some) paths that start at s.

In addition to being more expressive than CTL, CTL* is more complex. Whereas CTL and PDL are decidable by algorithms that run in deterministic exponential time, the complexity of CTL* is that of deterministic *doubly* exponential time. The lower bound here was established by Moshe Vardi and Larry Stockmeyer [1985], and the upper bound by Emerson and Charanjit Jutla [1988; 1999]. Methods from tree automata theory are used to prove decidability results in this context. Models can be viewed as infinite branching trees, or at least can be "unravelled" into such tree structures. Associated with each formula α is an automaton A_α that accepts a tree model iff it it satisfies α at its root. Thus the satisfiability problem for many logics can be reduced to the *emptiness* problem for automata on infinite trees that was shown to be decidable in [Rabin, 1969] (see section 6.2). This technique was first developed in the 1980 Masters thesis of Robert Streett (see [1982]) who used it to prove the decidability of PDL with the **repeat** construct.

The logic CTL* was defined semantically, and a sound and complete axiomatisation of it was hard to find. Eventually one was provided by Mark Reynolds [2001].

A property of paths not expressible in linear time logic, or even in CTL*, is that a formula be true at every *even* state along the path (and possibly at others). Sets of sequences that have this property can be generated by formal grammars, or characterised by finite-state automata that process infinite strings. Pierre Wolper [1983] showed that any regular grammar gives rise to a temporal connective creating formulas that are true just of paths generated by that grammar in a certain way. He also showed that the linear time connectives G, F, X and U can each be expressed by such a grammar, and dubbed this formalism ETL for "*Extended Temporal Logic*". The idea can be applied to branching time systems, and leads to a logic ECTL* into which CTL* can be translated (see [Thomas, 1989]).

Surveys of computational temporal logic, and its various applications to reasoning about programs, are given in [Emerson, 1990] and [Stirling, 1992].

A different kind of use of modalities of the branching-time type was made by Glynn Winskel [1985] in constructing *powerdomains*. These structures arise in the denotational semantics of programs, and are intended to provide domain-theoretic analogues of powersets. In dynamic logic a non-deterministic program is modelled as a binary transition relation R on a set S of possible program states. Alternatively this can be viewed as a function from S to its powerset $\mathcal{P}(S)$, taking each state $x \in S$ to the set $\{y : xRy\}$ of states that can be reached by different possible executions of the program. Analogously, given a domain $\mathcal{D}$, a non-deterministic program may be modelled as a function from $\mathcal{D}$ to its powerdomain.

There are several different powerdomain constructions, and Winskel shows how

to build them out of formulas of some modal languages associated with $\mathcal{D}$. This involves tree-like models of the languages that represent certain computations. For the "Smyth" powerdomain a modality $\square$ is used that it read "inevitably". $\square\alpha$ has the same meaning in these models as the CTL-modality $\forall F\alpha$, i.e. along every future path there is a state at which α holds. The construction of the "Hoare" powerdomain uses $\diamond$, for "possibly", with $\diamond\alpha$ meaning that there is a future path with α true somewhere, i.e. $\exists F\alpha$. For the "Plotkin" powerdomain, both of these modalities are involved.

7.4   The Modal μ-Calculus

Mathematics and computer science abound with concepts and objects that are defined recursively, or self-referentially. Many of these have an elegant formulation as special *fixed points* of certain operations. The μ-calculus L_μ of Kozen [1982; 1983] admits formulas that are interpreted as fixed points, and is expressively more powerful than any of the modal program logics considered above.

Let $\Theta : \mathcal{P}(S) \to \mathcal{P}(S)$ be an operation on the powerset of a set S. Tarski applied the term "*fixpoint*" to any subset T of S such that $\Theta(T) = T$. If Θ is *monotonic* in the sense that $T \subseteq T'$ implies $\Theta(T) \subseteq \Theta(T')$, then Θ has a *least* fixpoint $\mu\Theta$ and a *greatest* fixpoint $\nu\Theta$, given by

$$\mu\Theta = \bigcap\{T \subseteq S : \Theta(T) \subseteq T\},$$
$$\nu\Theta = \bigcup\{T \subseteq S : T \subseteq \Theta(T)\}.$$

The fact that Θ has a fixpoint was first shown by Tarski and B. Knaster in 1927. In 1939 Tarski generalised this to any monotonic function on a complete lattice, showing that its fixpoints also form a complete lattice, with greatest and least elements specified by the lattice versions of the definitions just given (see [Tarski, 1955b] for this historical background).

Pratt [1981] introduced the idea of using a "minimisation" operator in a PDL-like context, but interpreted μ as a *least root* operator rather than a least fixpoint one. He developed a language of terms intended to denote elements of a Boolean algebra, with a term of the form $\mu Q.\tau(Q)$ interpreted as the least solution of the equation "$\tau(Q) = 0$". A syntactic restriction was imposed on τ to ensure that at least one solution exists. A translation of PDL into the resulting calculus was given, and the system was shown to have the finite model property by a refinement of the McKinsey method. A deterministic exponential time algorithm was given for the problem of deciding satsfiability terms.

Pratt's work provided the inspiration for Kozen's development of the calculus $L\mu$, whose language is generated from some collection Π of atomic programs (or action labels) π. $L\mu$-formulas are constructed from propositional variables using the truth-functional connectives, the modalities $[\pi]$ and $\langle\pi\rangle$ for $\pi \in \Pi$, and the constructions $\mu p.\alpha$ and $\nu p.\alpha$, where p is a propositional variable and α is a formula. The operations μp and νp function like quantifiers, binding occurrences of p in α. $\mu p.\alpha$ and $\nu p.\alpha$ are only allowed to be formed when α is *positive* in the sense that all

free occurrences of p in α are within the scope of an *even* number of negations $\neg$. This condition is satisfied for instance by any formula constructed from variables using only $\top$, $\bot$, $\wedge$, $\vee$, $[\pi]$, $\langle \pi \rangle$, μp and νp. The "binder" ν is definable in terms of μ by taking $\nu p.\alpha$ as $\neg \mu p.\neg\alpha(\neg p/p)$. Vice versa, μ could be defined in terms of ν.

An $L\mu$ model $\mathcal{M} = (S, \{\xrightarrow{\pi}: \pi \in \Pi\}, \Phi)$ is just like a Kripke model for dynamic logic, or a labelled transition system for Hennessy–Milner logic augmented by a valuation Φ to interpret the variables p. $\mathcal{M}$ gives each formula α the interpretation $\mathcal{M}(\alpha) = \{x \in S : \mathcal{M} \models_x \alpha\}$. If α contains the variable p, then varying the interpretation of p causes the interpretation of α to vary, and in this way α induces an operation on $\mathcal{P}(S)$. To make this precise, for $T \subseteq S$ let $\mathcal{M}_{p:=T}$ be the model that is identical to $\mathcal{M}$ except in interpreting p as T, i.e. $\mathcal{M}_{p:=T}(p) = T$. Then the operation induced by α on $\mathcal{P}(S)$ relative to $\mathcal{M}$ is the function

$$\Theta_\alpha^{\mathcal{M}} : T \longmapsto \mathcal{M}_{p:=T}(\alpha).$$

If α is positive, then Θ_α is monotonic. Assuming inductively that Θ_α has been specified, $\mathcal{M}(\mu p.\alpha)$ and $\mathcal{M}(\nu p.\alpha)$ are defined to be the least and greatest fixpoints $\mu\Theta_\alpha^{\mathcal{M}}$ and $\nu\Theta_\alpha^{\mathcal{M}}$ given by the Tarski–Knaster Theorem.

The meaning of $\mu p.\alpha$ and $\nu p.\alpha$ for particular α can be hard to fathom, but it helps to think of them as solutions of the equation "$p = \alpha$" and repeatedly replace p by α in α itself. It turns out that $\mu p.(\alpha \vee \langle \pi \rangle p)$ has the same interpretation in a model as the PDL-formula $\langle \pi^* \rangle \alpha$, while $\nu p.(\alpha \wedge [\pi]p)$ has the same meaning as $[\pi^*]\alpha$. Also $\mu p.\langle \pi \rangle p$ is true at x_0 iff there is an *infinite* sequence $x_0 \xrightarrow{\pi} x_1 \xrightarrow{\pi} \cdots$ in $\mathcal{M}$, which is the condition for truth of the formula **repeat**(π). Using these observations it can be shown that the logic PDL with the **repeat** construct has a simple translation into the μ-calculus.

A CTL-model can be viewed as an $L\mu$-model with a single transition relation $\xrightarrow{\pi}$, and with a *path* being a sequence $x_0 \xrightarrow{\pi} x_1 \xrightarrow{\pi} \cdots$ in the model. CTL translates into L_μ by translating $\exists(\alpha U \beta)$ as $\mu p.\beta \vee (\alpha \wedge \langle \pi \rangle p)$ and $\forall(\alpha U \beta)$ as $\mu p.\beta \vee (\alpha \wedge [\pi]p \wedge \langle \pi \rangle \top)$. The L_μ-formula $\nu p.\alpha \wedge [\pi][\pi]p$ means "along all paths, α is true at every even state", a property expressible in ECTL* but not CTL*. Mads Dam [1994] has constructed algorithms for translating both CTL* and ECTL* into L_μ.

Kozen proposed a finite axiomatisation of L_μ which, for the binder μ, has the axiom schema

$$\alpha(\mu p.\alpha/p) \to \mu p.\alpha$$

and the inference rule:

$$\text{from } \alpha(\beta/p) \to \beta \text{ infer } (\mu p.\alpha) \to \beta \text{ if } p \text{ is not free in } \beta.$$

Validity of the axiom follows from the fact that $T = \mu\Theta_\alpha^{\mathcal{M}}$ is a solution of the "inequality" $\Theta(T) \subseteq T$, and soundness of the rule is due to $\mu\Theta_\alpha^{\mathcal{M}}$ being the least such solution. Kozen was able to prove the completeness of a limited fragment of L_μ for which he also showed the finite model property and an exponential time decision procedure. The full L_μ was proved decidable by Kozen and Parikh [1984]

by reduction to Rabin's SnS. Streett and Emerson [1984; 1989] used tree automata to improve this to a deterministic triple-exponential time decision algorithm and establish the full finite model property. Emerson and Jutla [1988; 1999] sharpened the complexity result further to a deterministic exponential time algorithm, which is the best possible result since it is the lower bound for PDL and therefore for the μ-calculus. Kozen [1988] gave a different proof of the finite model property using techniques from the theory of well-quasi orders, and proved a completeness theorem for L_μ using an infinitary rule of inference.

The problem of whether L_μ is complete for Kozen's originally proposed axiomatisation proved challenging, and remained open for some time. It was eventually solved in the affirmative by Igor Walukiewicz [1995; 2000].

The formalism of the μ-calculus originates in some unpublished notes of Jaco de Bakker and Dana Scott from 1969. Kozen's inference rule derives from the Fixpoint Induction rule of [Park, 1969]. Another early independent formulation of a modal program logic with a greatest and least fixpoint operators appears in [Emerson and Clarke, 1980]. For a recent survey of the field of modal μ-calculi, see [Bradfield and Stirling, 2001].

7.5 Solovay on Provability in Arithmetic as a Modality

Let PA be the first-order system of Peano Arithmetic that is the subject of Gödel's incompleteness theorems, and let $PA \vdash \sigma$ signify that sentence σ is provable in PA. Gödel showed that this notion can be "arithmetised" and expressed in the language of PA itself. There is a PA-formula $Bew(v)$ with one free variable v such that in general $PA \vdash \sigma$ iff the sentence $Bew(\ulcorner\sigma\urcorner)$ is *true* (i.e. true of the standard PA-model $(\omega, +, \cdot, 0, 1)$). Here $\ulcorner\sigma\urcorner$ is the numeral for the Gödel number of σ. Now all PA-provable sentences are true, so for every σ the sentence

$$Bew(\ulcorner\sigma\urcorner) \rightarrow \sigma$$

is true. But it is not always *PA-provable*, a fact which is a manifestation of the first incompleteness theorem. Gödel gave an example of this in his [1933], observing that if the modality "provable" is taken to mean provable in PA then some principles of S4 do not hold:

> For example, $B(Bp \rightarrow p)$ never holds for that notion, that is it holds for no system S that contains arithmetic. For otherwise, for example, $B(0 \neq 0) \rightarrow 0 \neq 0$ and therefore also $\neg B(0 \neq 0)$ would be provable in S, that is, the consistency of S would be provable in S.

Provability in S of the consistency of S would contradict the second incompleteness theorem.

The question therefore arises as to which modal principles do hold if $\Box$ is read as "PA-provable". To make this precise, define a *realisation* to be a function ϕ assigning to each propositional variable p some PA-sentence p^ϕ. This extends

inductively to all modal formulas by taking $\top^\phi$ to be $(0 = 0)$, realising the non-modal connectives as themselves, and defining

$$(\Box\alpha)^\phi := Bew(\ulcorner\alpha^\phi\urcorner).$$

A modal formula α is *PA-valid* if $PA \vdash \alpha^\phi$ for every realisation ϕ. The question becomes that of determining which modal formulas are *PA*-valid.

The set of all *PA*-valid formulas is a normal logic, known as G (for Gödel).[62] To show that it is normal it is necessary to verify that the following hold in general:

$PA \vdash Bew(\ulcorner\sigma \to \sigma'\urcorner) \to (Bew(\ulcorner\sigma\urcorner) \to Bew(\ulcorner\sigma'\urcorner));$

If $PA \vdash \sigma$, then $PA \vdash Bew(\ulcorner\sigma\urcorner).$

These results were distilled by Martin Löb [1955] from properties of *Bew* that were established in [Hilbert and Bernays, 1939]. Löb then proved

$$PA \vdash Bew(\ulcorner\sigma\urcorner) \to Bew(\ulcorner Bew(\ulcorner\sigma\urcorner)\urcorner),$$

which shows that $\Box p \to \Box\Box p$ is *PA*-valid and hence a G-theorem. However the other S4-axiom $\Box p \to p$ is not *PA*-valid, and indeed not even the formula $\Box\bot \to \bot$ is a G-theorem, since $(\Box\bot \to \bot)^\phi$ is

$$Bew(\ulcorner 0 \neq 0\urcorner) \to 0 \neq 0,$$

which is not *PA*-provable by Gödel's reasoning above.

Robert Solovay [1976] demonstrated that G is identical to Segerberg's logic K4W, discussed in section 5.3, which is characterised by the class of finite strictly ordered (i.e. transitive and irreflexive) Kripke frames. The validity of the axiom W, i.e.

$$\Box(\Box p \to p) \to \Box p,$$

follows from an answer given in [Löb, 1955] to a question raised by Leon Henkin in 1952 about the status of sentences that assert their own provability. Any *PA*-formula $F(v)$ has *fixed points*: sentences σ for which

$$PA \vdash \sigma \leftrightarrow F(\ulcorner\sigma\urcorner)$$

(this is usually called the *Diagonalisation Lemma*). A fixed point of $Bew(v)$ has

$$PA \vdash \sigma \leftrightarrow Bew(\ulcorner\sigma\urcorner)$$

so is equivalent to the assertion of its own provability. Must it in fact be provable?[63] Löb answered this in the affirmative by proving that

$$\text{if } PA \vdash Bew(\ulcorner\sigma\urcorner) \to \sigma, \text{ then } PA \vdash \sigma.$$

---

[62] Also known as GL for Gödel–Löb.

[63] This is a generalisation of Henkin's question: see [Smoryński, 1991] for discussion.

Equivalently, if $Bew(\ulcorner Bew(\ulcorner\sigma\urcorner) \to \sigma\urcorner)$ is true then so is $Bew(\ulcorner\sigma\urcorner)$, i.e. the sentence

$$Bew(\ulcorner Bew(\ulcorner\sigma\urcorner) \to \sigma\urcorner) \to Bew(\ulcorner\sigma\urcorner)$$

is true. But more strongly it can be shown that this sentence is PA-provable for any σ, including $\sigma = \alpha^\phi$, giving the PA-validity of W.

Solovay's completeness theorem for G is a remarkable application of the machinery of arithmetisation and recursive functions to show that any finite strictly ordered frame (K, R) can be "embedded into Peano Arithmetic". A recursive function $h : \omega \to K$ is defined that is in fact constant, but which cannot be proven to be constant in PA. Each element x of K is represented by a sentence σ_x expressing "$\lim_{n\to\infty} h(n) = x$". This sentence is consistent with PA, i.e. $PA \nvdash \neg\sigma_x$. The construction has a flavour of self-referential paradox similar to that of Gödel's incompleteness proof, because the sentences σ_x are used to define the function h itself. But that is resolved by some version of diagonalisation.[64] The structure of the ordering R is represented in PA by the fact that if xRy then

$$PA \vdash \sigma_x \to \neg Bew(\ulcorner\neg\sigma_y\urcorner),$$

and if not xRy then

$$PA \vdash \sigma_x \to Bew(\ulcorner\neg\sigma_y\urcorner).$$

Any model $\mathcal{M}$ on this frame determines a realisation ϕ by putting

$$p^\phi = \bigvee\{\sigma_x : \mathcal{M} \models_x p\}.$$

Then the truth conditions in $\mathcal{M}$ are PA-representable by the fact that for any modal formula α,

if $\mathcal{M} \models_x \alpha$ then $PA \vdash \sigma_x \to \alpha^\phi$; while

if $\mathcal{M} \nvDash_x \alpha$ then $PA \vdash \sigma_x \to \neg\alpha^\phi$ and so $PA \vdash \alpha^\phi \to \neg\sigma_x$.

Since $PA \nvdash \neg\sigma_x$, the last case gives $PA \nvdash \alpha^\phi$, showing α is not PA-valid. Therefore any PA-valid formula must be true in all models on finite strictly ordered frames, and therefore be a G-theorem.

A modal formula α is called ω-valid if α^ϕ is *true* for all realisations ϕ. The set G* of all ω-valid formulas is a logic that includes G, but also includes $\Box p \to p$, since $Bew(\ulcorner\sigma\urcorner) \to \sigma$ is always true. However Gödel's example shows that $Bew(\ulcorner Bew(\ulcorner\bot^\phi\urcorner) \to \bot^\phi\urcorner)$ is not true, so G* does not contain $\Box(\Box p \to p)$, and therefore is not a normal logic. Solovay extended his analysis of G to prove that G* can be axiomatised by taking all theorems of G and instances of $\Box\alpha \to \alpha$ as axioms, and detachment as the only rule of inference.

Another natural reading of $\Box$ in this context is "true and provable", formalised by modifying the definition of realisation to

$$(\Box\alpha)^\phi := \alpha^\phi \wedge Bew(\ulcorner\alpha^\phi\urcorner).$$

---

[64]Solovay's argument used Kleene's Recursion Theorem on fixed points in the enumeration of partial recursive functions.

The fact that "provable" implies "true" might make it seem that "true and provable" has the same status as "provable", but this is not so because of the existence of true but unprovable sentences of *PA*. In general, $Bew(\ulcorner\sigma\urcorner)$ is *PA*-provable iff $\sigma \wedge Bew(\ulcorner\sigma\urcorner)$ is *PA*-provable, and the two are equivalent in the sense that

$$Bew(\ulcorner\sigma\urcorner) \leftrightarrow \sigma \wedge Bew(\ulcorner\sigma\urcorner)$$

is *true*, but this equivalence is not itself *PA-provable* unless σ is, by Löb's theorem.

The modal logic of formulas *PA*-valid under this modified realisation turns out to be the system S4Grz characterised by finite partial orderings (see section 5.3). This was proved in [Goldblatt, 1978] by showing that replacing $\Box\alpha$ by $\alpha \wedge \Box\alpha$ gives a proof-invariant translation of S4Grz into G, and then applying Solovay's theorem for G.[65] Since the intuitionistic propositional calculus IPC can be translated into S4Grz (by the result of Grzegorczyk mentioned in section 5.3), these translations can be composed to obtain a translation $\alpha \mapsto \alpha^\tau$ of propositional formulas into modal formulas such that α is provable in IPC iff α^τ is *PA*-valid. In fact α^τ is *PA*-valid iff it is ω-valid [Goldblatt, 1978, theorem 5].

Research into the modal logic of provability since the 1970s has contributed much to our understanding of the phenomena of self-reference and diagonalisation that underly the incompleteness of *PA* and other systems. An account of the origins of the subject has been given by George Boolos and Giovanni Sambin [1991], and extensive expositions are provided in the books of Boolos [1979; 1993] and Craig Smoryński [1985]. The most recent survey is that of Giorgi Japaridze and Dick de Jongh [1998].

7.6   *Grothendieck Topology as Intuitionistic Modality*

By composing his semantic analysis of S4 with the McKinsey–Tarski translation of IPC into S4, Kripke [1965a] derived a relational model theory for intuitionistic logic based on structures $\mathfrak{S} = (K, R)$ in which R is a quasi-ordering, i.e. reflexive and transitive. He interpreted the members of K informally as "evidential situations" temporally ordered by R. His paper presented a semantics for *predicate* logic, proving completeness by the method of tableaux[66]. It also showed that attention can be confined to structures that are partially ordered, i.e. antisymmetric as well. By identifying elements $x, y \in K$ whenever xRy and yRx we pass to a partially ordered quotient $\mathfrak{S}'$ which validates the same intuitionistic formulas as $\mathfrak{S}$. More strongly, any model on $\mathfrak{S}$ has an equivalent model on $\mathfrak{S}'$. This contrasts with the modal semantics on these structures: it can happen that $\mathfrak{S}'$ validates the modal axiom Grz while $\mathfrak{S}$ does not (see section 5.3).

Segerberg [1968b] studied the propositional fragment of this model theory, using only partially ordered frames from the outset. He constructed canonical models

---

[65] The result was independently found by A. Kuznetsov and A. Muzavitski (*Abstracts of Reports of the Fourth All-Union Conference on Mathematical Logic*, Kishiniev, 1976, p. 73, in Russian).

[66] An extension of intuitionistic predicate logic that is incomplete for Kripke's semantics was found by Hiroakira Ono [1973], and an incomplete extension of intuitionistic propositional logic was obtained by Valentin Šehtman [1977].

and applied the filtration method to prove the finite model property for a number of logics, including some that are weaker than or independent of IPC. The fact that IPC is characterised by the finite partially ordered frames, which also characterise S4Grz under the modal semantics, provides a clear picture of why IPC translates into S4Grz and not just S4.

Here is a brief description of the relational models for IPC. Given a partial ordering $\mathfrak{S} = (K, \leq)$, a subset X of K will be called *increasing* if it is closed "upwards" under the ordering, i.e. whenever $x \in X$ and $x \leq y$, then $y \in X$. The definition of a model $\mathcal{M} = (\mathfrak{S}, \Phi)$ requires that the set $\{x \in K : \Phi(p, x) = \top\}$ be increasing for all propositional variables p. Formally this requirement is dictated by the modal translation of p as $\Box p$, while informally it conveys the idea that once p is established as true in a given evidential situation then it remains true in the future. The truth conditions for implication and negation are

$$\begin{aligned} \mathcal{M} \models_x \alpha \to \beta \quad &\text{iff} \quad \text{for all } y \geq x, \text{ if } \mathcal{M} \models_y \alpha \text{ then } \mathcal{M} \models_y \beta, \\ \mathcal{M} \models_x \neg\alpha \quad &\text{iff} \quad \text{for all } y \geq x, \text{ not } \mathcal{M} \models_y \alpha. \end{aligned}$$

The modelling of $\wedge$ and $\vee$ is as for classical logic. By induction it is demonstrable that for each formula α the set $\mathcal{M}(\alpha) = \{x \in K : \mathcal{M} \models_x \alpha\}$ is increasing.

The topological and algebraic modellings of IPC from section 3.2 are in evidence here. The increasing sets form a topology on K, and the associated Heyting algebra of open sets satisfies a formula α iff α is valid in $\mathfrak{S}$, i.e. iff $\mathcal{M}(\alpha) = K$ for all models $\mathcal{M}$ on $\mathfrak{S}$. At the same time α is valid in $\mathfrak{S}$ iff it is satisfied by the Brouwerian algebra of *closed* subsets of this space, with the least element $\emptyset$ of the algebra being designated. This follows from properties of the set

$$\overline{\mathcal{M}}(\alpha) = \{x \in K : \text{not } \mathcal{M} \models_x \alpha\}$$

of points at which α fails to hold in model $\mathcal{M}$. $\overline{\mathcal{M}}(\alpha)$ is closed, being the complement of the open set $\mathcal{M}(\alpha)$, and takes the designated value $\emptyset$ iff α is true in the model $\mathcal{M}$. These "falsity sets" can be reconstructed by applying the Brouwerian operations that correspond to the propositional connectives:

$$\overline{\mathcal{M}}(\alpha \wedge \beta) = \overline{\mathcal{M}}(\alpha) \cup \overline{\mathcal{M}}(\beta)$$
$$\overline{\mathcal{M}}(\alpha \vee \beta) = \overline{\mathcal{M}}(\alpha) \cap \overline{\mathcal{M}}(\beta)$$
$$\overline{\mathcal{M}}(\alpha \to \beta) = \overline{\mathcal{M}}(\alpha) \div \overline{\mathcal{M}}(\beta)$$
$$\overline{\mathcal{M}}(\neg\alpha) = \overline{\mathcal{M}}(\alpha) \div K.$$

This analysis accounts for the dual nature of the Brouwerian algebraic semantics.

Modal systems based on intuitionistic logic typically take $\Box$ and $\Diamond$ as independent connectives that are not interdefinable using $\neg$. Logics of this kind, using one or both of $\Box$ and $\Diamond$, have been studied by a number of authors, for a variety of philosophical and technical motivations, beginning with a paper published by F. B. Fitch in [1948]. The history of much of this work is reviewed in the dissertation of Alex Simpson [1994, §3.3]. Here we will consider another system which has a particular mathematical significance associated with topos theory.

A topos is a category $\mathcal{E}$ that may be thought of, roughly speaking, as a model of intuitionistic higher order logic or set theory. It includes a special entity Ω, the *object of truth values*, with morphisms

$$\cap, \cup, \Rightarrow : \Omega \times \Omega \to \Omega, \quad \neg : \Omega \to \Omega \tag{3}$$

satisfying categorical formulations of the laws of Heyting algebra. A "global element" of Ω is a morphism of the form $1 \to \Omega$, where 1 is the terminal object of $\mathcal{E}$. In the category **Set** of all sets and functions 1 is a one-element set and morphisms $1 \to X$ correspond precisely to actual elements of the set X. Thus global elements of Ω in a topos are also called *truth values*. The morphisms (3) induce operations on the collection $\mathcal{E}(1, \Omega)$ of truth values that make it into a Heyting algebra, which is just the two-element Boolean algebra in the case of **Set**. But for each topological space S there exists a topos in which $\mathcal{E}(1, \Omega)$ is (isomorphic to) the Heyting algebra $\mathcal{O}(S)$ of open subsets of S.

Grothendieck generalised the notion of a topology on a set to that of a topology on a category, by generalising the notion of an open covering of a set. He used this as a basis on which to formulate sheaf theory. F. William Lawvere and Miles Tierney showed that the theory could be developed axiomatically by starting with a topos $\mathcal{E}$ having a morphism $j : \Omega \to \Omega$, called a *topology on $\mathcal{E}$*, satisfying properties that allow the construction of a certain sub-topos of "j-sheaves". The pair $(\mathcal{E}, j)$ will be called a *site*. The axioms for j are categorical versions of the requirement that an operation on a lattice be

$$
\begin{aligned}
\textit{multiplicative}: &\quad j(x \cdot y) = jx \cdot jy, \\
\textit{idempotent}: &\quad j(jx) = jx, \quad \text{and} \\
\textit{inflationary}: &\quad x \le jx.
\end{aligned}
$$

In the address at which he first announced this new theory Lawvere [1970] stated that

> A Grothendieck "topology" appears most naturally as a modal operator of the nature "it is locally the case that".

Intuitively, a property holds *locally* at a point x of a topological space if it holds at all points "near" to x, or throughout some neighbourhood of x. Alternatively, a property holds locally of an object if it is covered by open sets for each of which the property holds. For example a locally constant function is one whose domain is covered by open sets on each of which the function is constant.

Define a *local operator*[67] on a Heyting algebra $\mathfrak{H}$ to be any operation j that is multiplicative, idempotent and inflationary, and call the pair $\mathfrak{A} = (\mathfrak{H}, j)$ a *local algebra*. The general theory of these algebras has been studied by Donald Macnab [1976; 1981], who showed that local operators can be alternatively defined by the single equation

$$(x \Rightarrow jy) = (jx \Rightarrow jy).$$

---

[67] Also known in the literature as a "nucleus".

Any local algebra is a candidate for modelling a modal logic based on the intuitionistic calculus IPC. Since j is multiplicative and has $j1 = 1$, this will be a *normal* logic when $\Box$ is interpreted as j, but there has been some uncertainty as to whether a modality modelled by j is of universal or existential character. Note that a local operator has a mixture of the properties of topological interior and closure operators. It fulfills all of the axioms of an interior operator except $\mathsf{I}x \leq x$, satisfying instead the inflationary condition which is possessed by closure operators. But topological closure operators are additive ($\mathsf{C}(x+y) = \mathsf{C}x + \mathsf{C}y$), a property not required of j.

Let $\mathcal{J}$ be the set of all modal propositional formulas satisfied by all local algebras with 1 designated. The proof theory and semantics (algebraic, relational, neighbourhood, topos-theoretic) of this logic was investigated in [Goldblatt, 1981] where the symbol ∇ was used in place of $\Box$ and interpreted as a "geometric" modality. It was shown that $\mathcal{J}$ can be axiomatised by adding to the axioms and rules for IPC the three axioms

$$\nabla(p \to q) \to (\nabla p \to \nabla q)$$
$$\nabla\nabla p \to \nabla p$$
$$p \to \nabla p.$$

The last axiom allows derivation of the rule *from α infer $\nabla\alpha$*. There are a number of alternative axiomatisations of $\mathcal{J}$, one of which is to add to IPC the axioms

$$(p \to q) \to (\nabla p \to \nabla q)$$
$$\nabla\nabla p \to \nabla p$$
$$\nabla\top.$$

As Macnab's characterisation of local operators suggests, $\mathcal{J}$ can also be specified by the single axiom

$$(p \to \nabla q) \leftrightarrow (\nabla p \to \nabla q).$$

In the presence of classical Boolean logic, the middle axiom $\nabla\nabla p \to \nabla p$ in the first group is deducible from the other two, and the logic becomes the rather uninteresting system K+$(p \to \nabla p)$ whose only connected validating frames are the two one-element frames $\mathfrak{S}_\bullet$ and $\mathfrak{S}_\circ$ (see section 6.1). But in the absence of the law of excluded middle we have a modal logic with many interesting models. In particular it has relational models based on structures $\mathfrak{S} = (K, \leq, \prec)$ which refine the Kripke semantics for IPC. Here $\leq$ is a partial ordering of K and $\prec$ is a binary relation interpreting ∇ as a universal quantifier in the familar way:

$$\mathcal{M} \models_x \nabla\alpha \text{ iff } \mathcal{M} \models_y \alpha \text{ for all } y \text{ such that } x \prec y.$$

To ensure that $\mathcal{M}(\nabla\alpha)$ is $\leq$-increasing it is required that $x \leq y \prec z$ implies $x \prec z$. The logic $\mathcal{J}$ is characterised by the class of such frames in which $\prec$ is a subrelation of $\leq$ that is *dense* in the sense that $x \prec y$ implies $\exists z(x \prec z \prec y)$.

There is a canonical frame $\mathfrak{S}^{\mathcal{J}}$ of this kind that characterises $\mathcal{J}$, and the logic also has the finite model property with respect to such frames. In addition there is a characterisation of $\mathcal{J}$ by *neighbourhood* frames $(K, \leq, N)$ (see 5.3), where N_x is a filter in the lattice of $\leq$-increasing subsets of K, and the following conditions hold:

$$x \leq y \text{ implies } N_x \subseteq N_y,$$
$$\{y : x \leq y\} \in N_x,$$
$$\{y : U \in N_y\} \in N_x \text{ implies } U \in N_x.$$

If $\nabla\alpha$ is *defined* to be the formula $\neg\neg\alpha$, then the axioms of $\mathcal{J}$ become theorems of IPC. Lawvere [1970] observed that

> There is a standard Grothendieck topology on any topos, namely double negation, which is more appropriately put into words as "it is cofinally the case that".

Now if Y and Z are subsets of a partially ordered set $(K, \leq)$, then Z is *cofinal with Y* if every element of Y has an element of Z greater than it, i.e.

$$\forall y \in Y \; \exists z \in Z \; y \leq z.$$

The Kripke modelling of IPC has

$$\mathcal{M} \models_x \neg\neg\alpha \quad \text{iff} \quad \mathcal{M}(\alpha) \text{ is cofinal with } \{y : x \leq y\},$$

which explains Lawvere's interpretation of double negation as a modality. On the algebraic level, putting $j(x) = --x$ in a Heyting algebra $\mathfrak{H}$ defines a local operator whose set $\{x : --x = x\}$ of fixpoints is a *Boolean* subalgebra of $\mathfrak{H}$. On the categorical level, putting $j = \neg \circ \neg$ defines a topology on any topos $\mathcal{E}$ for which the associated subtopos $\mathcal{E}_{\neg\neg}$ of sheaves is a model of classical Boolean logic. These constructions are mathematical manifestations of the *double-negation translation* of classical propositional calculus into IPC, originating in a paper of A. N. Kolmogorov [1925], which works by inserting $\neg\neg$ in front of each subformula.

For any partially-ordered set $\mathfrak{S} = (K, \leq)$ there is a topos $\mathcal{E}_{\mathfrak{S}}$ whose objects are certain "set-valued functors" $(P, \leq) \to \mathbf{Set}$, and whose algebra $\mathcal{E}_{\mathfrak{S}}(1, \Omega)$ of truth values is isomorphic to the Heyting algebra of all increasing subsets of $\mathfrak{S}$. In the case that $\mathfrak{S}$ is an appropriate set of "forcing conditions", the topos $(\mathcal{E}_{\mathfrak{S}})_{\neg\neg}$ of "double-negation sheaves" becomes a model showing that the continuum hypothesis (for example) is independent of the axioms for topos theory including classical logic (see [Tierney, 1972]).

If $j : \Omega \to \Omega$ is a Lawvere–Tierney topology on topos $\mathcal{E}$, then the site $(\mathcal{E}, j)$ can be used to interpret modal formulas as truth values $1 \to \Omega$ in $\mathcal{E}$. The morphism j induces a local operator $f \mapsto j \circ f$ on the Heyting algebra $\mathcal{E}(1, \Omega)$ of truth values in $\mathcal{E}$. If a formula is satisfied by the resulting local algebra then it is said to be *valid in the site* $(\mathcal{E}, j)$.

The modal formulas that are valid in all sites are precisely the $\mathcal{J}$-theorems. This is shown in [Goldblatt, 1981] by the construction out of any $\mathcal{J}$-frame $\mathfrak{S} = (P, \leq, \prec)$

of a particular site $(\mathcal{E}_{\mathfrak{S}}, j_{\mathfrak{S}})$ that validates exactly the same modal formulas as does $\mathfrak{S}$. $\mathcal{E}_{\mathfrak{S}}$ is the topos of functors $(P, \leq) \to \mathbf{Set}$ as above. The relation $\prec$ is used to define $j_{\mathfrak{S}}$. Applying this construction to the canonical frame $\mathfrak{S}^{\mathcal{J}}$ produces a *canonical site* that characterises the logic $\mathcal{J}$.

It is possible to study topoi from a logical perspective, building these categories out of the syntactic and proof-theoretic machinery of formal languages of types. By including a $\mathcal{J}$-style modality in these languages the Lawvere–Tierney sheaf categories can be constructed in such a way. This approach to the theory of sheaves and topoi has been developed by John Bell [1988].

There have been several independently motivated introductions of versions of the system $\mathcal{J}$. A Gentzen-style calculus studied by Haskell Curry [1952] for proof-theoretic purposes has rules for a possibility modality $\diamond$ that gives a variant of $\mathcal{J}$ when $\diamond$ is identified with ∇. Recently the logic has re-emerged in a different guise as the Propositional Lax Logic (PLL) of Matt Fairtlough and Michael Mendler [1995; 1997]. This is a system based on intuitionistic logic that is intended to formalise reasoning about the behaviour of hardware devices, like circuits, subject to certain "constraints". A modality $\bigcirc$ is used, with $\bigcirc\alpha$ having the intuitive interpretation " for some constraint c, α holds under c". This appears to be an *existential* reading of the modality, but the authors suggest that $\bigcirc$ "has a flavour both of possibility and necessity". Their proposed axioms are

$$(p \to q) \to (\bigcirc p \to \bigcirc q)$$
$$\bigcirc \bigcirc p \to \bigcirc p$$
$$p \to \bigcirc p,$$

showing that the system is indeed a version of $\mathcal{J}$ with $\bigcirc$ in place of ∇. They give a relational semantics for PLL using structures $(K, \leq, R)$ with R being a quasi-ordered subrelation of $\leq$. The connective $\bigcirc$ is interpreted by the *universal-existential* clause

$$\mathcal{M} \models_x \bigcirc\alpha \text{ iff for all } y \geq x \text{ there exists } z \text{ such that } yRz \text{ and } \mathcal{M} \models_z \alpha.$$

It is shown that $(K, \leq, R)$ validates the same formulas as the *neighbourhood $\mathcal{J}$-frame* $(K, \leq N)$ of the above kind, where a $\leq$-increasing set U is a neighbourhood of x (i.e. $U \in N_x$) iff

$$\text{for all } y \geq x \text{ there exists } z \text{ such that } yRz \text{ and } z \in U.$$

In other words, $U \in N_x$ iff U is R-cofinal with $\{y : x \leq y\}$.

Yet another manifestation of $\mathcal{J}$ is the *CL-logic* of Nick Benton, Gavin Bierman and Valeria de Paiva [1998]. This is designed to analyse a typed lambda calculus, due to Eugenio Moggi [1991], which gives a denotational semantics for programs using a constructor T that produces a *type of computations*. The denotation of a program computing values of type A is itself an element of the type TA. The CL-logic is an intuitionistic propositional calculus corresponding to this type

system, and has a "curious possibility-like modality $\diamond$" corresponding to the type constructor T. The axioms given for $\diamond$ are

$$\diamond p \to ((p \to \diamond q) \to \diamond q)$$
$$p \to \diamond p,$$

again equivalent to the axiomatisation of $\mathscr{J}$ when $\diamond$ is identified with ∇.

Double negation constitutes just one way of combining non-modal connectives to define a modality fulfilling the $\mathscr{J}$ axioms. Other possibilities are to define $\nabla\alpha$ to be any of $\beta \vee \alpha$, $\beta \to \alpha$, or $(\beta \to \alpha) \to \alpha$, where β is some fixed (but arbitrary) formula. Peter Aczel [2001] has studied the interpretation of $\nabla\alpha$ as the second-order formula $\forall p((\alpha \to p) \to p)$, where the variable p ranges over all propositions. He calls this the "Russell–Prawitz modality" because of its relevance to certain definitions of the connectives $\wedge$, $\vee$, $\neg$, $\exists$ in terms of $\to$ and $\forall$ that were introduced by Bertrand Russell and later shown by Dag Prawitz to be derivable as equivalences in second-order intuitionistic logic.

7.7 Modal Logic for Coalgebras

The mathematics of modality has recently been applied in theoretical computer science to the category-theoretic notion of a *coalgebra*. This application is still "under construction" but can already be seen as a natural evolution of some of the trends that have been described in this article.

If $T : \mathbf{C} \to \mathbf{C}$ is a functor on a category $\mathbf{C}$, then an *algebra* for T is defined to be a pair (A, τ_A) comprising a $\mathbf{C}$-object A and a $\mathbf{C}$-arrow τ_A from TA to A. A *morphism* from T-algebra $TA \xrightarrow{\tau_A} A$ to T-algebra $TB \xrightarrow{\tau_B} B$ is a $\mathbf{C}$-arrow $A \xrightarrow{f} B$ such that $f \circ \tau_A = \tau_B \circ Tf$. This is a categorization of the classical notion of a homomorphism of abstract algebras. To explain that properly is beyond our scope, and the interested reader should consult such sources as [Mac Lane, 1971, especially §VI.8] and [Manes, 1976] for enlightenment. But the idea can be illustrated by considering the category **Malg** of (normal) modal algebras and their homomorphisms (section 6.5), which is the category of algebraic models of the smallest normal modal logic K. There is a functor $T^{\mathrm{K}} : \mathbf{Set} \to \mathbf{Set}$ on the category of sets and functions such that $T^{\mathrm{K}}A$ is the underlying set of the free modal algebra $\mathfrak{F}_A$ generated by the set A. If A is itself the underlying set of some modal algebra $\mathfrak{A}$, then there is a unique function $T^{\mathrm{K}}A \xrightarrow{\tau_{\mathfrak{A}}} A$ that is a homomorphism from $\mathfrak{F}_A$ onto $\mathfrak{A}$ leaving members of A fixed. The map $\mathfrak{A} \mapsto (A, \tau_{\mathfrak{A}})$ then gives an isomorphism between **Malg** and the category of T^{K}-algebras and their morphisms.

Note that free modal algebras can be constructed as *Lindenbaum algebras*: if a set A is viewed as a collection of propositional variables, then $T^{\mathrm{K}}A$ is the set of equivalence classes of propositional modal formulas in these variables, with formulas α and β being equivalent when $\alpha \leftrightarrow \beta$ is a K-theorem. This construction is important even when $A = \emptyset$, for there are infinitely many variable-free formulas constructible from the constants $\top$ and $\bot$ by the truth-functional connectives and

the modalities $\Box$ and $\Diamond$. The free algebra $\mathfrak{F}_\emptyset$ is an *initial object* in the category **Malg**, because for each modal algebra $\mathfrak{A}$ there a unique homomorphism from $\mathfrak{F}_\emptyset$ to $\mathfrak{A}$, since each constant formula has a uniquely determined value in $\mathfrak{A}$. The T^{K}-algebra corresponding to $\mathfrak{F}_\emptyset$ is an initial object in the category of T^{K}-algebras.

Now category theory has a principle of duality that creates a new concept out of a given one by "reversing the arrows", with the new concept being named by attaching the prefix "co" to the name of the old one. This leads to the notion of a T-*coalgebra* as an arrow of the form $A \xrightarrow{\tau_A} TA$, with a *coalgebraic morphism* from coalgebra $A \xrightarrow{\tau_A} TA$ to coalgebra $B \xrightarrow{\tau_B} TB$ being an arrow $A \xrightarrow{f} B$ such that $\tau_B \circ f = Tf \circ \tau_A$, as in

$$
\begin{array}{ccc}
A & \xrightarrow{f} & B \\
\tau_A \downarrow & & \downarrow \tau_B \\
TA & \xrightarrow{Tf} & TB
\end{array}
$$

Any modal frame can be viewed as a coalgebra for the powerset functor $\mathcal{P} : \textbf{Set} \to \textbf{Set}$. A $\mathcal{P}$-coalgebra $A \xrightarrow{\tau_A} \mathcal{P}A$ defines a binary relation R on the set A by

$$xRy \quad \text{if and only if} \quad y \in \tau_A(x),$$

giving the frame (A, R), with $\tau_A(x) = \{y : xRy\} \in \mathcal{P}A$. But this last equation can also be read as a *definition* of τ_A given R, so there is an exact correspondence between frames and $\mathcal{P}$-coalgebras. Moreover, a function $f : A \to B$ is a coalgebraic morphism from $A \xrightarrow{\tau_A} \mathcal{P}A$ to coalgebra $B \xrightarrow{\tau_B} \mathcal{P}B$ precisely when it is a p-morphism (section 5.3) between the corresponding frames.

Refining this analysis shows that models on frames can be identified with coalgebras for a functor T^Π on **Set** that has $T^\Pi A = \mathcal{P}A \times \mathcal{P}\Pi$, where Π is the set of propositional variables. A model $\mathcal{M} = (A, R, \Phi)$ corresponds to the coalgebra $A \xrightarrow{\tau_\mathcal{M}} \mathcal{P}A \times \mathcal{P}\Pi$ having

$$\tau_\mathcal{M}(x) = \big\langle \{y : xRy\}, \{p : \Phi(p, x) = \top\} \big\rangle.$$

Similar coalgebraic presentations can be given for a range of structures that arise in the theory of computation. These include state-based systems from automata theory and process algebra; various data structures like lists, trees and streams; and classes in object-oriented programming languages. Many such examples can be found in the papers of [Reichel, 1995; Jacobs, 1996; Jacobs and Rutten, 1997; Rutten, 1995; 2000; Jacobs, 2002]. Here we illustrate with the case of a collection $\{R_i : i \in I\}$ of observation relations associated with the Hennessy–Milner logic described in section 7.2. This can be viewed as a coalgebra for a functor $(\mathcal{P}-)^I$ that takes each set A to the set $(\mathcal{P}A)^I$ of all functions from I to $\mathcal{P}A$. Using the notation $x \xrightarrow{i} y$ in place of $\langle x, y \rangle \in R_i$, we find that a system $\{R_i : i \in I\}$ of relations on a set A corresponds to the coalgebra $A \xrightarrow{\tau_I} (\mathcal{P}A)^I$ for which $\tau_I(x)$ is the function $i \mapsto \{y : x \xrightarrow{i} y\}$. A coalgebra for $(\mathcal{P}-)^I$ can also be regarded as providing the state-transition relation for a non-deterministic automaton with

input set I and state set A. For each state x in A, $\tau_I(x)(i)$ is the set of possible next states that can be reached by making a transition from x on input i. For this reason, the τ-arrow of any kind of coalgebra is often called a *transition structure*, and its domain is thought of as a *state set*. (We can identify (A, τ_A) with its transition structure, since A is determined as the domain of τ_A.)

Examples such as these have spurred the establishment of a general theory of **Set**-based coalgebras, by analogy with the classical theory of universal algebras, This "universal coalgebra" was initiated and developed extensively by Jan Rutten [1996; 2000]. Another valuable source of material is the lecture notes of Peter Gumm [1999]. The theory makes significant use of a definition of *bisimulation* for coalgebras that was introduced in [Aczel and Mendler, 1989]. A relation $R \subseteq A \times B$ is a bisimulation from $A \xrightarrow{\tau_A} TA$ to $B \xrightarrow{\tau_B} TB$ when there exists a transition structure $R \xrightarrow{\tau_R} TR$ on R such that the projection functions from (R, τ_R) to (A, τ_A) and (B, τ_B) are coalgebraic morphisms. There is always a largest such bisimulation $\sim_{AB}$, known as the *bisimilarity* relation from (A, τ_A) to (B, τ_B). This abstracts the relation of observational equivalence of processes discussed in section 7.2.

Another fundamental notion is that of a *final*, or *terminal*, coalgebra, categorically dual to the notion of initial algebra discussed for modal algebras above. A T-coalgebra (F, τ_F) is called final if for each T-coalgebra (A, τ_A) there is a unique coalgebraic morphism $(A, \tau_A) \xrightarrow{f_A} (F, \tau_F)$. In the process algebra context the states of a final coalgebra are thought of as representing all possible "observable behaviours" of processes, because observationally equivalent processes are identified by the unique morphism to a final coalgebra. More precisely, for any states x and y of coalgebra (A, τ_A), if $x \sim y$ then $f_A(x) = f_A(y)$, and the converse is also true under a mild restriction on T [Rutten and Turi, 1993, Corollary 2.9].

It is a well known observation of Joachim Lambek that the transition structure τ_F of a final T-coalgebra is an isomorphism between F and TF. So it follows from Cantor's Theorem that there cannot exist any final $\mathcal{P}$-coalgebra, since there is no bijection from any set A onto its powerset $\mathcal{P}A$. Thus the category of modal frames and p-morphisms has no final object. More generally there is no final coalgebra for the functor $(\mathcal{P}-)^I$ whose coalgebras are non-deterministic transition systems with input set I. On the other hand, we can model *finitely branching* non-determinism by using the finitary powerset functor $\mathcal{P}_\omega$, where $\mathcal{P}_\omega A$ is the set of all finite subsets of A. A $(\mathcal{P}_\omega -)^I$-coalgebra is an *image-finite* transition system in the sense, described in section 7.2, that the set $\{y : x \xrightarrow{i} y\}$ of possible next states is finite for each state x and each input i. There does exist a final $(\mathcal{P}_\omega -)^I$-coalgebra: this follows from general results about the existence of final coalgebras [Aczel and Mendler, 1989; Barr, 1993; Kawahara and Mori, 2000; Rutten, 2000]. In particular, a final T-coalgebra exists whenever T is *bounded*, which means that there is some cardinal number κ such that any state of a T-coalgebra belongs to some subcoalgebra with no more than κ states. The functor $\mathcal{P}_\omega$ is bounded with $\kappa = \aleph_0$, and for each set I, $(\mathcal{P}_\omega -)^I$ is bounded with $\kappa = \max\{\aleph_0, \operatorname{card} I\}$.

Devising a suitable syntax and semantics for T-coalgebras is a matter that depends on the nature of the functor T involved. A natural desideratum is a satisfaction relation $\tau_A, x \models \alpha$, expressing "formula α is true/satisfied at state x in coalgebra τ_A", that provides a logical characterisation of bisimilarity in the following form:

$$x \sim_{AB} y \quad \text{iff} \quad \text{for all formulas } \alpha, \ \tau_A, x \models \alpha \ \text{iff} \ \tau_B, y \models \alpha.$$

If this holds we will say that the logic, or the functor T, has the *Hennessy–Milner (HM) property* (see ($*$) in section 7.2).

The first explicit coalgebraic logic with this property was introduced by Lawrence Moss [1999] for a broad class of functors that have final coalgebras. The language involved was *infinitary*, allowing formation of the conjunction of any set of formulas. For certain functors it was shown that this language has sufficient expressive power to characterise each state of the final coalgebra uniquely by a single formula.

Finitary modal languages with the HM-property were developed by Alexander Kurz [1998; 2001], Martin Rößiger [1998; 2001] and Bart Jacobs [2000] for coalgebras of *polynomial* functors. A functor is polynomial if it can be inductively constructed from the identity functor $A \mapsto A$ and functors $A \mapsto C$ with some constant value C, by forming products $A \mapsto T_1 A \times T_2 A$, disjoint unions $A \mapsto T_1 A + T_2 A$, and "exponential" functors $A \mapsto (TA)^I$ with fixed exponent I. The value C of a constant functor can be thought of as a set of "outputs" or "observable values" and an exponent I as an "input" set. For example, consider the functor having $TA = (C \times A)^I$ with fixed sets C and I. The corresponding modal language has a modality $[i]$ for each $i \in I$. Given a state x in a T-coalgebra (A, τ_A), and an "input" $i \in I$, we obtain a pair $\tau_A(x)(i) \in C \times A$ whose second projection $\pi_2(\tau_A(x)(i))$ is a new state from A. We declare a modal formula $[i]\alpha$ to be true at x when α is true at this next state:

$$\tau_A, x \models [i]\alpha \quad \text{iff} \quad \tau_A, \pi_2(\tau_A(x)(i)) \models \alpha.$$

Note that the first projection $\pi_1(\tau_A(x)(i))$ here is an output value from C. The language for T-coalgebras in this case has formulas $(i)c$ for each $c \in C$ with the semantics

$$\tau_A, x \models (i)c \quad \text{iff} \quad \pi_1(\tau_A(x)(i)) = c.$$

Similarly, the logic for a general polynomial functor T has modal formulas $[p]\alpha$ and "observational" formulas $(p)c$ built from certain *path expressions* p that syntactically reflect the internal structure and inductive formation of T. The Lemmon–Scott canonical model construction (section 5.1) can be adapted to such logics, and Kurz and Rößiger proved that the canonical model is a final T-coalgebra in the case that the constant sets C occurring in the definition of T are all finite. Jacobs showed that under this same restriction a contravariant duality of the kind considered in section 6.5 can be constructed between the category of T-coalgebras and a certain category of Boolean algebras with operators corresponding to the path-modalities $[p]$.

Another approach to polynomial coalgebraic logic was introduced in [Goldblatt, 2001b; 2003b] by working with terms for algebraic expressions, like $\pi_1(\tau_A(x)(i))$, that have a single state-valued variable x. Boolean combinations of equations between observable-valued terms were shown to give a class of formulas that has the Hennessy–Milner property. Bisimilar states were also characterised as those that assign the same values to all observable-valued terms. Equations with the same semantics as the above formulas $[p]\alpha$ and $(p)c$ can be defined in this language.

Of course the idea of a formula or term having a single state-valued variable is an implicitly modal one, and goes all the way back to Meredith's U-calculus interpretation of propositional modal formulas as formulas of first-order logic that have a single free variable (Sections 4.4 and 6.3). At the same time this equational approach is closer to classical universal algebra and model theory, and leads to natural coalgebraic constructions of ultraproducts [Goldblatt, 2003d] and ultrafilter extensions [Goldblatt, 2003a].

Coalgebras for polynomial functors can be thought of as generalised deterministic automata. Non-determinism can also be accommodated by using the powerset functor $\mathcal{P}$ along with the polynomial operations to form the so-called *Kripke polynomial functors* of [Rößiger, 2000]. There are finitary modal logics for these as well, but the HM-property now only holds for coalgebras that are imagine-finite, which essentially means that the finitary powerset functor $\mathcal{P}_\omega$ is used in place of $\mathcal{P}$ in their construction.

The original modal language and semantics of Hennessy and Milner (section 7.2) provides any functor of the form $(\mathcal{P}_\omega-)^I$ with a finitary logic having the HM-property. Its syntax can be extended by allowing formation of conjunctions of sets of fewer than κ formulas, for some fixed infinite cardinal number κ. The result is a logic with the HM-property for the functor $(\mathcal{P}_\kappa-)^I$, where $\mathcal{P}_\kappa A$ is the set of all subsets of A with fewer than κ elements. $(\mathcal{P}_\kappa-)^I$ is bounded and has a final coalgebra, for any infinite κ. By going further and forming conjunctions of *arbitrary* sets of formulas [Milner, 1989], an HM-logic is obtained for the functor $(\mathcal{P}-)^I$. But now the collection of formulas becomes a proper class, rather than a set. Also, there is no longer any final coalgebra. These two facts are connected: it can be shown [Goldblatt, 2004] that if a functor T has an HM-logic whose class of formulas is *small* (i.e. a set), then there must be a final T-coalgebra. Consequently, there is no such small HM-logic for a functor of the form $(\mathcal{P}-)^I$.

The formulation and analysis of logics for various categories of coalgebras is the subject of current research. The assessment of the impact of these investigations on the evolution of modal logic is a task for the historians of the future.

BIBLIOGRAPHY

[Aczel and Mendler, 1989] Peter Aczel and Nax Mendler. A final coalgebra theorem. In D. H. Pitt *et al.*, editors, *Category Theory and Computer Science. Proceedings 1989*, volume 389 of *Lecture Notes in Computer Science*, pages 357–365. Springer-Verlag, 1989.

[Aczel, 1988] Peter Aczel. *Non-Well-Founded Sets*. CSLI Lecture Notes No. 14. CSLI Publications, Stanford University, 1988.

[Aczel, 2001] Peter Aczel. The Russell-Prawitz modality. *Mathematical Structures in Computer Science*, 11(4):541–554, 2001.

[Astroh and Klüwer, 1998] Michael Astroh and Johan W. Klüwer. Bibliography: Hugh MacColl. *Nordic Journal of Philosophical Logic*, 3(1), 1998. Electronically available at: `www.hf.uio.no/filosofi/njpl/vol3no1/index.html`.

[Barr, 1993] Michael Barr. Terminal coalgebras in well-founded set theory. *Theoretical Computer Science*, 114:299–315, 1993.

[Barwise, 1989] Jon Barwise. *The Situation in Logic*. CSLI Lecture Notes No. 17. CSLI Publications, Stanford University, 1989.

[Bayart, 1958] A. Bayart. Correction de la logique modale du premier et du second ordre S5. *Logique et Analyse*, 1(1):28–45, 1958.

[Bayart, 1959] A. Bayart. Quasi-adéquation de la logique modal du second ordre S5 et adéquation de la logique modal du premier ordre S5. *Logique et Analyse*, 2(6–7):99–121, 1959.

[Becker, 1930] Oskar Becker. Zur logik der modalitäten. *Jahrbuch für Philosophie und Phänomenologische Forschung*, 11:497–548, 1930.

[Bell, 1988] J. L. Bell. *Toposes and Local Set Theories*. Oxford University Press, 1988.

[Ben-Ari *et al.*, 1983] Mordechai Ben-Ari, Amir Pnueli, and Zohar Manna. The temporal logic of branching time. *Acta Informatica*, 20:207–226, 1983.

[Benton *et al.*, 1998] P. N. Benton, G. M. Bierman, and V. C. V. de Paiva. Computational types from a logical perspective. *Journal of Functional Programming*, 8:177–193, 1998.

[Bergstra *et al.*, 2001] J. A. Bergstra, A. Ponse, and S. A. Smolka. *Handbook of Process Algebra*. Elsevier, 2001.

[Beth, 1955] E. W. Beth. Semantic entailment and formal derivability. *Mededelingen van de Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde, N. R.*, 18:309–342, 1955. Reprinted in *The Philosophy of Mathematics*, edited by Jaakko Hintikka, Oxford University Press, 1969.

[Birkhoff, 1933] Garrett Birkhoff. On the combination of subalgebras. *Proceedings of the Cambridge Philosophical Society*, 29:441–464, 1933.

[Birkhoff, 1935] Garrett Birkhoff. On the structure of abstract algebras. *Proceedings of the Cambridge Philosophical Society*, 31:433–454, 1935.

[Birkhoff, 1944] Garrett Birkhoff. Subdirect unions in universal algebra. *Bulletin of the American Mathematical Society*, 50:764–768, 1944.

[Birkhoff, 1948] Garrett Birkhoff. *Lattice Theory*. American Mathematical Society, second edition, 1948.

[Blok, 1978a] W. J. Blok. On the degree of incompleteness of modal logics. *Bulletin of the Section of Logic of the Polish Academy of Sciences*, 7(4):167–175, December 1978.

[Blok, 1978b] W. J. Blok. On the degree of incompleteness of modal logics and the covering relation in the lattice of modal logics. Report 78-07, Department of Mathematics, University of Amsterdam, July 1978.

[Blok, 1980] W. J. Blok. The lattice of modal logics: An algebraic investigation. *The Journal of Symbolic Logic*, 45:221–236, 1980.

[Bochenski, 1961] I. M. Bochenski. *A History of Formal Logic*. University of Notre Dame Press, 1961.

[Boole, 1854] George Boole. *An Investigation of the Laws of Thought*. Walton and Maberley, 1854. Reprinted by Dover Publications 1958.

[Boolos and Sambin, 1985] George Boolos and Giovanni Sambin. An incomplete system of modal logic. *Journal of Philosophical Logic*, 14:351–358, 1985.

[Boolos and Sambin, 1991] George Boolos and Giovanni Sambin. Provability: The emergence of a mathematical modality. *Studia Logica*, 50:1–23, 1991.

[Boolos, 1979] George Boolos. *The Unprovability of Consistency*. Cambridge University Press, 1979.

[Boolos, 1993] George Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

[Bradfield and Stirling, 2001] Julian Bradfield and Colin Stirling. Modal logics and mu-calculi: An introduction. In J. A. Bergstra, A. Ponse, and S. A. Smolka, editors, *Handbook of Process Algebra*, pages 293–330. Elsevier, 2001.

[Bull, 1964] R. A. Bull. A note on the modal calculi S4.2 and S4.3. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 10:53–55, 1964.

[Bull, 1965] R. A. Bull. An algebraic study of Diodorean modal systems. *The Journal of Symbolic Logic*, 30:58–64, 1965.

[Bull, 1966] R. A. Bull. That all normal extensions of S4.3 have the finite model property. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:341–344, 1966.

[Bull, 1968] R. A. Bull. An algebraic study of tense logics with linear time. *The Journal of Symbolic Logic*, 33:27–38, 1968.

[Bull, 1969] R. A. Bull. Note on a paper in tense logic. *The Journal of Symbolic Logic*, 34:215–218, 1969.

[Carnap, 1946] Rudolph Carnap. Modalities and quantification. *The Journal of Symbolic Logic*, 11:33–64, 1946.

[Carnap, 1947] Rudolph Carnap. *Meaning and Necessity*. The University of Chicago Press, 1947. Enlarged Edition 1956.

[Chagrov and Zakharyaschev, 1997] Alexander Chagrov and Michael Zakharyaschev. *Modal Logic*. Oxford University Press, 1997.

[Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.

[Clarke and Emerson, 1981] Edmund M. Clarke and E. Allen Emerson. Design and synthesis of synchronisation skeletons using branching time temporal logic. In D. Kozen, editor, *Logics of Programs*, volume 131 of *Lecture Notes in Computer Science*, pages 52–71. Springer-Verlag, 1981.

[Cook, 1971] S. A. Cook. The complexity of theorem proving procedures. In *Proceedings of the Third Annual ACM Symposium on the Theory of Computing*, pages 151–158, 1971.

[Copeland, 1996a] B. J. Copeland, editor. *Logic and Reality: Essays on the Legacy of Arthur Prior*. Oxford University Press, 1996.

[Copeland, 1996b] B. J. Copeland. Prior's life and legacy. In B. J. Copeland, editor, *Logic and Reality: Essays on the Legacy of Arthur Prior*, pages 1–40. Oxford University Press, 1996.

[Copeland, 2002] B. J. Copeland. The genesis of possible worlds semantics. *Journal of Philosophical Logic*, 31(2):99–137, 2002.

[Cresswell, 1967] M. J. Cresswell. A Henkin completeness theorem for T. *Notre Dame Journal of Formal Logic*, 8:186–190, 1967.

[Cresswell, 1972] M. J. Cresswell. The completeness of S1 and some related systems. *Notre Dame Journal of Formal Logic*, 13:485–496, 1972.

[Cresswell, 1973] M. J. Cresswell. *Logics and Languages*. Methuen, 1973.

[Cresswell, 1984] M. J. Cresswell. An incomplete decidable modal logic. *The Journal of Symbolic Logic*, 49:520–527, 1984.

[Cresswell, 1995] M. J. Cresswell. S1 is not so simple. In Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher, editors, *Modality, Morality and Belief. Essays in Honor of Ruth Barcan Marcus*, pages 29–40. Cambridge University Press, 1995.

[Curry, 1950] Haskell B. Curry. *A Theory of Formal Deducibility*. Notre Dame Mathematical Lectures, no. 6. University of Notre Dame, 1950.

[Curry, 1952] Haskell B. Curry. The elimination theorem when modality is present. *The Journal of Symbolic Logic*, 17:249–265, 1952.

[Dam, 1994] Mads Dam. CTL* and ECTL* as fragments of the modal μ-calculus. *Theoretical Computer Science*, 126:77–96, 1994.

[Davis, 1954] Chandler Davis. Modal operators, equivalence relations, and projective algebras. *American Journal of Mathematics*, 76:747–762, 1954.

[de Jongh and Troelstra, 1966] D. H. J. de Jongh and A. S. Troelstra. On the connection of partially ordered sets with some pseudo-Boolean algebras. *Indagationes Mathematicae*, 28:317–329, 1966.

[de Jongh and van Ulsen, 1998–1999] Dick de Jongh and Paul van Ulsen. Beth's nonclassical valuations. *Philosophia Scientiae*, 3(4):279–302, 1998–1999.

[de Rijke and Venema, 1995] Maarten de Rijke and Yde Venema. Sahlqvist's theorem for Boolean algebras with operators with an application to cylindric algebras. *Studia Logica*, 54:61–78, 1995.

[Došen, 1989] Kosta Došen. Duality between modal algebras and neighbourhood frames. *Studia Logica*, 48:219–234, 1989.

[Dugundji, 1940] James Dugundji. Note on a property of matrices for Lewis and Langford's calculi of propositions. *The Journal of Symbolic Logic*, 5:150–151, 1940.

[Dummett and Lemmon, 1959] M. A. E. Dummett and E. J. Lemmon. Modal logics between S4 and S5. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 5:250–264, 1959.

[Emerson and Clarke, 1980] E. Allen Emerson and Edmund M. Clarke. Characterizing correctness properties of parallel programs using fixpoints. In J. W. de Bakker and J. van Leeuwen, editors, *Automata, Languages and Programming. Proceedings 1980*, volume 85 of *Lecture Notes in Computer Science*, pages 169–181. Springer-Verlag, 1980.

[Emerson and Halpern, 1982] E. Allen Emerson and Joseph Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. In *Annual ACM Symposium on Theory of Computing (STOC)*, pages 169–180, 1982.

[Emerson and Halpern, 1983] E. Allen Emerson and Joseph Y. Halpern. "Sometimes" and "not never" revisited: On branching versus linear time. In *Proceedings of the Annual ACM Symposium on Principles of Programming Languages*, pages 127–140, 1983.

[Emerson and Halpern, 1985] E. Allen Emerson and Joseph Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. *Journal of Computer and Systems Sciences*, 30:1–24, 1985.

[Emerson and Halpern, 1986] E. Allen Emerson and Joseph Y. Halpern. "Sometimes" and "not never" revisited: On branching versus linear time temporal logic. *Journal of the Association for Computing Machinery*, 33:151–178, 1986.

[Emerson and Jutla, 1988] E. Allen Emerson and Charanjit S. Jutla. The complexity of tree automata and logics of programs. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 328–337, 1988.

[Emerson and Jutla, 1999] E. Allen Emerson and Charanjit S. Jutla. The complexity of tree automata and logics of programs. *SIAM Journal on Computing*, 29:132–158, 1999.

[Emerson, 1990] E. Allen Emerson. Temporal and modal logic. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume B: Formal Models and Semantics*, pages 995–1072. Elsevier, 1990.

[Erdős, 1959] Paul Erdős. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[Èsakia, 1974] L. L. Èsakia. Topological Kripke models. *Soviet Mathematics Doklady*, 15:147–151, 1974.

[Everett and Ulam, 1946] C. J. Everett and S. Ulam. Projective algebras I. *American Journal of Mathematics*, 68:77–88, 1946.

[Fagin et al., 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. The MIT Press, 1995.

[Fairtlough and Mendler, 1995] Matt Fairtlough and Michael Mendler. An intuitionistic modal logic with applications to the formal verification of hardware. In Leszek Pacholoski and Jerzy Tiuryn, editors, *Computer Science Logic. Proceedings 1994*, volume 933 of *Lecture Notes in Computer Science*, pages 354–368. Springer-Verlag, 1995.

[Fairtlough and Mendler, 1997] Matt Fairtlough and Michael Mendler. Propositional lax logic. *Information and Computation*, 137:1–33, 1997.

[Feys, 1965] Robert Feys. *Modal Logics*. E. Nauwelaerts, 1965.

[Fine, 1971] Kit Fine. The logics containing S4.3. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 17:371–376, 1971.

[Fine, 1974] Kit Fine. An incomplete logic containing S4. *Theoria*, 40:23–29, 1974.

[Fine, 1975a] Kit Fine. Normal forms in modal logic. *Notre Dame Journal of Formal Logic*, 16:229–234, 1975.

[Fine, 1975b] Kit Fine. Some connections between elementary and modal logic. In Stig Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 15–31. North-Holland, Amsterdam, 1975.

[Fisher and Ladner, 1977] Michael J. Fisher and Richard E. Ladner. Propositional modal logic of programs. In *Proceedings of the Ninth Annual ACM Symposium on The Theory of Computing*, pages 286–294, 1977.

[Fischer and Ladner, 1979] Michael J. Fischer and Richard E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and Systems Sciences*, 18:194–211, 1979.

[Fitch, 1948] F. B. Fitch. Intuitionistic modal logic with quantifiers. *Portugaliae Mathematicae*, 7:113–118, 1948.

[Føllesdal, 1994] Dagfinn Føllesdal. Stig Kanger in memoriam. In Dag Prawitz, Brian Skyrms, and Dag Westerståhl, editors, *Logic, Methodology and Philosophy of Science IX*, pages 885–888. North-Holland, Amsterdam, 1994.

[Gabbay *et al.*, 1980] D. M. Gabbay, A. Pneuli, S. Selah, and J. Stavi. On the temporal analysis of fairness. In *Proceedings of the 7th Annual ACM Symposium on Principles of Programming Languages*, pages 163–173, 1980.

[Gabbay *et al.*, 1994] Dov M. Gabbay, Ian Hodkinson, and Mark Reynolds. *Temporal Logic. Mathematical Foundations and Computational Aspects, volume 1.* Oxford University Press, 1994.

[Gabbay, 1972] Dov M. Gabbay. On decidable, finitely axiomatizable, modal and tense logics without the finite model property, part I. *Israel Journal of Mathematics*, 11:478–495, 1972.

[Gabbay, 1976] Dov M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. D. Reidel, 1976.

[Gerson, 1975a] Martin Gerson. An extension of S4 complete for the neighborhood semantics but incomplete for the relational semantics. *Studia Logica*, 34:333–342, 1975.

[Gerson, 1975b] Martin Gerson. The inadequacy of the neighborhood semantics for modal logic. *The Journal of Symbolic Logic*, 40:141–148, 1975.

[Gerson, 1976] Martin Gerson. A neighborhood frame for T with no equivalent relational frame. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 22:29–34, 1976.

[Givant and Venema, 1999] S. Givant and Y. Venema. The preservation of Sahlqvist equations in completions of Boolean algebras with operators. *Algebra Universalis*, 41:47–84, 1999.

[Gödel, 1931] Kurt Gödel. Review of *Becker 1930*. *Monatshefte für Mathematik und Physik (Literaturberichte)*, 38:5, 1931. English translation in Solomon Feferman *et al.*, editors, *Kurt Gödel, Collected Works, Volume 1*, p. 217. Oxford University Press, 1986.

[Gödel, 1933] Kurt Gödel. Eine interpretation des intuitionistischen aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933. English translation, with an introductory note by A. S. Troelstra, in Solomon Feferman *et al.*,editor, *Kurt Gödel, Collected Works, Volume 1*, pages 296–303. Oxford University Press, 1986.

[Goldblatt and Thomason, 1975] R. I. Goldblatt and S. K. Thomason. Axiomatic classes in propositional modal logic. In J. N. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 163–173. Springer-Verlag, 1975.

[Goldblatt, *et al.*, 2003] Robert Goldblatt, Ian Hodkinson, and Yde Venema. On canonical modal logics that are not elementarily determined. *Logique et Analyse*, 181:???–???, 2003.

[Goldblatt *et al.*, 2004] Robert Goldblatt, Ian Hodkinson, and Yde Venema. Erdős graphs resolve Fine's canonicity problem. *The Bulletin of Symbolic Logic*, 10(2):186–208, 2004.

[Goldblatt, 1974] Robert Goldblatt. *Metamathematics of Modal Logic*. PhD thesis, Victoria University, Wellington, February 1974. Included in [Goldblatt, 1993].

[Goldblatt, 1975a] Robert Goldblatt. First-order definability in modal logic. *The Journal of Symbolic Logic*, 40(1):35–40, 1975.

[Goldblatt, 1975b] Robert Goldblatt. Solution to a completeness problem of Lemmon and Scott. *Notre Dame Journal of Formal Logic*, 16:405–408, 1975.

[Goldblatt, 1978] Rob Goldblatt. Arithmetical necessity, provability, and intuitionistic logic. *Theoria*, 44:38–46, 1978. Reprinted in [Goldblatt, 1993].

[Goldblatt, 1980] Robert Goldblatt. Diodorean modality in Minkowski spacetime. *Studia Logica*, 39:219–236, 1980. Reprinted in [Goldblatt, 1993].

[Goldblatt, 1981] Robert Goldblatt. Grothendieck topology as geometric modality. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 27:495–529, 1981. Reprinted in [Goldblatt, 1993].

[Goldblatt, 1986] Robert Goldblatt. Review of initial papers on dynamic logic by Pratt, Fischer and Ladner, Segerberg, Parikh and Kozen. *The Journal of Symbolic Logic*, 51:225–227, 1986.

[Goldblatt, 1989] Robert Goldblatt. Varieties of complex algebras. *Annals of Pure and Applied Logic*, 44:173–242, 1989.

[Goldblatt, 1991a] Robert Goldblatt. The McKinsey axiom is not canonical. *The Journal of Symbolic Logic*, 56:554–562, 1991.

[Goldblatt, 1991b] Robert Goldblatt. On closure under canonical embedding algebras. In H. Andréka, J.D. Monk, and I. Németi, editors, *Algebraic Logic*, volume 54 of *Colloquia Mathematica Societatis János Bolyai*, pages 217–229. North-Holland, Amsterdam, 1991.

[Goldblatt, 1993] Robert Goldblatt. *Mathematics of Modality*. CSLI Lecture Notes No. 43. CSLI Publications, Stanford, CA, 1993. Distributed by Chicago University Press.

[Goldblatt, 1995] Robert Goldblatt. Elementary generation and canonicity for varieties of Boolean algebras with operators. *Algebra Universalis*, 34:551–607, 1995.

[Goldblatt, 1999] Robert Goldblatt. Reflections on a proof of elementarity. In Jelle Gerbrandy, Maarten Marx, Maarten de Rijke, and Yde Venema, editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, Vossiuspers. Amsterdam University Press, 1999. ISBN 90 5629 104 1.

[Goldblatt, 2000] Robert Goldblatt. Algebraic polymodal logic: A survey. *Logic Journal of the IGPL*, Special Issue on Algebraic Logic edited by István Németi and Ildikó Sain, 8(4):393–450, July 2000. Electronically available at: `www3.oup.co.uk/igpl`.

[Goldblatt, 2001a] Robert Goldblatt. Quasi-modal equivalence of canonical structures. *The Journal of Symbolic Logic*, 66:497–508, 2001.

[Goldblatt, 2001b] Robert Goldblatt. What is the coalgebraic analogue of Birkhoff's variety theorem? *Theoretical Computer Science*, 266:853–886, 2001.

[Goldblatt, 2003a] Robert Goldblatt. Enlargements of polynomial coalgebras. In Rod Downey et al., editor, *Proceedings of the 7th and 8th Asian Logic Conferences*, pages 152–192. World Scientific, 2003.

[Goldblatt, 2003b] Robert Goldblatt. Equational logic of polynomial coalgebras. In Philippe Balbiani, Nobu-Yuki Suzuki, Frank Wolter, and Michael Zakharyaschev, editors, *Advances in Modal Logic, Volume 4*, pages 149–184. King's College Publications, King's College London, 2003. `www.aiml.net`.

[Goldblatt, 2003c] Robert Goldblatt. Mathematical modal logic: A view of its evolution. *Journal of Applied Logic*, 1(5–6):309–392, 2003.

[Goldblatt, 2003d] Robert Goldblatt. Observational ultraproducts of polynomial coalgebras. *Annals of Pure and Applied Logic*, 123:235–290, 2003.

[Goldblatt, 2004] Robert Goldblatt. Final coalgebras and the Hennessy-Milner property. *Annals of Pure and Applied Logic*, 2004. To appear.

[Grzegorczyk, 1967] Andrzej Grzegorczyk. Some relational systems and the associated topological spaces. *Fundamenta Mathematicae*, 60:223–231, 1967.

[Gumm, 1999] H. Peter Gumm. Elements of the general theory of coalgebras. LUATCS'99, Rand Africaans University, Johannesburg, South Africa, 60 pp. `www.Mathematik.uni-marburg.de/~gumm/Papers/publ.html`, 1999.

[Halmos, 1962] P. R. Halmos. *Algebraic Logic*. Chelsea, New York, 1962.

[Halpern and Moses, 1985] Joseph Y. Halpern and Yoram Moses. A guide to the modal logics of knowledge and belief: Preliminary draft. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 480–490, 1985.

[Halpern and Moses, 1992] Joseph Y. Halpern and Yoram Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[Harel et al., 1982] David Harel, Dexter Kozen, and Rohit Parikh. Process logic: Expressiveness, decidability, completeness. *Journal of Computer and Systems Sciences*, 25:144–170, 1982.

[Harel et al., 1983] D. Harel, A. Pnueli, and J. Stavi. Propositional dynamic logic of nonregular programs. *Journal of Computer and Systems Sciences*, 26:222–243, 1983.

[Harel, 1979] David Harel. *First-Order Dynamic Logic*, volume 68 of *Lecture Notes in Computer Science*. Springer-Verlag, 1979.

[Harel, 1984] David Harel. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 497–604. D. Reidel, 1984.

[Harrop, 1958] R. Harrop. On the existence of finite models and decision procedures for propositional calculi. *Proceedings of the Cambridge Philosophical Society*, 54:1–13, 1958.

[Hasle and Øhrstrøm, 2004] Peter Hasle and Peter Øhrstrøm. The flow of time into logic – and computer science. *Bulletin of the European Association for Theoretical Computer Science*, 82:191–226, February 2004.

[Henkin et al., 1971] Leon Henkin, J. Donald Monk, and Alfred Tarski. *Cylindric Algebras I*. North-Holland, Amsterdam, 1971.

[Henkin, 1949] Leon Henkin. The completeness of the first-order functional calculus. *The Journal of Symbolic Logic*, 14:159–166, 1949.

[Henkin, 1950] Leon Henkin. Completeness in the theory of types. *The Journal of Symbolic Logic*, 15:81–91, 1950.

[Hennessy and Liu, 1995] M. Hennessy and X. Liu. A modal logic for message passing processes. *Acta Informatica*, 32:375–393, 1995.

[Hennessy and Milner, 1980] Matthew Hennessy and Robin Milner. On observing nondeterminism and concurrency. In J. W. de Bakker and J. van Leeuwen, editors, *Automata, Languages and Programming. Proceedings 1980*, volume 85 of *Lecture Notes in Computer Science*, pages 299–309. Springer-Verlag, 1980.

[Hennessy and Milner, 1985] Matthew Hennessy and Robin Milner. Algebraic laws for nondeterminism and concurrency. *Journal of the Association for Computing Machinery*, 32:137–161, 1985.

[Hilbert and Bernays, 1939] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*. Springer, 1939.

[Hintikka, 1957] K. J. J. Hintikka. Quantifiers in deontic logic. *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum*, 23(4), 1957.

[Hintikka, 1961] K. J. J. Hintikka. Modality and quantification. *Theoria*, 27:119–128, 1961.

[Hintikka, 1969] K. J. J. Hintikka. Review of "The morning star paradox" by Stig Kanger. *The Journal of Symbolic Logic*, 34:305–306, 1969.

[Hoare, 1969] C. A. R. Hoare. An axiomatic basis for computer programming. *Communications of the Association for Computing Machinery*, 12:576–580, 583, 1969.

[Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, 1968.

[Hughes, 1990] G. E. Hughes. Every world can see a reflexive world. *Studia Logica*, 49:175–181, 1990.

[Huntington, 1937] Edward V. Huntington. Postulates for assertion, conjunction, negation, and equality. *Proceedings of the American Academy of Arts and Sciences*, 72:1–44, 1937.

[Jacobs and Rutten, 1997] Bart Jacobs and Jan Rutten. A tutorial on (co)algebras and (co)induction. *Bulletin of the European Association for Theoretical Computer Science*, 62:222–259, 1997.

[Jacobs, 1996] Bart Jacobs. Objects and classes, coalgebraically. In B. Freitag, C. B. Jones, C. Lengauer, and H.-J. Schek, editors, *Object-Orientation with Parallelism and Persistence*, pages 83–103. Kluwer Academic Publishers, 1996.

[Jacobs, 2000] Bart Jacobs. Towards a duality result in coalgebraic modal logic. *Electronic Notes in Theoretical Computer Science*, 33, 2000.

[Jacobs, 2002] Bart Jacobs. Exercises in coalgebraic specification. In R. Backhouse, R. Crole, and J. Gibbons, editors, *Algebraic and Coalgebraic Methods in the Mathematics of Program Construction*, volume 2297 of *Lecture Notes in Computer Science*, pages 237–280. Springer, 2002.

[Japaridze and de Jongh, 1998] Giorgi Japaridze and Dick de Jongh. The logic of provability. In Samuel R. Buss, editor, *Handbook of Proof Theory*, volume 137 of *Studies in Logic*, pages 475–546. Elsevier, 1998.

[Jónsson and Tarski, 1948] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators. *Bulletin of the American Mathematical Society*, 54:79–80, January 1948.

[Jónsson and Tarski, 1951] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators, part I. *American Journal of Mathematics*, 73:891–939, 1951.

[Jónsson and Tarski, 1952] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators, part II. *American Journal of Mathematics*, 74:127–162, 1952.

[Jónsson, 1967] Bjarni Jónsson. Algebras whose congruence lattices are distributive. *Math. Scand.*, 21:110–121, 1967.

[Jónsson, 1993] Bjarni Jónsson. A survey of Boolean algebras with operators. In *Algebras and Orders*, volume 389 of *NATO ASI Series*, pages 239–286. Kluwer Academic Publishers, 1993.

[Jónsson, 1994] Bjarni Jónsson. On the canonicity of Sahlqvist identities. *Studia Logica*, 53:473–491, 1994.

[Kamp, 1968] J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, University of California at Los Angeles, 1968.

[Kanger, 1957a] Stig Kanger. The morning star paradox. *Theoria*, 23:1–11, 1957.

[Kanger, 1957b] Stig Kanger. *Provability in Logic*. University of Stockholm–Almqvist & Wiksell, 1957.

[Kaplan, 1966] David Kaplan. Review of "Semantical analysis of modal logic I. Normal modal propositional calculi", by Saul A. Kripke. *The Journal of Symbolic Logic*, 31:120–122, 1966.

[Kawahara and Mori, 2000] Yasuo Kawahara and Masao Mori. A small final coalgebra theorem. *Theoretical Computer Science*, 233:129–145, 2000.

[Kneale and Kneale, 1962] William Kneale and Martha Kneale. *The Development of Logic*. Oxford University Press, 1962.

[Kolmogorov, 1925] A. N. Kolmogorov. On the principle of excluded middle (Russian). *Matematicheskii Sbornik*, 32:646–667, 1925. English translation by Jean van Heijenoort in *van Heijenhoort 1967*, pages 414–437.

[Kozen and Parikh, 1984] Dexter Kozen and Rohit Parikh. A decision procedure for the propositional μ-calculus. In E. Clarke and D. Kozen, editors, *Logics of Programs. Proceedings 1983*, volume 164 of *Lecture Notes in Computer Science*, pages 313–325. Springer-Verlag, 1984.

[Kozen and Tiuryn, 1990] Dexter Kozen and Jerzy Tiuryn. Logics of programs. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume B: Formal Models and Semantics*, pages 789–840. Elsevier, 1990.

[Kozen, 1982] Dexter Kozen. Results on the propositional μ-calculus. In M. Nielsen and E. M. Schmidt, editors, *Automata, Languages and Programming. Ninth Colloquium 1982*, volume 140 of *Lecture Notes in Computer Science*, pages 348–359. Springer-Verlag, 1982.

[Kozen, 1983] Dexter Kozen. Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

[Kozen, 1988] Dexter Kozen. A finite model theorem for the propositional μ-calculus. *Studia Logica*, 47:233–241, 1988.

[Kracht, 1999] Marcus Kracht. *Tools and Techniques in Modal Logic*, volume 142 of *Studies in Logic*. Elsevier, 1999.

[Kripke, 1959a] Saul A. Kripke. A completeness theorem in modal logic. *The Journal of Symbolic Logic*, 24:1–14, 1959.

[Kripke, 1959b] Saul A. Kripke. Semantic analysis of modal logic (abstract). *The Journal of Symbolic Logic*, 24:323–324, 1959.

[Kripke, 1962] Saul A. Kripke. The undecidability of monadic modal quantification theory. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 8:113–116, 1962.

[Kripke, 1963a] Saul A. Kripke. Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9:67–96, 1963.

[Kripke, 1963b] Saul A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.

[Kripke, 1965a] Saul A. Kripke. Semantical analysis of intuitionistic logic I. In J. N. Crossley and M. A. E. Dummett, editors, *Formal Systems and Recursive Functions*, pages 92–130. North-Holland, Amsterdam, 1965.

[Kripke, 1965b] Saul A. Kripke. Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 206–220. North-Holland, Amsterdam, 1965.

[Kripke, 1967] Saul A. Kripke. Review of [Lemmon, 1966b]. *Mathematical Reviews*, 34:1021–1022, 1967. MR 34 #5661.

[Kurz, 1998] Alexander Kurz. Specifying coalgebras with modal logic. *Electronic Notes in Theoretical Computer Science*, 11, 1998.

[Kurz, 2001] Alexander Kurz. Specifying coalgebras with modal logic. *Theoretical Computer Science*, 260:119–138, 2001.

[Ladner, 1977] Richard E. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal on Computing*, 6:467–480, 1977.

[Larsen, 1990] Kim G. Larsen. Proof systems for satisfiability in Hennessy-Milner logic with recursion. *Theoretical Computer Science*, 72:265–288, 1990.

[Lawvere, 1970] F. W. Lawvere. Quantifiers and sheaves. *Actes des Congrès International des Mathématiques*, 1:329–334, 1970.

[Lemmon and Scott, 1966] E. J. Lemmon and D. Scott. Intensional logic. Preliminary draft of initial chapters by E. J. Lemmon, Stanford University, July 1966. Later published as [Lemmon, 1977].

[Lemmon, 1957] E. J. Lemmon. New foundations for Lewis modal systems. *The Journal of Symbolic Logic*, 22:176–186, 1957.

[Lemmon, 1966a] E. J. Lemmon. Algebraic semantics for modal logics I. *The Journal of Symbolic Logic*, 31:46–65, 1966.

[Lemmon, 1966b] E. J. Lemmon. Algebraic semantics for modal logics II. *The Journal of Symbolic Logic*, 31:191–218, 1966.

[Lemmon, 1977] E. J. Lemmon. *An Introduction to Modal Logic*, volume 11 of *American Philosophical Quarterly Monograph Series*. Basil Blackwell, Oxford, 1977. Written in collaboration with Dana Scott. Edited by Krister Segerberg.

[Lewis and Langford, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic*. The Century Co., 1932. Second Edition, Dover Publications, 1959.

[Lewis, 1912] C. I. Lewis. Implication and the algebra of logic. *Mind (New Series)*, 21:522–531, 1912.

[Lewis, 1918] C. I. Lewis. *A Survey of Symbolic Logic*. University of California Press, Berkeley, CA, 1918. Second Edition, Dover Publications, 1960.

[Löb, 1955] M. H. Löb. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic*, 20:115–118, 1955.

[Loux, 1979] Michael J. Loux, editor. *The Possible and the Actual. Readings in the Metaphysics of Modality*. Cornell University Press, 1979.

[Łukasiewicz and Tarski, 1930] Jan Łukasiewicz and Alfred Tarski. Untersuchungen über den aussagenkalkül. *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 23:30–50, 1930. English translation by J. H. Woodger as *Investigations into the Sentential Calculus* in [Tarski, 1956], 38–59.

[Lyndon, 1950] Roger C. Lyndon. The representation of relational algebras. *Annals of Mathematics*, 51(3):707–729, 1950.

[Mac Lane, 1971] Saunders Mac Lane. *Categories for the Working Mathematician*. Springer-Verlag, 1971.

[MacColl, 1880] Hugh MacColl. Symbolic Reasoning. *Mind*, 5:54, 1880.

[MacColl, 1906] Hugh MacColl. *Symbolic Logic and its Applications*. Longmans, Green and Co., London, 1906.

[Macnab, 1976] D. S. Macnab. *An Algebraic Study of Modal Operators on Heyting Algebras, with Applications to Topology and Sheafification*. PhD thesis, University of Aberdeen, 1976.

[Macnab, 1981] D. S. Macnab. Modal operators on Heyting algebras. *Algebra Universalis*, 12:5–29, 1981.

[Makinson, 1966] D. C. Makinson. On some completeness theorems in modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:379–384, 1966.

[Makinson, 1969] D. C. Makinson. A normal modal calculus between T and S4 without the finite model property. *The Journal of Symbolic Logic*, 34:35–38, 1969.

[Makinson, 1970] D. C. Makinson. A generalisation of the concept of relational model for modal logic. *Theoria*, 36:331–335, 1970.

[Makinson, 1971] D. C. Makinson. Some embedding theorems for modal logic. *Notre Dame Journal of Formal Logic*, 12:252–254, 1971.

[Maksimova, 1992] L. L. Maksimova. Modal logics and varieties of modal algebras: The Beth properties, interpolation and amalgamation. *Algebra and Logic*, 31:90–105, 1992.

[Manes, 1976] Ernest G. Manes. *Algebraic Theories*. Springer-Verlag, 1976.

[Marx and Venema, 1997] Maarten Marx and Yde Venema. *Multi-Dimensional Modal Logic*. Kluwer Academic Publishers, 1997.

[Mates, 1986] Benson Mates. *The Philosophy of Leibniz*. Oxford University Press, 1986.

[McKenzie, 1972] Ralph McKenzie. Equational bases and nonmodular lattice varieties. *Transactions of the American Mathematical Society*, 174:1–43, 1972.

[McKinsey and Tarski, 1944] J. C. C. McKinsey and Alfred Tarski. The algebra of topology. *Annals of Mathematics*, 45:141–191, 1944.

[McKinsey and Tarski, 1946] J. C. C. McKinsey and Alfred Tarski. On closed elements in closure algebras. *Annals of Mathematics*, 47:122–162, 1946.

[McKinsey and Tarski, 1948] J. C. C. McKinsey and Alfred Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *The Journal of Symbolic Logic*, 13:1–15, 1948.

[McKinsey, 1941] J. C. C. McKinsey. A solution of the decision problem for the Lewis systems S2 and S4 with an application to topology. *The Journal of Symbolic Logic*, 6:117–134, 1941.

[McKinsey, 1945] J. C. C. McKinsey. On the syntactical construction of systems of modal logic. *The Journal of Symbolic Logic*, 10:83–94, 1945.

[Meredith, 1956] C. A. Meredith. Interpretations of different modal logics in the 'property calculus'. Mimeograph, Department of Philosophy, University of Canterbury, Christchurch, New Zealand, August 1956. Recorded and Expanded by A. N. Prior.

[Meyer, 1975] Albert R. Meyer. Weak monadic second order theory of successor is not elementary-recursive. In R. Parikh, editor, *Logic Colloquium. Symposium on Logic Held at Boston, 1972-73*, volume 453 of *Lecture Notes in Mathematics*, pages 132–154. Springer-Verlag, 1975.

[Milner *et al.*, 1993] Robin Milner, Joachim Parrow, and David Walker. Modal logics for mobile processes. *Theoretical Computer Science*, 114:149–171, 1993.

[Milner, 1971] Robin Milner. An algebraic notion of simulation between programs. In *Proceedings of the Second International Joint Conference on Artificial Intelligence*, pages 481–496, 1971.

[Milner, 1980] Robin Milner. *A Calculus of Communicating Systems*, volume 92 of *Lecture Notes in Computer Science*. Springer-Verlag, 1980.

[Milner, 1983] Robin Milner. Calculi for synchrony and asynchrony. *Theoretical Computer Science*, 25:267–310, 1983.

[Milner, 1989] Robin Milner. *Communication and Concurrency*. Prentice-Hall, 1989.

[Moggi, 1991] Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93:55–92, 1991.

[Montague, 1960] Richard Montague. Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry*, 3:259–269, 1960. Reprinted in [Montague, 1974].

[Montague, 1968] Richard Montague. Pragmatics. In R. Klibansky, editor, *Contemporary Philosophy: A Survey*, pages 102–122. La Nuova Italia Editrice, 1968. Reprinted in [Montague, 1974].

[Montague, 1970] Richard Montague. Pragmatics and intensional logic. *Synthèse*, 22:68–94, 1970. Reprinted in [Montague, 1974].

[Montague, 1974] Richard Montague. *Formal Philosophy (Selected Papers)*. Yale University Press, 1974. Edited by Richmond Thomason.

[Moss, 1999] Lawrence S. Moss. Coalgebraic logic. *Annals of Pure and Applied Logic*, 96:277–317, 1999.

[Nishimura, 1980] Hirokazu Nishimura. Descriptively complete process logic. *Acta Informatica*, 14:359–369, 1980.

[Ono, 1973] Hiroakira Ono. Incompleteness of semantics for intermediate predicate logics i. Kripke's semantics. *Proceedings of the Japan Academy*, 49(9):711–713, 1973.

[Parikh, 1978a] Rohit Parikh. The completeness of propositional dynamic logic. In *Mathematical Foundations of Computer Science 1978*, volume 64 of *Lecture Notes in Computer Science*, pages 403–415. Springer Verlag, 1978.

[Parikh, 1978b] Rohit Parikh. A decidability result for second order process logic. In *Proceedings of the 19th Annual IEEE Symposium on Foundations of Computer Science*, pages 177–183, 1978.

[Park, 1969] David Park. Fixpoint induction and proofs of program properties. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 5*, pages 59–78. Edinburgh University Press, 1969.

[Park, 1981] David Park. Concurrency and automata on infinite sequences. In P. Deussen, editor, *Theoretical Computer Science*, volume 104 of *Lecture Notes in Computer Science*, pages 167–183. Springer-Verlag, 1981.

[Pnueli, 1977] Amir Pnueli. The temporal logic of programs. In *Proceedings of the 18th Annual IEEE Symposium on Foundations of Computer Science*, pages 46–57, 1977.

[Pnueli, 1981] Amir Pnueli. The temporal semantics of concurrent programs. *Theoretical Computer Science*, 13:45–60, 1981.

[Ponse *et al.*, 1995] Alban Ponse, Maarten de Rijke, and Yde Venema, editors. *Modal Logic and Process Algebra: a Bisimulation Perspective*. CSLI Lecture Notes No. 53. CSLI Publications, Stanford, CA, 1995.

[Pratt, 1976] Vaughan R. Pratt. Semantical considerations on Floyd-Hoare logic. In *Proceedings of the 17th Annual IEEE Symposium on Foundations of Computer Science*, pages 109–121, 1976.

[Pratt, 1979] Vaughan R. Pratt. Process logic: Preliminary report. In *Proceedings of the 6th Annual ACM Symposium on Principles of Programming Languages*, pages 93–100, 1979.

[Pratt, 1980a] Vaughan R. Pratt. Application of modal logic to programming. *Studia Logica*, 39:257–274, 1980.

[Pratt, 1980b] Vaughan R. Pratt. A near-optimal method for reasoning about action. *Journal of Computer and System Sciences*, 20:231–254, 1980.

[Pratt, 1981] Vaughan R. Pratt. A decidable mu-calculus: Preliminary report. In *Proceedings of the 22nd Annual IEEE Symposium on Foundations of Computer Science*, pages 421–427, 1981.

[Prior and Fine, 1977] Arthur N. Prior and Kit Fine. *Worlds, Times and Selves*. Duckworth, 1977.

[Prior, 1956] Arthur N. Prior. Modality and quantification in S5. *The Journal of Symbolic Logic*, 21(1):60–62, 1956.

[Prior, 1957] Arthur N. Prior. *Time and Modality*. Clarendon Press, Oxford, 1957.

[Prior, 1958] Arthur N. Prior. The syntax of time-distinctions. *Franciscan Studies*, 18:105–120, 1958.

[Prior, 1962a] Arthur N. Prior. Possible worlds. *Philosophical Quarterly*, 12:36–43, 1962.

[Prior, 1962b] Arthur N. Prior. Tense-logic and the continuity of time. *Studia Logica*, 13:133–148, 1962.

[Prior, 1967] Arthur N. Prior. *Past, Present and Future*. Oxford University Press, 1967.

[Quine, 1947] W. V. Quine. The problem of interpreting modal logic. *The Journal of Symbolic Logic*, 12:43–48, 1947.

[Rabin, 1969] Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

[Read, 1998] Stephen Read. Hugh MacColl and the algebra of strict implication. *Nordic Journal of Philosophical Logic*, 3(1):59–83, 1998. Electronically available at: `www.hf.uio.no/filosofi/njpl/vol3no1/index.html`.

[Reichel, 1995] Horst Reichel. An approach to object semantics based on terminal co-algebras. *Mathematical Structures in Computer Science*, 5:129–152, 1995.

[Rescher and Urquhart, 1971] N. Rescher and A. Urquhart. *Temporal Logic*. Springer-Verlag, 1971.

[Reynolds, 2001] M. Reynolds. An axiomatization of full computation tree logic. *The Journal of Symbolic Logic*, 66(3):1011–1057, 2001.

[Rößiger, 1998] Martin Rößiger. From modal logic to terminal coalgebras. Technical Report MATH-AL-3-1998, Technische Universität Dresden, 1998.

[Rößiger, 2000] Martin Rößiger. Coalgebras and modal logic. *Electronic Notes in Theoretical Computer Science*, 33, 2000.

[Rößiger, 2001] Martin Rößiger. From modal logic to terminal coalgebras. *Theoretical Computer Science*, 260:209–228, 2001.

[Russell, 1905] Bertrand Russell. Necessity and possibility, 1905. Paper read to the Oxford Philosophical Society on 22 October 1905. Published in *The Collected Papers of Bertrand Russell, Volume 4*, pages 507–520, Routledge, 1994.

[Rutten and Turi, 1993] Jan J.M.M. Rutten and Daniele Turi. On the foundations of final semantics: Non-standard sets, metric spaces, partial orders. In J. W. de Bakker, W.-P. de Roever, and G. Rozenberg, editors, *Semantics: Foundations and Applications. REX Workshop*, volume 666 of *Lecture Notes in Computer Science*, pages 477–530. Springer Verlag, 1993.

[Rutten, 1995] J.J.M.M. Rutten. A calculus of transition systems (towards universal coalgebra). In Alban Ponse, Maarten de Rijke, and Yde Venema, editors, *Modal Logic and Process Algebra*, CSLI Lecture Notes No. 53, pages 231–256. CSLI Publications, Stanford, CA, 1995.

[Rutten, 1996] J.J.M.M. Rutten. Universal coalgebra: a theory of systems. Technical Report CS-R9652, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1996.

[Rutten, 2000] J.J.M.M. Rutten. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249(1):3–80, 2000.

[Sahlqvist, 1975] Henrik Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. In S. Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 110–143. North-Holland, Amsterdam, 1975.

[Sambin and Vaccaro, 1989] G. Sambin and V. Vaccaro. A new proof of Sahlqvist's theorem on modal definability and completeness. *The Journal of Symbolic Logic*, 54:992–999, 1989.

[Savitch, 1970] W. J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and Systems Sciences*, 4:177–192, 1970.

[Scott, 1970] Dana Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, 1970.

[Scroggs, 1951] Schiller Joe Scroggs. Extensions of the Lewis system S5. *The Journal of Symbolic Logic*, 16:112–120, 1951.

[Segerberg, 1968a] Krister Segerberg. Decidability of S4.1. *Theoria*, 34:7–20, 1968.

[Segerberg, 1968b] Krister Segerberg. Propositional logics related to Heyting's and Johansson's. *Theoria*, 34:26–61, 1968.

[Segerberg, 1970] Krister Segerberg. Modal logics with linear alternative relations. *Theoria*, 36:301–322, 1970.

[Segerberg, 1971] Krister Segerberg. *An Essay in Classical Modal Logic*, volume 13 of *Filosofiska Studier*. Uppsala Universitet, 1971.

[Segerberg, 1973] Krister Segerberg. Franzén's proof of Bull's theorem. *Ajatus*, 35:216–221, 1973.

[Segerberg, 1975] Krister Segerberg. That every extension of S4.3 is normal. In Stig Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 194–196. North-Holland, Amsterdam, 1975.

[Segerberg, 1977] Krister Segerberg. A completeness theorem in the modal logic of programs. *Notices of the American Mathematical Society*, 24:A–552, 1977.

[Segerberg, 1982] Krister Segerberg. A completeness theorem in the modal logic of programs. In T. Traczyck, editor, *Universal Algebra and Applications. Papers presented at the seminar held at the Stefan Banach International Mathematical Center, 1978*, pages 31–46. PWN–Polish Scientific Publishers, Warsaw, 1982. Banach Center Publications, vol. 9.

[Šehtman, 1977] V. B. Šehtman. On incomplete propositional logics. *Soviet Mathematics Doklady*, 18:985–989, 1977.

[Shehtman, 1983] V. B. Shehtman. Modal logics of domains on the real plane. *Studia Logica*, 42:63–80, 1983.

[Shoenfield, 1967] Joseph R. Shoenfield. *Mathematical Logic*. Addison-Wesley, 1967.

[Simpson, 1994] Alex K. Simpson. *The Proof Theory and Semantics of Intuitionistic Modal Logics*. PhD thesis, University of Edinburgh, 1994. Available at `www.dcs.ed.ac.uk/home/als/Research/`.

[Smoryński, 1985] C. Smoryński. *Self-Reference and Modal Logic*. Springer-Verlag, 1985.

[Smoryński, 1991] C. Smoryński. The development of self-reference: Löb's theorem. In Thomas Drucker, editor, *Perspectives on the History of Mathematical Logic*, pages 110–133. Birkhäuser, 1991.

[Sobociński, 1964] Boleslaw Sobociński. Remarks about axiomatizations of certain modal systems. *Notre Dame Journal of Formal Logic*, 5:71–80, 1964.

[Solovay, 1976] R. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25:287–304, 1976.

[Spaan, 1993] Edith Spaan. *Complexity of Modal Logics*. PhD thesis, University of Amsterdam, 1993.

[Stirling, 1987] Colin Stirling. Modal logics for communicating systems. *Theoretical Computer Science*, 49:311–347, 1987.

[Stirling, 1992] Colin Stirling. Modal and temporal logics. In S. Abramsky, Dov M. Gabbay, and T. S. E. Maibaum, editors, *Handbook of Logic in Computer Science, Volume 2*, pages 477–563. Oxford University Press, 1992.

[Stone, 1936] M. H. Stone. The theory of representations for Boolean algebras. *Transactions of the American Mathematical Society*, 40:37–111, 1936.

[Stone, 1937–1938] M. H. Stone. Topological representations of distributive lattices and Brouwerian logics. *Casopis pro Pestovani Matematiky a Fysiky*, 67:1–25, 1937–1938.

[Streett and Emerson, 1984] Robert S. Streett and E. Allen Emerson. The propositional mu-calculus is elementary. In J. Paradaens, editor, *Automata, Languages and Programming. Proceedings 1984*, volume 172 of *Lecture Notes in Computer Science*, pages 465–472. Springer-Verlag, 1984.

[Streett and Emerson, 1989] Robert S. Streett and E. Allen Emerson. An automata theoretic decision procedure for the propositional mu-calculus. *Information and Computation*, 81:249–264, 1989.

[Streett, 1982] Robert S. Streett. Propositional dynamic logic of looping and converse is elementarily decidable. *Information and Control*, 54:121–141, 1982.

[Tang, 1938] Tsao-Chen Tang. Algebraic postulates and a geometric interpretation for the Lewis calculus of strict implication. *Bulletin of the American Mathematical Society*, 44:737–744, 1938.

[Tarski, 1938] Alfred Tarski. Der aussagenkalkül und die topologie. *Fundamenta Mathematicae*, 31:103–134, 1938. English translation by J. H. Woodger as *Sentential Calculus and Topology* in *Tarski 1956*, 421–454.

[Tarski, 1941] Alfred Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6:73–89, 1941.

[Tarski, 1946] Alfred Tarski. A remark on functionally free algebras. *Annals of Mathematics*, 47:163–165, 1946.

[Tarski, 1955a] Alfred Tarski. Contributions to the theory of models III. *Koninklijkle Nederlandse Akademie van Wetenschappen, Proceedings, Series A*, 58:56–64, 1955. Indagationes Mathematicae, vol. 17.

[Tarski, 1955b] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[Tarski, 1956] Alfred Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford University Press, 1956. Translated into English and edited by J. H. Woodger.

[Thomas, 1989] Wolfgang Thomas. Computation tree logic and regular ω-languages. In W. P. de Roever J. W. de Bakker and G. Rozenberg, editors, *Linear Time, Branching Time, and Partial Order in Logics and Models for Concurrency*, volume 354 of *Lecture Notes in Computer Science*, pages 690–713. Springer-Verlag, 1989.

[Thomason, 1972a] S. K. Thomason. Noncompactness in propositional modal logic. *The Journal of Symbolic Logic*, 37:716–720, 1972.

[Thomason, 1972b] S. K. Thomason. Semantic analysis of tense logic. *The Journal of Symbolic Logic*, 37:150–158, 1972.

[Thomason, 1974a] S. K. Thomason. An incompleteness theorem in modal logic. *Theoria*, 40:30–34, 1974.

[Thomason, 1974b] S. K. Thomason. Reduction of tense logic to modal logic, I. *The Journal of Symbolic Logic*, 39:549–551, 1974.

[Thomason, 1975a] S. K. Thomason. Categories of frames for modal logic. *The Journal of Symbolic Logic*, 40:439–442, 1975.

[Thomason, 1975b] S. K. Thomason. The logical consequence relation of propositional tense logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 21:29–40, 1975.

[Thomason, 1975c] S. K. Thomason. Reduction of second-order logic to modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 21:107–114, 1975.

[Thomason, 1975d] S. K. Thomason. Reduction of tense logic to modal logic II. *Theoria*, 41:154–169, 1975.

[Tierney, 1972] Myles Tierney. Sheaf theory and the continuum hypothesis. In F. W. Lawvere, editor, *Toposes, Algebraic Geometry and Logic*, volume 274 of *Lecture Notes in Mathematics*, pages 13–42. Springer-Verlag, 1972.

[Urquhart, 1981] Alasdair Urquhart. Decidability and the finite model property. *Journal of Philosophical Logic*, 10:367–370, 1981.

[van Benthem, 1975] J. F. A. K. van Benthem. A note on modal formulas and relational properties. *Journal of Symbolic Logic*, 40(1):55–58, 1975.

[van Benthem, 1976a] J. F. A. K. van Benthem. *Modal Correspondence Theory*. PhD thesis, University of Amsterdam, 1976.

[van Benthem, 1976b] J. F. A. K. van Benthem. Modal formulas are either elementary or not $\Sigma\Delta$-elementary. *The Journal of Symbolic Logic*, 41:436–438, 1976.

[van Benthem, 1978] J. F. A. K. van Benthem. Two simple incomplete modal logics. *Theoria*, 44:25–37, 1978.

[van Benthem, 1979] J. F. A. K. van Benthem. Syntactic aspects of modal incompleteness theorems. *Theoria*, 45:67–81, 1979.

[van Benthem, 1980] J. F. A. K. van Benthem. Some kinds of modal completeness. *Studia Logica*, 39:125–141, 1980.

[van Benthem, 1983] J. F. A. K. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.

[van Benthem, 1984] J. F. A. K. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 167–247. D. Reidel, 1984.

[Vardi and Stockmeyer, 1985] Moshe Y. Vardi and Larry Stockmeyer. Improved upper and lower bounds for modal logics of programs. In *Proceedings of the 17th Annual ACM Symposium on the Theory of Computing*, pages 240–251. ACM, 1985.

[von Wright, 1951] G. H. von Wright. *An Essay in Modal Logic*. North-Holland, Amsterdam, 1951.

[Walukiewicz, 1995] Igor Walukiewicz. Completeness of Kozen's axiomatisation of the proposi-
    tional μ-calculus. In D. Kozen, editor, *Proceedings of the Tenth Annual IEEE Symposium on
    Logic in Computer Science*, pages 14–24. IEEE Computer Society Press, 1995.
[Walukiewicz, 2000] Igor Walukiewicz. Completeness of Kozen's axiomatisation of the proposi-
    tional μ-calculus. *Information and Computation*, 157:142–182, 2000.
[Winskel, 1985] Glynn Winskel. On powerdomains and modality. *Theoretical Computer Science*,
    36:127–137, 1985.
[Wolper, 1983] Pierre Wolper. Temporal logic can be more expressive. *Information and Control*,
    56:72–99, 1983.

EPISTEMIC LOGIC

Paul Gochet and Pascal Gribomont

INTRODUCTION

Epistemic logic grew in the Middle Ages. As early as the mid-twelfth century Garlandus and Abelard attempted to formulate an epistemic conception of entailment-propositions. Inspired by the efforts of Burley and Ockham, epistemic logic then blossomed during the first two decades of the fourteenth century. Intense research into epistemic logic is known to have been pursued at Oxford in about 1330; the second half of the fourteenth century witnessed the formulation of general rules for epistemic entailment-propositions by men like Strode and Peter of Mantua.

Medieval scholars conducted research into the relationship of truth to knowing, believing, and having faith. They discovered the *de dicto-de re* constructions anticipated by Aristotle and recognized inferences whose validity depends on epistemic/doxastic modalities. Even problems connected with iterated modalities or substitutivity in intentional contexts were given due consideration. The most active period of epistemic logic in the Middle Ages was during the fifteenth century. The main figures, Paul of Venice, Paul of Pergula, Gaetanus of Thiene, Frachantian of Vicenza, were affiliated with northern Italian universities. Readers interested in learning more about this period are encouraged to study Ivan Boh's classic monograph entitled *Epistemic Logic in the Later Middle Ages* [1993].

Jaakko Hintikka inaugurated contemporary research into epistemic logic with his book *Knowledge and Belief, an Introduction to the Logic of the Two Notions*. A thorough survey of epistemic logic from 1962 to 1978 has been provided by Lenzen [Lenzen, 1978]. Three major contributions to epistemic logic deserve particular attention: *Glauben, Wissen und Wahrscheinlichkeit. Systeme der epistemischen Logik* [Lenzen, 1980]; *Reasoning About Knowledge* [Fagin *et al.*, 1995] (the main source of our first and last sections); and *Epistemic Logic for AI and Computer Science* [Meyer and van der Hoek, 1995]. Section 1 of the present monograph is meant to be a general introduction to the subject for the newcomer. Only knowledge of first-order logic and modal propositional logic is presumed.

Sections 2 to 7 deal with special issues that have been intensively discussed over the last twenty years by logicians, philosophers, computer scientists, AI researchers and economists. These five sections follow the same pattern. We first state a major problem, or puzzle, which has prompted intense research in the field. We then describe the logical formalism which has been developed to solve this problem or puzzle. As semantics is more intuitive than axiomatics, we start our formal

presentation by spelling out the model theory. We then turn to proof theory. Detailed examples are provided when needed to facilitate comprehension.

Over the last twenty years logic (modal, temporal and epistemic) has become an efficient tool in the hands of computer scientists. The last section of the present work focuses on some important applications of epistemic logic to computer science. It shows how epistemic logic supplements temporal and other formal systems designed to specify and verify concurrent programs.

1   INTRODUCTION TO FORMAL EPISTEMIC LOGIC

1.1   Epistemic Interpretation Of Propositional Modal Logic

The modal operator $\Box$, sometimes also denoted by L, has received various interpretations, such as "It is necessary that . . . " or "It is mandatory that . . . ". Most of these are compatible with the semantics of Kripke structures. The simplest approach to epistemic logic is probably to view it as a modal logic and to interpret $\Box$ as "It is known that . . . ". With Kripke semantics, this leads to the Possible-Worlds Model of epistemic logic.

The propositional "modal-epistemic" logic has several variants, all based on the system K of modal logic [Hughes and Cresswell, 1984; 1996]. Formulas are built with a set of (elementary) propositions, usually denoted by $p, q, r, \ldots$, the Boolean connectives $\neg, \wedge, \vee, \supset, \equiv$ and the modal operator $\Box$. More precisely,

- Propositions are formulas.

- If A and B are formulas, then $\neg A$, $(A \wedge B)$, $(A \vee B)$, $(A \supset B)$, $(A \equiv B)$ and $\Box A$ are formulas.

- Nothing else is a formula.

Formulas are interpreted on *Kripke structures*. A Kripke structure M, also called a *model*, is made up of

1. a *frame*, that is, a directed graph; the nodes are named *states* or *worlds* and the arrows determine an *accessibility relation*, or *possibility relation*;

2. an *assignment function* π_s attached to each state s, that maps every atomic proposition on a truth value.

Comment. Some authors use the word "model" instead of "structure"; other use "model" for a structure (frame *cum* interpretation) which assigns the value *true* to some formula(s) or theory.

The interpretation rules assign truth values to formulas, for every state of the structure. The notations $(M, s) \models \varphi$, i.e. "(M, s) satisfies φ" (resp. $(M, s) \not\models \varphi$) mean that formula φ is true (resp. is false) at state s of structure M. Interpretation rules are

- If p is a proposition, $(M, s) \models p$ if and only if $\pi_s(p) = \mathbf{T}$.   [Basic rule]

- If φ and ψ are formulas, $(M, s) \models \neg\varphi$ if and only if $(M, s) \not\models \varphi$; $(M, s) \models \varphi \wedge \psi$ if and only if $(M, s) \models \varphi$ and $(M, s) \models \psi$; the other Boolean connectives are handled in a similar way.   [Classical rule]

- If φ is a formula, $(M, s) \models \Box\varphi$ if and only if $(M, s') \models \varphi$ for each state s' accessible from s.   [Modal rule]

Kripke semantics has long proved to be a convenient approach to assigning meaning to formulas in various logics. This is probably because the principle of Kripke structure is elementary but nevertheless versatile enough to account for many subtle distinctions in interpretation. Besides, in many cases, the modal rule has an intuitive meaning. For instance, when the modal operator is interpreted as a necessity operator, possible states from state s truly appear as the set of states which seem "possible" when looking from the "real" state s and, quite naturally, a formula φ is classified as possible when true in at least one of these states. It is classified as necessary when true in all these states.

Whether such an intuitive meaning applies to the knowledge operator is a matter of opinion. Let us assume that some state t is accessible from state s in some structure M. This could mean that, on the basis of the information available to the "epistemic agent" at state s, he or she cannot rule out state t as being the "real" state. So, if the agent *knows* p from state s, then p must be true at state t and at every state accessible from s (probably including s itself). On the contrary, the agent does not know p from state s if p happens to be false in some state accessible from state s, for instance t or s.

A more promising way to investigate whether Kripke semantics is appropriate for epistemic logic is to determine which formulas are valid, i.e. always true, and which are not. If the partition is intuitively acceptable, then the semantics will be acceptable too. We first introduce the notions of validity and satisfiability in a more formal way.

A formula is *satisfiable in structure M* if it is true at some state of M. A formula is *satisfiable* if it is satisfiable in some Kripke structure.

A formula is *valid in structure M* if it is true at all states of M. It is *valid* if it is valid in all Kripke structures. The symbol $\models$ is classically used to denote truth at a state, validity in some structure, and full validity:

- $M \models \varphi$ if and only if $(M, s) \models \varphi$ for each state $s \in M$.

- $\models \varphi$ if and only if $M \models \varphi$ for each Kripke structure M.

The knowledge operator is usually noted K instead of $\Box$, so "φ is known" is formalized into $\mathsf{K}\varphi$. Note that the formula $\neg\mathsf{K}\neg\varphi$ means "φ is not ruled out", that is, $\neg\varphi$ is not known.

Figure 1 contains seven axioms[1] and two inference rules, each listed with its usual abbreviation and name. These axioms and rules have interesting intuitive meanings and therefore have been used as "benchmarks" for formal epistemic logics. Some of them are widely accepted as intuitively valid, and so must be valid in any appropriate formal system, whereas others are a bit more controversial. The latter will be valid in some formal systems and simply satisfiable in other systems.

$P:$	Classical tautologies are valid	Tautology property
$K:$	$[\mathsf{K}\varphi \wedge \mathsf{K}(\varphi \supset \psi)] \supset \mathsf{K}\psi$	Distribution property
$T:$	$\mathsf{K}\varphi \supset \varphi, \ \varphi \supset \neg\mathsf{K}\neg\varphi$	Knowledge property
$B:$	$\varphi \supset \mathsf{K}\neg\mathsf{K}\neg\varphi$	Brouwerian property
$4:$	$\mathsf{K}\varphi \supset \mathsf{K}\mathsf{K}\varphi$	Positive introspection property
$5:$	$\neg\mathsf{K}\varphi \supset \mathsf{K}\neg\mathsf{K}\varphi$	Negative introspection property
$D:$	$\neg\mathsf{K}\ false$	Consistency property
$MP:$	$\dfrac{\varphi, \ \varphi \supset \psi}{\psi}$	*Modus ponens*
$KG:$	$\dfrac{\varphi}{\mathsf{K}\varphi}$	Knowledge generalization

Figure 1. Some typical epistemic statements

A first point is that it seems desirable for epistemic logic to be an extension of classical logic. Axiom P and rule MP are respected by Kripke semantics. So there is no problem here. The substitution property also holds, so the validity of, say, formula $\mathsf{K}p \supset (\mathsf{K}p \vee \neg\mathsf{K}\neg q)$ is a consequence of the validity of $\varphi \supset (\varphi \vee \psi)$.

A second important point is that Kripke semantics also enforces axiom K and rule KG. Axiom K provides an epistemic variant of *Modus ponens*: if φ and $\varphi \supset \psi$ are known, then ψ is known too. Besides, if φ is valid, then φ is known. The epistemic agent can be described by Kripke semantics only if assumed to be *logically omniscient*, meaning all logical consequences of known formulas are also known, in particular that all valid formulas are known (rule KG). That might be a non-realistic assumption as far as human reasoning is considered, but it is a natural assumption in computer science and in every application where knowledge is externally ascribed to the agent.

Axioms T, B, 4, 5 and D are valid in some Kripke structures but not in all. As a first example, let us consider M with states s and t and the accessibility relation $R_M = \{(s,t)\}$ with $\pi_s(p) = \mathbf{F}$ and $\pi_t(p) = \mathbf{T}$ (see Fig. 2). We have $(M,t) \models p$ but $(M,t) \models \mathsf{K}false$ and $(M,t) \models \mathsf{K}\neg p$ since no state is accessible from t. So axiom D and T are not valid in this structure. Axiom B is not valid either since

---

[1]More precisely, these formulas are axiom schemes, and become axioms when specific formulas are used in place of φ and ψ.

$(M, t) \models \mathsf{K}p$, so $(M, s) \models \neg\mathsf{K}\neg\mathsf{K}p$, but $(M, s) \not\models p$. Axiom 4 and 5 are valid in M. As a second example, assume structure N with states s, t and u and the accessibility relation such that every state is accessible from all states, except that s is not accessible from t and t is not accessible from s; the state function is such that $\pi_s(p) = \pi_u(p) = \mathbf{T}$ and $\pi_t(p) = \mathbf{F}$. Axioms T, B and D are valid in N but axioms 4 and 5 are not: $(N, s) \not\models \mathsf{K}p \supset \mathsf{K}\mathsf{K}p$ and $(N, u) \not\models \neg\mathsf{K}p \supset \mathsf{K}\neg\mathsf{K}p$.

Figure 2. Two Kripke structures

It is possible to turn some or all of these axioms into valid formulas if specific constraints are imposed upon the accessibility relation, for stronger constraints lead to stronger systems with more valid formulas. For instance, it is quite clear that axiom T is valid in all reflexive structures, that is, in structures where the accessibility relation is reflexive. In fact, axiom T expresses the reflexivity of the accessibility relation. This *knowledge property* is usually desirable, but sometime only the weaker *consistency property* is assumed. Indeed, $(M, s) \models \neg\mathsf{K}\neg true$ simply means that at least one state t is accessible from s (where *true* is satisfied); s is not necessarily accessible from itself. Otherwise stated, axiom D expresses that the accessibility relation $\mathbf{K}$ is serial: for each s there exists t such that $(s, t) \in \mathbf{K}$. Similarly, axiom B expresses symmetry and axiom 4 expresses transitivity of the accessibility relation; axiom 5 expresses that the relation is *Euclidean*: if (s, t) and (s, u) belong to the relation, then (t, u) also belongs to the relation.[2] These axioms are not independent. For instance, as all reflexive relations are also serial, D is a logical consequence of T. Similarly, sets $\{T, B, 4\}$ and $\{T, 5\}$ are logically equivalent since a relation is reflexive and Euclidean if and only if it is an equivalence relation.

---

[2]If an axiom expresses a relational property, then all structures where the accessibility relation enjoys this property satisfy all instances of the axiom. The converse is not true, but every "offending" model can be converted in an equivalent model where the accessibility relation enjoys the property; see [Fagin *et al.*, 1995] for more details.

Axioms T, B, 4, 5 and D, or some of them, together with axioms P and K and rules MP and KG, provide a family of sound and complete axiomatic systems for interesting epistemic logics. A '5' is usually used to designate the foremost member of this family. A formula is $S5$-valid if it is valid in all structures where the accessibility relation is an equivalence, that is, a reflexive, symmetric and transitive relation. All axioms of figure 1 are $S5$-valid. So, in some sense, system $S5$ is the strongest epistemic system based on Kripke semantics. The weakest system is usually named K; the accessibility relation is not constrained so none of the five axioms is K-valid. Some useful systems are listed in Fig. 3. Recall that all systems are based on axioms P and K, and rules MP and KG.

name	basic axioms	also valid	invalid	constraints
K			$T, B, 4, 5, D$	
D	D		$T, B, 4, 5$	serial
$K45$	$4, 5$		D, T, B	transitive, Euclidean
$KD45$	$D, 4, 5$		T, B	serial, transitive, Euclidean
T	T	D	$B, 4, 5$	reflexive
B	T, B	D	$4, 5$	reflexive, symmetric
$S4$	$T, 4$	D	$B, 5$	reflexive, transitive
$S5$	$T, 5$	$D, B, 4$		reflexive, symmetric, transitive

Figure 3. Some Propositional Epistemic Systems

All these systems are sound and complete, that is, the set of theorems is exactly the set of valid formulas. For instance, a formula is a T-theorem, that is, can be produced from axioms P, K and T and rules MP and KG, if and only if this formula is valid in every Kripke structure where the accessibility relation is reflexive.

Another desirable property inherited by these systems is decidability. In classical propositional logic, it is easy to obtain a model of any consistent formula. For instance, one can use the truth table method. All the epistemic systems listed in Fig. 3 enjoy that property, since any consistent formula admits a finite model, i.e. a Kripke model with finitely many states. However, the complexity of the satisfiability problem is likely to get worse when the epistemic operator is introduced into the propositional logic. Indeed, pure propositional logic can be seen as epistemic logic with one-state reflexive Kripke structures, for which $\mathsf{K}\varphi$ and $\neg\mathsf{K}\neg\varphi$ both reduce to φ. It can be proved that the satisfiability problem for most epistemic systems is PSPACE-complete, but for $K45$, $KD45$ and $S5$, it remains NP-complete, just as for classical propositional logic.[3] It is due to the fact that

---

[3]These classical notions of complexity theory will not be commented here. Just recall that, roughly speaking, NP-complete and PSPACE-complete decision problems are challenging cases for automated theorem proving. Reasonably efficient theorem provers may exist for the corresponding theories, but "unfavourable cases", that is, formulas which are not handled in a moderate amount of time (and space, for the PSPACE-complete case) seem unavoidable. The

for these systems, if a formula has a model, it has a small model. We refer the reader to [Fagin *et al.*, 1995] for details, and simply mention an interesting fact about $S5$. An equivalence relation on a set determines a partition of this set. So an $S5$-Kripke model is partitioned into smaller Kripke structures, which are also models. In these submodels the accessibility relation is universal, that is, every state is accessible from all states. As a result, a formula is $S5$-satisfiable if and only if it has a universal Kripke model. Furthermore, the search for a universal Kripke model for φ can be restricted to structures with at most $|\varphi|$ states, where $|\varphi|$ denotes the size of φ, that is, the number of symbol occurrences (propositions, connectives, epistemic operators) in φ.

1.2   *Multi-Agent Epistemic Logics*

The epistemic version of modal logic introduced in the first section allows one to reason about the knowledge of a single agent. In many applications, especially in computer science, artificial intelligence and game theory, it is also useful to reason about the knowledge of several agents.[4] A simple way to do that is to use a specific epistemic operator for every agent. For instance,

$$\mathsf{K}_1 \neg \mathsf{K}_2 p$$

intuitively means "Agent 1 knows that agent 2 does not know p".

Useful multi-agent epistemic logics can be obtained easily by extending the notion of Kripke structure. An n-agent Kripke structure M is simply a set S_M of states with n accessibility relations. These relations share the same domain S_M but are otherwise independent. An n-agent Kripke structure is reflexive (symmetric, serial, ...) if all its accessibility relations are reflexive (symmetric, serial, ...). Sound and complete axiomatizations for the n-agent case are obtained from the single-agent case in a straightforward way. For instance, the system T_n is the system T where the axiom T, that is, $\mathsf{K}\varphi \supset \varphi$ is replaced by the axiom set $\{\mathsf{K}_i\varphi \supset \varphi : i = 1, \ldots, n\}$, where K_i denotes the epistemic operator associated with agent i. A formula is a T_n-theorem if and only if it is T_n-valid, that is, valid in all reflexive n-agent Kripke structures. Similar soundness and completeness results hold for the other systems listed in Fig. 3.

The satisfiability problem remains decidable in the multi-agent case, and all consistent formulas have finite models. However, the satisfiability problem is PSPACE-complete for all the multi-agent versions of the systems listed in Fig. 3, including $S5_n$, $K45_n$ and $KD45_n$, as soon as $n \geq 2$. This is a small price to pay for a definite increase in expressive power and a wider class of application problems. Nevertheless, it is often necessary in most problems to distinguish between two kinds of knowledge, and therefore to increase the expressive power of our systems

---

more complex EXPTIME-complete case will be encountered later; reasonably efficient theorem provers do not exist in this case.

[4]As pointed out by W. van der Hoek, the issue in many games is to act as to maximize your own knowledge, but at the same time, to maximize your opponent's ignorance.

somewhat further. The "particular knowledge" of an agent, say 1, can increase when the agent observes something, or is told something by another (reliable) agent, say 2. The former case is adequately modelled by the formula

$$\mathsf{K}_1 p$$

that, in most systems used in practice, can be extended into

$$p \wedge \mathsf{K}_1 p \wedge \mathsf{K}_1 \mathsf{K}_1 p \wedge \mathsf{K}_1 \mathsf{K}_1 \mathsf{K}_1 p \wedge \ldots$$

whereas the second should be modelled by

$$p \wedge \mathsf{K}_1 p \wedge \mathsf{K}_2 p \wedge \mathsf{K}_1 \mathsf{K}_1 p \wedge \mathsf{K}_1 \mathsf{K}_2 p \wedge \mathsf{K}_2 \mathsf{K}_1 p \wedge \ldots$$

However, our systems rightly forbid us to deduce, say, $\mathsf{K}_1 \mathsf{K}_2 p$ from $\mathsf{K}_1 p \wedge \mathsf{K}_2 p$. It is indeed possible for some fact to be known by two agents, each of them thinking that he or she alone knows it. So, in the multi-agent case, we need to model explicitly that knowledge acquired by agent 1 from agent 2 is "common knowledge" between them, when this is the case. Note that, if the knowledge of p had been acquired by undetected eavesdropping on agent 2, it will not be common knowledge between agent 1 and agent 2 since, for instance, $\mathsf{K}_2 \mathsf{K}_1 p$ would not be true. Before introducing the notion of common knowledge in a formal way, let us mention another use of common knowledge in practical application. In most problems about knowledge, not only the "clues" specific to the problem are to be used in reasoning, but more general ground rules are to be used as well. For instance, it is implicitly admitted that, if agent 1 says something to agent 2, then agent 2 will reliably hear and record it, and therefore know it. Ground rules are typically pieces of knowledge common to whole sets of agents.

Let S be a subset of the set $\{1, 2, \ldots, n\}$ of agents. We introduce the new formula $\mathsf{C}_S \varphi$ to express that φ is common knowledge among the members of S. Let S^* be the (infinite) set of finite sequences of S-elements.[5] The length of a sequence $\sigma \in S^*$ is the number $\ell(\sigma)$ of its elements. These are denoted by $\sigma_1, \sigma_2, \ldots, \sigma_{\ell(\sigma)}$. From the semantic point of view, the following definition is rather clear:

$$\mathsf{C}_S \varphi \ =_{def} \ \bigwedge_{\sigma \in S^*} \mathsf{K}_{\sigma_1} \ldots \mathsf{K}_{\sigma_{\ell(\sigma)}} \varphi$$

but infinite conjunctions are not syntactically acceptable, so the operator C_S will be defined axiomatically. The appropriate axioms are easily obtained from the semantic description. We first observe that the formula

$$\mathsf{C}_S \varphi \ \equiv \ (\varphi \ \wedge \ \bigwedge_{i \in S} \mathsf{K}_i \mathsf{C}_S \varphi)$$

---

[5]For instance, an element of $\{1, 3\}^*$ is $(1, 1, 3, 1, 3)$.

must be valid for each formula φ. This is, therefore, a sound axiom scheme CK (Common Knowledge).[6] Furthermore, let us assume that, for some formulas φ and ψ, the formula

$$\varphi \supset \bigwedge_{i \in S} \mathsf{K}_i(\psi \wedge \varphi)$$

is valid in some structure M. It clearly follows that, for each natural number p, the formulas

$$\varphi \supset \bigwedge_{i_1 \in S} \mathsf{K}_{i_1} \ldots \bigwedge_{i_p \in S} \mathsf{K}_{i_p}(\psi \wedge \varphi)$$

and therefore

$$\varphi \supset \bigwedge_{i_1 \in S} \mathsf{K}_{i_1} \ldots \bigwedge_{i_p \in S} \mathsf{K}_{i_p} \psi$$

will also be valid in M. As a result, the formula

$$\varphi \supset \mathsf{C}_S \psi$$

is valid in M too. This leads to a sound rule of inference, called the *induction rule* (*IR*):

$$\frac{\varphi \supset \bigwedge_{i \in S} \mathsf{K}_i(\psi \wedge \varphi)}{\varphi \supset \mathsf{C}_S \psi}$$

It can be proved that the addition of Axiom CK and Rule IR are enough to turn the formal systems listed in Fig. 3 into (sound and) complete axiomatic systems for epistemic logic with common knowledge. Extended thus, System K_n is denoted by K_n^C. Similar notation is used for the other systems.

Common knowledge within a group could be interpreted as the knowledge that "any fool" in the group will have. It is sometimes interesting to define the *distributed knowledge* within a group. That would be the knowledge that a "wise person", capable of making all the group's implicit knowledge explicit, in turn identified with the knowledge possessed by at least one member of the group. Including distributed knowledge in formal systems is easy. First, the distributed knowledge within a group of one member reduces to this member's knowledge.[7] So, in self-explanatory notation,

$$\mathsf{D}_{\{i\}}\varphi \equiv \mathsf{K}_i \varphi$$

is a valid formula. Moreover, the larger the group, the greater the knowledge that this group possesses:

$$\mathsf{D}_S \varphi \supset \mathsf{D}_{S'} \varphi$$

if $S \subset S'$. Lastly, the operator D_S inherits from the properties of the knowledge operator. So new axioms are obtained from those of Fig. 1, by replacing K with D_S.

---

[6] Observe that the conjunction is now finite and thus acceptable here.

[7] For common knowledge, this is true only if axiom $S4$ is valid.

With these additions, the formal systems are extended in a sound and complete way for distributed knowledge.[8]

The introduction of common knowledge increases the complexity of the satisfiability problem, which becomes EXPTIME-complete for all systems.[9] The introduction of distributed knowledge does not increase the complexity of the satisfiability problem.

1.3   Epistemic Logic As First-order Logic

Classical first-order logic can be used to model nearly anything in an elementary way, and the propositional epistemic systems introduced above are no exception. Propositional formulas of an n-agent epistemic logic can be converted into first-order formulas in a straightforward way. If the set of elementary propositions is $p_1, \ldots, p_r$, the monadic predicate symbols will be $P_1, \ldots, P_r$. Furthermore, if the agents are $1, \ldots, n$, the dyadic predicate symbols will be $R_1, \ldots, R_n$. No other predicate symbols or function symbols are needed. A proposition p_i is translated into the first-order formula $P_i(x)$. Translation respects Boolean connectives. For instance, if propositional epistemic formulas φ and ψ are translated into first-order formulas φ^* and ψ^*, then the formula $\varphi \wedge \neg\psi$ will be translated into $\varphi^* \wedge \neg\psi^*$. Lastly, the formula $\mathsf{K}_j\varphi$ will be translated into $\forall y \, [R_j(x,y) \supset \varphi^*[x/y]]$.[10] The rules inductively define the so-called standard translation process, which leads to first-order formulas containing the single free variable x. This syntactic correspondence is supplemented with a semantic correspondence. More specifically, if φ is interpreted on a Kripke structure M, a corresponding first-order structure M^* for φ^* is obtained as follows. The domain of the structure M^* is the set of states S_M; $P_i(x)$ is interpreted as $\mathbf{T}$ for a valuation $V(x) = s$ if and only if $\pi_s(p_i) = \mathbf{T}$. Similarly, $R_j(x,y)$ is interpreted as $\mathbf{T}$ for a valuation $V(x) = s, V(y) = t$ if and only if (s,t) belongs to the accessibility relation associated with agent j in structure M. The idea behind this correspondence is to obtain $(M, s) \models \varphi$ if and only if $(M^*, V) \models \varphi^*$ where $V(x) = s$, with the consequence that φ^* (and therefore $\forall x \, \varphi^*$) will be valid if and only if φ is valid. It is possible to take into account usual restrictions about the accessibility relations. For instance, if they are reflexive, formula φ^* is replaced by $(\bigwedge_{j=1}^{n} \forall y \, R_j(y,y)) \supset \varphi^*$.

This elementary translation technique does not extend to the common knowledge operator C_S. To see this, we define the notion of S-reachability. A state t is S-reachable from state s in structure M if a finite sequence of states $s_0, \ldots, s_k$ exists such that $s_0 = s$, $s_k = t$ and, for each $\ell = 1, \ldots, k$, there exists $j \in S$ such that $(s_{\ell-1}, s_\ell)$ is an ordered pair of the accessibility relation R_j^M. It is clear that $(M, s) \models \mathsf{C}_S\varphi$ holds if and only if $(M, t) \models \varphi$ holds for each state t that is

---

[8]To prove this, a more formal semantics of distributed knowledge would be needed.

[9]Recall that, for systems including axiom $S4$, common knowledge reduces to knowledge if $n = 1$, so, say, $S4_1^C$ reduces to $S4_1$.

[10]If A is a formula, then $A[x/y]$ denotes the formula obtained by replacing all free occurrences of x by y.

S-reachable from state s. It is also clear that S-reachability is a binary relation, which is easily described in terms of the accessibility relations associated with the members of S. Indeed, it is the smallest relation X such that, first, if $(a, b) \in R_j$ with $j \in S$, then $(a, b) \in X$ and, second, if $(a, b) \in R_j$ with $j \in S$ and if $(b, c) \in X$, then $(a, c) \in X$. The problem is that we can specify in a few axioms that a binary relation does satisfy these requirements, but we cannot specify that a binary relation is the smallest one that satisfies these requirements.

It is theoretically interesting to determine that the (ordinary) knowledge operators can be eliminated by switching to first-order logic, but it is more useful to combine these operators in order to obtain a first-order epistemic logic with both quantification and knowledge operators. From the syntactic point of view, the combination is straightforward. From the semantic point of view, the natural idea is to extend classical first-order logic to modal or epistemic first-order logic just as classical propositional logic has already been extended to modal or epistemic propositional logic. In the propositional case, the extension has been easy. The attachment of a propositional interpretation to each node of a graph whose arcs are labelled with agents provides a propositional Kripke structure. In order to get a first-order Kripke structure, the propositional interpretation has to be replaced by a first-order interpretation. Observe that, in the propositional case, the lexicon, that is, the set of elementary propositions, is the same for all states, although the truth values associated with the propositions at distinct states may be distinct. A (classical) first-order interpretation contains a domain, a valuation function that assigns values to predicate symbols and function symbols (including individual constants) and a valuation that assigns a value to the variables. Just as in the propositional case, we assume that the set of symbols is the same for all states, but the interpretation of these symbols can be distinct in distinct states.

We assume an important restriction: the interpretation domain will be common to all states and so will be the valuation. This restriction is introduced to avoid problems in the interpretation of formulas like $\mathsf{K}_i P(x)$ and, on the whole, attempts to relax this "common-domain, rigid variables" (CDRV) assumption have raised more problems than they have solved. Nevertheless, the CDRV assumption has some unforeseen consequences, mainly due to the fact that quantifiers and knowledge operators do not always commute. For instance,

$$\mathsf{K}_i P(t) \;\supset\; \mathsf{K}_i \exists x \, P(x)$$

is valid, but

$$\mathsf{K}_i P(t) \;\supset\; \exists x \, \mathsf{K}_i P(x)$$

is not. Indeed, the term t might be *nonrigid*, that is, be interpreted as distinct elements of the interpretation domain in distinct states, whereas the variable x, as any variable, is a *rigid* term, interpreted as the same element of the domain in all states. Equality also raises a problem. Formula

$$(t = u) \;\supset\; \mathsf{K}_i(t = u)$$

is valid if t and u are rigid terms (this is the knowledge of equality axiom), but not otherwise.[11] This contradicts the feeling that all instances of

$$(t = u) \supset (\varphi(t) \equiv \varphi(u))$$

are valid. Indeed, the former is equivalent to an instance of the latter, where $\varphi(u)$ is $\mathsf{K}_i(t = u)$ and $\varphi(t)$ is therefore $\mathsf{K}_i(t = t)$, that is, *true*.

This suggests that the axiomatization of first-order epistemic logic is not straightforward. However, a fairly simple axiomatic system can be proved sound and complete. For the first-order version of K_n, such an axiomatic system comprises:

- The axioms and rules of the propositional system K_n;

- All instances of
  $$\varphi(t) \supset \exists x \, \varphi(x)$$
  and of
  $$(t_1 = t_2) \supset (\varphi(t_1) \equiv \varphi(t_2))$$
  such that, if $\varphi(x)$ contains a knowledge operator, then t, t_1, t_2 are variables;

- The Generalization rule
  $$\frac{\varphi \supset \psi(x)}{\varphi \supset \forall x \, \psi(x)}$$
  where x has no free occurrence in φ;

- The knowledge of inequality axiom:
  $$(x_1 \neq x_2) \supset \mathsf{K}_i(x_1 \neq x_2)\,;$$

- All instances of the Barcan formula:
  $$\forall x \, \mathsf{K}_i\varphi \supset \mathsf{K}_i\forall x \, \varphi\,.$$

Similar results hold for the other systems listed in Fig. 3.[12]

2   MULTI-MODAL EPISTEMIC LOGIC

2.1   The Relationship Between Knowledge And Belief

The relationship between knowledge and belief has been of concern for philosophers at least since the time of Plato. Towards the end of the *Meno* Plato draws a

---

[11]Two constants, like "The president of the United States" and "George Bush", may denote the same man in the real world, but not in some other possible world.

[12]For system $S5_n$, the knowledge of the inequality axiom and the Barcan axiom are not needed since they can be derived from the other axioms. The converse of the Barcan axiom is derivable in all systems.

distinction between true opinion (*doxa*) and knowledge, writing that the "guide who only thinks that this is the road to Larissa but is quite right gets us to Larissa as effectively as if he actually does know it. The defect of opinion, even when correct, is that, unlike knowledge, it can be shaken by criticism, conflicting evidence, authority, etc. [Ryle, 1967, pp. 325–326]". In the *Theaetetus* Plato resumed his inquiry into knowledge and belief. He came to the conclusion that knowledge is true belief plus something else, i.e. plus a *logos*.

A. J. Ayer is more precise than Plato was about what kind of thing it is that must be added to true belief in order to make it into knowledge. His definition of knowledge reads as follows: "the necessary and sufficient conditions for knowing that something is the case are first that what one is said to know be true, secondly that one be sure of it, and thirdly that one should have the right to be sure [Ayer, 1956, p. 34]".[13] By Ayer's definition, "X knows that φ" implies "X believes that φ". Seven years later E. Gettier [1963] laid down a counter-example to the above-mentioned definition of knowledge. Consider a man X who mistakes a dog for a sheep when he looks at a field. (The dog has been astutely disguised as a sheep by the farmer.) Suppose that there happens to be a sheep in the field that X does not see. Then X believes the proposition stating that there is a sheep in the field. That proposition is true and moreover it is justified. Yet we are not ready to say that X knows that there is a sheep in the field.

Counter-examples to Ayer's definition *à la* Gettier show that one cannot equate "X justifiably knows φ" with "X justifiably believes φ and φ is true" where "and" is truth-functional. The following case set up by F. Voorbraak [1992, p. 220] brings that out clearly. Imagine a situation in which an agent justifiably believes that p but not q. Suppose that $\neg p \wedge q$ is true. The agent believes $p \vee q$ and, under some reasonable assumptions, justifiably believes $p \vee q$. Moreover $p \vee q$ is true. Yet we cannot say that this agent knows that $p \vee q$ since it is believed for the wrong reason.

To circumvent this objection, one might be inclined to define "X justifiably knows φ" as "X justifiably believes φ and φ is true for the same reason that φ is justifiably believed". But if this definition is adopted, the first conjunct is no longer independent of the second and we cannot construe "X justifiably knows φ" as a truth-functional conjunction of "X justifiably believes φ" and "φ is true".

In view of the interdependence of the two conjuncts defining "X justifiably knows φ", the question as to whether justified knowledge implies justified belief ceases to be trivial. Conceptual analysis of notions like knowledge and belief taken in isolation is not likely to lead us very far. A study of the whole network of propositional attitudes with the techniques of logic seems to be more promising. The formal approach was launched with Hintikka's seminal book *Knowledge and Belief* [1962]. The subject was taken over by computer scientists and people working in the area of artificial intelligence. In "A guide to the modal logics of knowledge and belief" [Halpern and Moses, 1992 first version in 1985], questions of complexity were raised [Spaan, 1993a; 1993b]. A precise methodology for a systematic exam-

---

[13]The quote from [Ayer, 1956] has been reprinted with the permission of Palgrave Macmillan.

ination of possible combinations of epistemic operators was developed by W. van
der Hoek in "Systems of Knowledge and Belief" [1993].

2.2 Knowledge And Belief, How To Capture The Distinction

In the first section we followed current practice and used $S5$ (i.e. $KT45$) to capture
the knowledge operator K axiomatically. We have seen that this amounts to taking
the accessibility relation $\mathbf{K}$ to be an equivalence relation. This way of conceiving
knowledge is in keeping with the practice of economists whose model of knowledge
is based on partitions [Aumann, 1976].

As it is generally held that what distinguishes knowledge from belief is that the
former as opposed to the latter must be true, it appears that we could obtain an
axiomatic system for belief by substituting the operator B for K and replacing the
T axiom by the weaker axiom $D : (\mathsf{B}\varphi \supset \neg B\neg\varphi)$ or $D' : (\neg B \perp)$ which require
that beliefs be merely consistent rather than true.

To capture belief we introduce into the model a new accessibility relation, $\mathbf{B}$,
which serves to interpret the belief operator semantically when we spell out the
recursive definition of truth for the bimodal language containing both K and B.
Axiom T imposes reflexivity on the accessibility relation $\mathbf{K}$. It requires that every
world be accessible from itself. Axiom D requires that each world be such that
a world is accessible by $\mathbf{B}$ from it. That means that in whatever state the agent
may be, there is always at least one state to which he or she has access.

To distinguish belief from knowledge in the case of several agents, drastic
changes are required. Besides introducing a designated world $\mathbf{w}_0$, we should also
drop symmetry and impose only seriality and Euclideanity on the accessibility
relation. We can, however, retain something of the equivalence relations which
captured our intuition of indistinguishability between worlds. If $\mathbf{R}$ is Euclidean,
then restricting $\mathbf{R}$ to the set of the relata w' of $\mathbf{R}$, i.e. to $\{w' : (w, w') \in \mathbf{R}\}$,
amounts to turning its restricted part into an equivalence relation. Hence worlds
which are considered to be possible by the agent still form an equivalence relation,
but the real world need not be among them. Seriality, as we saw before, guarantees
that the agent thinks that some worlds are possible.

2.3 Kraus And Lehmann's Bimodal System KB_{CD}

S. Kraus and D. Lehmann [1988, p. 157] took over Halpern's and Moses' enterprise
and set up a model involving a separate accessibility relation for each knower
and a separate accessibility relation for each believer. Two states s, t are in the
equivalence relation $\equiv_i$ if the knowledge of person i cannot enable him or her
to distinguish between s and u. The relation $\equiv_i$ of indistinguishability is an
equivalence relation which divides the set of states into equivalence classes of states.
Another relation $\approx_i$ is introduced to capture belief. Relation $\approx_i$ is Euclidean and
serial, but it is not necessarily symmetric and reflexive. S. Kraus and D. Lehmann
compare the size of the set of propositions that an agent knows with the size of

the set of propositions that he believes: "It is easier to believe something than to know it, because one knows only true things, so one's beliefs can enable him to distinguish between more states than one's knowledge and therefore there could be some states s, u such that $s \equiv_i u$ but not $s \approx_i u$" [Kraus and Lehmann, 1988, p. 157]. In other words, if two states cannot be distinguished on the basis of an agent's beliefs, *a fortiori* they cannot be distinguished on the basis of this agent's knowledge. Formally this can be expressed by $s \approx_i u \supset s \equiv_i u$. Yet the concept of knowledge is richer than the concept of belief as Plato's and Ayer's definitions show. The idea that the concept of knowledge contains the concept of belief plus something else is reflected by the axiom $\mathsf{K}_i\varphi \supset \mathsf{B}_i\varphi$.

How are the two formulas related? The axiom, which has been rejected by some authors, can be derived from Kraus' and Lehmann's unquestionable observation that belief is more fine-grained than knowledge.

Let Bxy stand for 'world x is indistinguishable by belief from world y'. Let Kxy stand for 'world x is indistinguishable by knowledge from world y'. Let $Sx\varphi$ stand for 'world x forces (makes true) formula φ'. Kraus-Lehmann observation can be written

$$\forall\, x \,\forall\, y \,(Bxy \supset Kxy)\,.$$

From this we derive

$$\forall\, x \,\forall\, y \,[(Kxy \supset \alpha) \supset (Bxy \supset \alpha)]\,, \text{ for each statement } \alpha,$$

and, in particular,

$$\forall\, x \,\forall\, y \,[(Kxy \supset Sy\varphi) \supset (Bxy \supset Sy\varphi)]\,, \text{ for each formula } \varphi.$$

Standard quantification calculus leads to

$$\forall\, x \,\forall\, y \,(Kxy \supset Sy\varphi) \;\supset\; \forall\, x \,\forall\, y \,(Bxy \supset Sy\varphi)\,, \text{ for each formula } \varphi,$$

which reduces to

$$\mathsf{K}\varphi \supset \mathsf{B}\varphi\,, \text{ for each formula } \varphi.$$

The axiomatic system set up by Kraus and Lehmann will be denoted by KB_{CD}. It is designed to axiomatically capture the relationship between individual knowledge (K_i), individual belief (B_i), common knowledge (C) and common belief (D).[14] Following in W. van der Hoek's footsteps, we shall restrict ourselves to the examination of the KB subsystem. The latter consists of any axiomatic system of the standard propositional calculus to which the following axioms are added:

---

[14]Most authors use D for distributed knowledge, not for common belief. Kaneko *et al.* introduced $\mathsf{C_B}$ for common belief.

$K_i(\varphi \supset \psi) \supset (K_i\varphi \supset K_i\psi)$	Axiom K for knowledge
$K_i\varphi \supset \varphi$	Axiom T for knowledge
$\neg K_i\varphi \supset K_i\neg K_i\varphi$	Axiom 5 (negative introspection)
$B_i(\varphi \supset \psi) \supset (B_i\varphi \supset B_i\psi)$	Axiom K for belief
$\neg B_i \perp$	Axiom D for belief
$K_i\varphi \supset B_i\varphi$	Axiom KB 1
$B_i\varphi \supset K_iB_i\varphi$	Axiom KB 2

The inference rules are *Modus Ponens*, Necessitation for K, and Uniform Substitution in axioms and theorems.

The "bridge axioms" $KB1$ and $KB2$ relating belief to knowledge fit in with our intuition very well. Unfortunately, as F. Voorbraak observes, they license the theorem $B_iK_i\varphi \supset K_i\varphi$ which states that whoever believes to know some proposition φ does know it. Since in virtue of axiom T, which holds for $KT45$, one can only know the truth, φ must be true. Hence $B_iK_i\varphi \supset K_i\varphi$ commits us to saying that "one cannot believe to know a false proposition" and this is counterintuitive [Voorbraak, 1993, p. 8]. We are not perfect believers. A problem of the same kind arises in one of B. van Linder's Logics for rational agents (1996). This was shown by L. Simon who worked out an axiomatic system of epistemic-doxastic logic free of the unwanted formula $B_iK_i\varphi \supset K_i\varphi$ [Simon, 1998].

2.4   The Derivation Of The Paradox Of The Perfect Believer

Before looking for a remedy, we shall spell out the proof of the unwanted theorem and identify the axioms it rests upon. The proof reads as follows:

1.	$K_i\varphi \supset B_i\varphi$	Axiom KB
2.	$K_i\neg K_i\varphi \supset B_i\neg K_i\varphi$	$KB1 : \neg K_i\varphi/\varphi$
3.	$B_i\varphi \supset \neg B_i\neg\varphi$	Axiom D
4.	$B_i\neg\varphi \supset \neg B_i\varphi$	3, contraposition
5.	$B_i\neg K_i\varphi \supset \neg B_iK_i\varphi$	$4 : K_i\varphi/\varphi$
6.	$\neg K_i\varphi \supset K_i\neg K_i\varphi$	Axiom 5
7.	$\neg K_i\varphi \supset B_i\neg K_i\varphi$	6, 2, hypothetical syllogism
8.	$\neg K_i\varphi \supset \neg B_iK_i\varphi$	7, 5, hypothetical syllogism
9.	$B_iK_i\varphi \supset K_i\varphi$	8, contraposition

The proof rests upon three modal axioms: $KB1$, D and 5. If we wish to make the derivation of the unwanted conclusion impossible, we have to remove one of them.[15]

---

[15]Another way would be to restrict the use of the Substitution rule.

2.5   Voorbraak's System OK & RIB

Voorbraak chose the first option. He removed the axiom $\mathsf{K}_i\varphi \supset \mathsf{B}_i\varphi$ ($KB1$). His axiom system uses the axioms of $S5$ to capture K and those of $KD45$ to capture B. As bridge axioms between K and B, he puts forward:

$$KB2: \quad \mathsf{B}_i\varphi \supset \mathsf{K}_i\mathsf{B}_i\varphi$$
$$KB3: \quad \mathsf{B}_i\varphi \supset \mathsf{B}_i\mathsf{K}_i\varphi$$

F. Voorbraak is well aware that he breaks up with a long philosophical tradition in abandoning $\mathsf{K}_i\varphi \supset \mathsf{B}_i\varphi$. He concedes that the notion of knowledge that he has axiomatized, let us call it "objective knowledge", is unusual in so far as "it applies to any agent which is capable of processing information" irrespective of whether conscious belief states can be ascribed to it. Hence it applies to a device like a thermostat or a television receiver.

F. Voorbraak's notion of knowledge may be unusual but it is by no means a metaphoric notion. The notion of knowledge analyzed by Dretske comes very close to Voorbraak's objective knowledge [Dretske, 1981]. F. Voorbraak's axiom $KB3$ could also evince some misgivings. If Goldbach believed in the truth of his conjecture, does it follow, as Voorbraak's axiom would have it, that Goldbach also believed that he knew its truth? Presumably not since he called it a conjecture. F. Voorbraak will answer quite rightly that he aims at formalizing "believing" in a sense close to that of "being convinced of" ("*überzeugt sein*" in [Lenzen, 1980, p. 28]).

To sum up, the system of objective knowledge and rational introspective belief (OK & RIB) is "the normal modal system in the language L_{KB} which is obtained by adding the schemes $\mathsf{B}\varphi \supset \mathsf{B}\mathsf{K}\varphi$ and $\mathsf{B}\varphi \supset \mathsf{K}\mathsf{B}\varphi$ to the $S5$ principles for K and $KD45$ principles for B [Voorbraak, 1993, p. 62]". F. Voorbraak also succeeded in providing a sound and complete axiomatization for the notion of "justified belief" which plays an important role in epistemology. The system that he proposes for that purpose is a system intermediary between $S4$ and $S5$, namely $S4.2$, i.e. $S4+\neg\mathsf{B}\neg\mathsf{B}\varphi \supset \mathsf{B}\neg\mathsf{B}\neg\varphi$ in which the accessibility relation is confluent, i.e. satisfies the condition: $\forall s \forall t \forall u((sRt \wedge sRu) \supset \exists v(tRv \wedge uRv))$. The axiom $\neg\mathsf{B}\neg\mathsf{B}\varphi \supset \mathsf{B}\neg\mathsf{B}\neg\varphi$ is called axiom of convergence or axiom G in [Hughes and Cresswell, 1996, p. 134], where it is written $\neg\Box\neg\Box\varphi \supset \Box\neg\Box\neg\varphi$. Lenzen said that "[t]here is strong evidence in favor of the assumption that $S4.2$ is *the* logic of knowledge [1979, p. 33]".

2.6   Can Beliefs Be Inconsistent?

The second option we can take to block the derivation of $\mathsf{B}_i\mathsf{K}_i\varphi \supset \mathsf{K}_i\varphi$ consists of removing axiom D, i.e. $\mathsf{B}\varphi \supset \neg\mathsf{B}\neg\varphi$, and allowing $\neg(\mathsf{B}\varphi \supset \neg\mathsf{B}\neg\varphi)$ which is equivalent to $(\mathsf{B}\varphi \wedge \mathsf{B}\neg\varphi)$. By factorization, the latter implies $\mathsf{B}(\varphi \wedge \neg\varphi)$. The second option forces us to admit that we can believe the impossible. Can we? In *Principles of Human Knowledge*, Berkeley replies negatively: "Believing that

which involves a contradiction is impossible (Berkeley, 1710), [Marcus, 1993])".
Berkeley's claim, however, can be disproved by the following structure: take a
model M and a possible world w which has no world accessible from it. In that
structure $\mathsf{B}\varphi \wedge \mathsf{B}\neg\varphi$ trivially holds (as G. Sandu pointed out to us). One can
however drop axiom D, i.e. accept $(\mathsf{B}\varphi \wedge \mathsf{B}\neg\varphi)$ and side with Berkeley nevertheless,
i.e. one can accept $\mathsf{B}\varphi \wedge \mathsf{B}\neg\varphi$, but reject $\mathsf{B}(\varphi \wedge \neg\varphi)$. This is actually R. Barcan
Marcus' position. She is willing to renounce the principle of factorization $(\mathsf{B}\varphi \wedge$
$\mathsf{B}\psi) \supset \mathsf{B}(\varphi \wedge \psi)$. This means that doxastic logic ceases to be a regular modal system
(see [Chellas, 1980, p. 235]).

R. Fagin and J. Y. Halpern have set up a cluster semantics for which it is
possible to model the frame of mind of an agent who believes φ and believes ψ
without believing $\varphi \wedge \psi$ [Fagin and Halpern, 1988, pp. 58-59]. As E. Thijsse puts
it, on their view, "an agent is similar to a community in which different persons
may have different opinions, yet no one will defend contradictions. In a nutshell,
beliefs stemming from various frames of mind need not be combined by the agent"
[Thijsse, 1992, p. 170]. As an alternative to the semantic compartmentalization
just described, J. Dubucs advocates a proof-theoretic approach of the problem.
We have to explain why the inference (1) "Infer $\varphi \wedge \psi$ from the justifiable presence
of φ and ψ in the same context" is performable while the inference (2) "Infer $\varphi \wedge \psi$
from the justifiable presence of φ in a context, and the justifiable presence of ψ in
another context" is not. He argues that the difference rests upon a difference in
the conjunctions used in (1) and (2). The first one is context-sensitive, the second
is context-free. To capture the difference, he spells out the inferences under con-
sideration in the formalism of Gentzen's Calculus of Sequents and takes advantage
of the possibility to tamper with structural rules: "[...] the very possibility to
model the context-sensitivity of the inferences rests on the (partial) removing of
the structural rules: in order to prevent dissonant cognizers from admitting bla-
tant inconsistencies, we have to restrict the scope of these rules" [Dubucs, 1991,
p. 54].

Let us now turn to the third way of stopping the unwanted derivation, i.e.
renouncing negative introspection.

2.7   *Negative Introspection And The Paradox Of Infallibility*

Negative introspection, i.e. axiom 5: $\neg\mathsf{K}_i\varphi \supset \mathsf{K}_i\neg\mathsf{K}_i\varphi$, has been under attack
for some time. W. Lenzen raised the following objection: "[i]f an individual a is
completely sure of p's truth but nevertheless goes wrong about p, then he evidently
does not know that p, although he believes to know that p; and hence he is far
from knowing that he does not know that p. Unlike belief, knowledge is not
purely subjective in character but requires at least one objective mark, *viz.* truth.
Hence we cannot — by mere introspection — ascertain whether we know that p."
[Lenzen, 1978, p. 79]. T. Williamson observes that there are familiar situations
which disprove negative introspection. It happens that, when an agent does not
know p, it does not know that it does not know p and he adds: "That is because

it cannot survey the totality of its knowledge. It is a failure of self-knowledge, not of rationality in any ordinary sense" [Williamson, 2000, p. 317].[16]

If we cease to consider knowledge in isolation, but relate it to belief, a new and powerful objection crops up. Combined with other principles that we are not willing to abandon, negative introspection entails the agent's infallibility [Williamson, 2001]. More precisely, from 1 to 5 below, we can derive 6,

1.	$\mathsf{B}_i\varphi \supset \mathsf{B}_i\mathsf{K}_i\varphi$	$(KB3)$
2.	$\mathsf{B}\varphi \supset \neg\mathsf{B}\neg\varphi$	(D)
3.	$\mathsf{K}_i\varphi \supset \mathsf{B}_i\varphi$	$(KB1)$
4.	$\neg\mathsf{K}_i\varphi \supset \mathsf{K}_i\neg\mathsf{K}_i\varphi$	(5)
5.	$\mathsf{K}_i\varphi \supset \varphi$	(T)
6.	$\mathsf{B}_i\varphi \supset \varphi$	(Ω)

i.e. whatever agent i believes is true. Since this is clearly untenable, one of the five premises has to yield. Williamson and Simon chose to sacrifice the fourth, i.e. axiom 5. Voorbraak sacrificed the third, i.e. axiom $KB1$. The proof (due to Williamson) that Ω follows from the premises is left as an exercise. It is made easier if we avail ourselves of axiom B: $\neg\varphi \supset \mathsf{K}_i\neg\mathsf{K}_i\varphi$, an instance of Brouwer's principle which is derivable from axioms T and 5. We shall return to $KB3$ in the next section.

Yet, were it not for the unwanted consequence (infallibility ascribed to the agent), we have very good reasons to take up $S5$ as a formal system for knowledge and $KD45$ as a formal system for belief while adopting $\mathsf{K}\varphi \supset \mathsf{B}\varphi$ and $\mathsf{B}\varphi \supset \mathsf{BK}\varphi$ as bridge laws between knowledge and belief. J. Halpern has shown that we can do so without falling prey to the infallibility predicament; what we have to do is to restrict axiom $KB1$ to objective, i.e. non modal formulas. This means, *e.g.*, that $\mathsf{K}\neg\mathsf{K}p \supset \mathsf{B}\neg\mathsf{K}p$ is not eligible as a substituend for $\mathsf{K}\varphi \supset \mathsf{B}\varphi$. The weakened system spelled out to capture this restriction is sound and complete [Halpern, 1996].

Several sets of axioms for multi-modal logic have been presented. Some of those sets contain members which look quite acceptable as long as they are taken in isolation but produce unexpected and unacceptable results when they are brought together. At this stage we need a general method to make a systematic exploration of all possible combinations of epistemic operators. Such a method has been worked out by W. van der Hoek who applied correspondence theory to multi-modal systems. Correspondence theory was independently created by H. Sahlqvist [1975] and J. van Benthem [1976; 1983].

2.8   Correspondence Theory For Comparing Multi-modal Logics

The problem that W. van der Hoek addresses can be stated in this way: given some epistemic logic KB* which we have proved sound and complete with respect to a class of canonical frames, we want to know whether adding a new axiom φ_1

---

[16]The quote from [Williamson, 2000] has been reprinted with the permission of Oxford University Press, Oxford, UK.

will make an unwanted theorem φ_2 derivable or not. W. van der Hoek shows that under some proviso the problem can be reformulated as the question whether the first-order property Φ_1 which corresponds to modal formula φ_1 implies the first-order property Φ_2 which corresponds to modal formula φ_2. Two definitions [van der Hoek, 1993, p. 181] are needed. We presume the definitions of canonical model and canonical frame [Hughes and Cresswell, 1984].

1. An axiom scheme φ belonging to the language of a normal modal logic L is canonical if and only if φ is satisfied by the canonical frame for L.

2. If Φ is a (first-order) property of the class $\mathbf{D}$ of canonical frames $\mathbf{F^c}$, φ corresponds with Φ [formally: $\varphi \sim_D \Phi$] if $\mathbf{F^c}$ makes the formula φ true if and only if it satisfies Φ.

With this apparatus, a canonical frame $\mathbf{F^c} = \langle \mathbf{W^c}, \mathbf{K^c}, \mathbf{B^c} \rangle$ can be constructed for KB* such that

- $\mathbf{F^c}$ makes the axioms of KB* true (hence "canonical") and

- the following equivalences are obtained with respect to the canonical frames of KB*:

  (a) $\forall x \, \forall y \, \forall z \, [(x\mathbf{B}y \wedge y\mathbf{K}z) \supset x\mathbf{B}z] \sim_{KB*^c} [\mathsf{B}_i\varphi \supset \mathsf{B}_i\mathsf{K}_i\varphi]$;

  (b) $\forall x \, \forall y \, (x\mathbf{B}y \supset x\mathbf{K}y) \sim_{KB*^c} [\mathsf{K}_i\varphi \supset \mathsf{B}_i\varphi]$;

  (c) $\forall x \, \exists y \, [x\mathbf{B}y] \sim_{KB*^c} [\neg B \perp]$;

  (d) $\forall x \, \forall y \, \forall z \, [(x\mathbf{K}y \wedge x\mathbf{K}z) \supset y\mathbf{K}z] \sim_{KB*^c} [\neg\mathsf{K}_i\varphi \supset \mathsf{K}_i\neg\mathsf{K}_i\varphi]$.

As an example we shall prove the first one in the left to right direction. We want to demonstrate that $\forall x \forall y \forall z((x\mathbf{B}y \wedge y\mathbf{K}z) \supset x\mathbf{B}z)$ implies $\mathsf{B}_i\varphi \supset \mathsf{B}_i\mathsf{K}_i\varphi$. Assume to the contrary that the consequent is false, i.e. that we have both $w_0 \models \mathsf{B}_i\varphi$ and $w_0 \not\models \mathsf{B}_i\mathsf{K}_i\varphi$. The first conjunct can be rewritten $\forall y(w_0\mathbf{B}y \supset y \models \varphi)$ and the second: $\neg\forall y\forall z((w_0\mathbf{B}y \wedge y\mathbf{K}z) \supset z \models \varphi)$. Pushing the negation inside we get $\exists y\exists z((w_0\mathbf{B}y \wedge y\mathbf{K}z) \wedge z \not\models \varphi)$. By instantiation we obtain $((w_0\mathbf{B}w_1 \wedge w_1\mathbf{K}w_2) \wedge w_2 \not\models \varphi)$. But in virtue of the left side, we have $((w_0\mathbf{B}w_1 \wedge w_1\mathbf{K}w_2) \supset w_0\mathbf{B}w_2)$. By *modus ponens* we get $w_0\mathbf{B}w_2$, which, together with $\forall y(w_0\mathbf{B}y \supset y \models \varphi)$, gives us $w_2 \models \varphi$ by trivial transformations. We have arrived at our contradiction as expected.

W. van der Hoek states this startling result: when taken together, the four axioms mentioned on the right side of the equivalences (a) - (d) entail the theorem $\mathsf{B}_i\varphi \supset \mathsf{K}_i\varphi$, which, when combined with its converse, i.e. axiom (b), produces the collapse of the distinction between knowledge and belief. Thanks to the correspondence theory the proof of W. van der Hoek's theorem is an easy exercise in first-order logic. It is outlined below.

Assume the negation of the first-order formula corresponding to $\mathsf{B}_i\varphi \supset \mathsf{K}_i\varphi$, i.e.

(e) $\neg\forall\, x\forall\, y(x\mathbf{K}y \supset x\mathbf{B}y)$

and add them to the four first-order formulas corresponding to the axioms; a contradiction is easily obtained:

1.	$\forall\, x\,\forall\, y\,\forall\, z\,(x\mathbf{B}y \wedge y\mathbf{K}z \supset x\mathbf{B}z)$	(a)
2.	$\forall\, x\,\forall\, y\,(x\mathbf{B}y \supset x\mathbf{K}y)$	(b)
3.	$\forall\, x\,\exists\, y\,x\mathbf{B}y$	(c)
4.	$\forall\, x\,\forall\, y\,\forall\, z\,(x\mathbf{K}y \wedge x\mathbf{K}z \supset y\mathbf{K}z)$	(d)
5.	$\exists\, x\,\exists\, y\,(x\mathbf{K}y \wedge \neg x\mathbf{B}y)$	(e), trivial transformation
6.	$a\mathbf{K}b \wedge \neg a\mathbf{B}b$	5, existential instantiation
7.	$\exists\, y\,a\mathbf{B}y$	3, universal instantiation
8.	$a\mathbf{B}c$	7, existential instantiation
9.	$a\mathbf{B}c \supset a\mathbf{K}c$	2, universal instantiation
10.	$a\mathbf{K}c \wedge a\mathbf{K}b \supset c\mathbf{K}b$	4, universal instantiation
11.	$a\mathbf{B}c \wedge c\mathbf{K}b \supset a\mathbf{B}b$	1, universal instantiation
12.	$a\mathbf{B}b$	$6, 8\text{-}11$, trivial transformation
13.	$a\mathbf{B}b \wedge \neg a\mathbf{B}b$	$6, 12$

After splitting (6) into two conjuncts we get a contradiction with (12).

The second part of W. van der Hoek's theorem reads as follows: for each proper subset of $\{a, b, c, d\}$, counter-models can be built which show that none of those sets of axioms entails the collapse of the distinction between knowledge and belief [van der Hoek, 1993, pp. 187-188]. Since everybody wants to retain (a) and (c), the number of viable alternatives between which we have to choose has been drastically reduced. This significant advance was made possible by applying correspondence theory.

2.9 Implicit Versus Explicit Beliefs

In 1984, H. J. Levesque [1984] developed a logic which distinguishes between implicit and explicit belief. The enterprise was taken over by R. Fagin and J. Y. Halpern who extended Levesque's logic and spelled out two distinct accounts of awareness. We shall restrict ourselves to the second one: general awareness. They define a Kripke structure for general awareness as a tuple $\mathbf{M} = (\mathbf{S}, \pi, \mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{B}_1, \ldots, \mathbf{B}_n)$. $\mathbf{S}$ is a set of states, π is a truth assignment to primitive propositions for each state member of $\mathbf{S}$, and $\mathbf{B}_i$ is the accessibility relation for the interpretation of implicit beliefs of agent i, $\mathbf{A}_i$ is an arbitrary set of formulas, the formulas of which agent i is aware. We can also view it as a function which assigns to agent i the set of formulas of which he is aware.

The set of formulas of which the agent is aware may contain contradictory formulas without the agent entertaining inconsistent beliefs: being aware is not believing. To complete the semantics, a recursive definition of truth in model $\mathbf{M}$

is provided. The following clauses interpret the operator of awareness (A), that of implicit belief (L) and that of explicit belief (B):

$$\mathbf{M}, s \models \mathsf{A}_i\varphi \text{ iff } \varphi \in \mathbf{A}_i.$$
$$\mathbf{M}, s \models \mathsf{L}_i\varphi \text{ iff } \mathbf{M}, t \models \varphi \text{ for all } t \text{ such that } s\mathbf{B}t.$$
$$\mathbf{M}, s \models \mathsf{B}_i\varphi \text{ iff } \varphi \in \mathbf{A}_i \text{ and } \mathbf{M}, t \models \varphi \text{ for all } t \text{ such that } s\mathbf{B}t.$$

R. Fagin and J. Y. Halpern observe that a complete axiomatization of the logic of awareness can be obtained by adding the axiom $\mathsf{B}_i\varphi \equiv (\mathsf{L}_i\varphi \wedge \mathsf{A}_i\varphi)$ to the axioms of $KD45$. The distinction between explicit and implicit beliefs was introduced to solve the problem of logical omniscience which will be examined in detail in section 6. The idea is that the agent is a perfect logician only at the level of his or her implicit beliefs, not of his or her explicit beliefs. This view strikes us as too optimistic. Could we not instead claim that we are sometimes illogical and irrational in what we implicitly believe? What about prejudices and beliefs to which we stick against factual evidence? To handle these, W. van der Hoek and J.-J.Ch. Meyer suggested to supplement the awareness function introduced by R. Fagin and J. Y. Halpern by a prejudice function P that gives a set of formulas for each world representing the beliefs the agent wants to stick to. With this utility, we can formalize the behaviour of an agent who reasons against the facts [van der Hoek and Meyer, 1989, pp. 186].

2.10 Implicit Contradictions Becoming Explicit

If we agree, as we think we should, that far from being implicitly logically omniscient we can be implicitly illogical, we have to reconcile this with the fact that we do not believe explicit contradictions. To tackle this problem successfully we first need to get a firm hold on the notion of belief itself.

Let us consider R. Barcan Marcus' account of knowledge and belief. On her view, knowledge is a relation between an agent and a real state of affairs. When we discover that some proposition p we held to be true is false, we do not say "we knew p". Instead we withdraw our previous claim and say "we thought we knew but we only believed p". Belief, R. Barcan Marcus says, is a relation between an agent and a possible state of affairs [Marcus, 1993, p. 145]. When we are shown that the content of a belief is an impossible state of affairs, we withdraw our claim to believe. We do not say "we believed p". We rather say "we thought we believed".

On the account of belief just presented, we cannot believe things which we know to be inconsistent, but we can believe propositions which, unknown to us, are inconsistent. A full treatment of the logic of belief cannot ignore the process of becoming aware of an inconsistency and the revision of belief which it triggers. To handle that question we need a dynamic approach to epistemic logic. This will be the topic of the next section.

2.11   Implicit Contradictions In The Language

Up to this point, we have considered contradictions hidden in theories. There is, however, a more pervasive, hidden contradiction to be taken care of. Tarski showed that the antinomy of the liar can be obtained in any language which makes four assumptions [Tarski, 1949, pp. 58-59], [Haack, 1978]:

1. The language is universal in the sense that it contains its own metalanguage.

2. The ordinary laws of logic hold.

3. An empirical premise such as "The sentence printed in this paper on page n, line m, is not true" can be formulated and asserted.

4. For any sentence p, a sentence of the form "'p' is true if and only if p" belongs to the language.

Ordinary language fulfills these four conditions. Hence, we seem to be committed to the contradictory statement "s is true if, and only if, s is not true", even before adopting a particular theory. F. Orilia came to grips with this problem and showed how we can reason in spite of the fact that we use a universal and thus possibly inconsistent language [2000, p. 292]. F. Orilia reconsiders assumption 2 and puts forward the idea that logical rules can be construed as default rules which can admit exceptions.

Although Orilia's approach bears some resemblance to Priest's proposal of a nonmonotonic version of his paraconsistent logic [Batens *et al.*, 2000], it differs from it insofar as for Orilia "there is no *a priori* decision regarding which inference rules [...] may have exceptions. It all depends on the entrenchment ordering" [Orilia, 2000, 19, p. 295]. Orilia supplies an algorithm for belief revision which enables us to remove both the alethic paradoxes and the inconsistencies which come to the surface as dispositional beliefs acquire the status of active beliefs. The algorithm takes due account of the entrenchment.

3   MULTI-AGENT SYSTEMS

3.1   The Modeling Of Knowledge Change Via Interpreted Systems

The muddy children puzzle involves several agents, the father and the children, interacting over time. If we want to describe such a multi-agent system, Kripke's possible world semantics can be used. There is however an alternative semantic framework designed to model interactions which occur in time, namely interpreted systems, which we shall briefly describe here [Fagin *et al.*, 1995].

Fagin *et al.* assume that if we look at the system at any point of time each agent is in some state, called *local state*, which encodes all the information to which the agent has access. Once it is granted that each agent is in some state, it becomes quite natural to consider that the whole system itself is in some state. If we allow

for the role of the environment, we can define the notion of the global state of a system in this way: a global state of a system with n agents is an $(n+1)$-tuple of the form $(s_e, s_1, \ldots, s_n)$ where s_e is the state of the environment and s_i is the local state of each agent i.

If L_i is used to denote a set of possible local states for agent i, we can take $G = L_e \times L_1 \times \cdots \times L_n$ to be the set of *global states*. If we want to represent the temporal evolution of the system, we have to complicate the picture a little bit and introduce the notion of a run over G where "run" denotes a function from the time domain (the natural numbers if we use discrete time) to G.

A *non-interpreted system* S is a subset of the Cartesian product of global states $S \subseteq L_e \times L_1 \times \cdots \times L_n$. An *interpreted system* is a system S together with an interpretation function ($\pi : P \mapsto 2^S$) which assigns a subset of S to every atomic sentence. The difference between a non-interpreted system and an interpreted system is very much the same as the difference between a Kripke frame and a Kripke model.

Accessibility relations are an essential ingredient of Kripke frames and models. There are no relations of that kind in systems. To bring together systems and frames we need something in systems which corresponds to the accessibility relations in frames. Consider two global states L and L' in systems. When can we say that they are epistemically indistinguishable for agent i, i.e. when can we say that they are related to one another by the equivalence relation $\sim$? The answer is provided by the following definition: two global states L and L' are epistemically indistinguishable whenever their respective local states ℓ and ℓ' are identical. Formally: $L \sim L'$ iff $\ell = \ell'$.

3.2   Hypercubes As A Proper Subset Of Equivalence Frames

Between standard Kripke models and interpreted systems made up of runs designed to describe temporal evolution, there is an intermediate class of systems which has been scrutinized at depth [Lomuscio, 1999]. It turns out that the systems of that intermediate class are ideally suited for the logical exploration of static epistemic properties which can only be ascribed to a network of several interacting agents such as the property of sharing knowledge.

That intermediate class H is dubbed the class of hypercube systems. Hypercube systems are a variety of interpreted systems in which the full Cartesian product of non-empty sets of local states is considered. Formally, $H = L_1 \times \cdots \times L_n$; the environment states are ignored. We are facing two ways of modeling knowledge:

1. The standard way provided by $S5$ Kripke models whose frames consist of $\langle W, \sim_1, \sim_2, \cdots, \sim_n \rangle$ where $\sim_i$ is an equivalence relation which serves to capture the intuition that "agent i considers t possible in world s if ... the two worlds are indistinguishable to the agent" [Fagin *et al.*, 1995, p. 18].

2. The new way provided by hypercube systems in which the identity relation between local states belonging to global states plays the role of the equivalence relations between worlds.

The question arises whether it is possible to characterize the relationship between hypercubes and frames in a rigorous way. As a preliminary step, Lomuscio defines a map f from hypercubes H to Kripke frames F as follows. If $H = L_1 \times \cdots \times L_n$, then $f(H) = (L_1 \times \cdots \times L_n, \sim_1, \sim_2, \cdots, \sim_n)$ where $\sim_i$ is defined as $(\ell_1, \ldots, \ell_n) \sim_i (\ell'_1, \ldots, \ell'_n)$ if and only if $\ell_i = \ell'_i$. Applying the map f just defined to a hypercube H, we obtain an equivalence frame, namely $F = (W, \sim_1, \sim_2, \cdots, \sim_n)$. The frame thus obtained however is not any arbitrary equivalence frame. It is, as Lomuscio stresses, a very peculiar equivalence frame, i.e. a frame which can be proved to have these two properties:

1. The *identity-intersection property.* The intersection of the equivalences relations boils down to the identity between worlds: $\bigcap_{i \in A} \sim_i = \text{id}_W$.

2. The *n-directedness* of the frame. For any world $w_i \in W$, there is a world $\underline{w}$ such that $\underline{w} \sim_i w_i$ for $i = 1, \ldots, n$.

It follows that the Kripke frames obtained up to now from hypercubes are a proper subset of the set of Kripke equivalence frames.

On the basis of further theorems about the relationship between hypercubes and Kripke equivalence frames, Lomuscio succeeds in isolating a class of Kripke equivalence frames G which is semantically equivalent to the class of hypercubes, i.e. which verifies (respectively falsifies) the same formulas, namely the class of ID-equivalence frames which enjoy the identity-intersection and the n-directedness properties. This result shows that the task of axiomatizing the set of hypercubes boils down to that of axiomatizing the G class of Kripke equivalence frames.

3.3   The Axiomatization Of Hypercubes

Equivalence frames are axiomatized by $S5$. As hypercubes can be seen as equivalence frames of a special sort (ID-frames) which satisfy two special conditions (identity-intersection and directedness), we expect that for the axiomatization of hypercubes the axioms of $S5$ will not suffice. This is actually the case: ID-frames validate more formulae than $S5$. They validate a generalized form of the convergence axiom which reads

$$\diamond_i \square_j \varphi \supset \square_j \diamond_i, \text{ where } i, j \in A, A = \{1, \ldots, n\}, i \neq j, j, n \geq 2.$$

If we interpret the convergence axiom in epistemic terms, it states that "if agent 1 considers possible that agent 2 knows the fact p, then agent 2 knows that agent 1 considers p possible". The epistemic meaning of the axiom is even better grasped by thinking out the situations it excludes, indeed the situations in which agent 1 considers possible that agent 2 knows p, while agent 2 considers possible that agent

1 knows not p. Hence the axiom of convergence, read epistemically, "imposes a sort of homogeneity on the knowledge considered possible by other agents" [Lomuscio, 1999] and makes up one of the background assumptions which underlie dialogues between peers.

3.4   A Survey Of Knowledge Sharing Among Ideal Agents

To capture the logical properties of knowledge sharing among ideal agents, the best policy is to take the standard epistemic logic $S5_2$ ($S5$ for two agents) as our starting point and to add new axioms. The scope of the task can be easily defined from the start.

If we restrict ourselves to the language of propositional logic enriched with two symbols denoting agents, '1' and '2', and two epistemic operators, 'agent x knows' ($\Box_x$) and 'agent x considers possible' ($\Diamond_x$), an exhaustive survey of the interaction axioms which can be built with these four terms, together with the implication symbol and propositional atoms, involves 16 axioms of the form $\triangle \supset \triangle$, where $\triangle$ stands for one of the operators $\Box_1, \Box_2, \Diamond_1, \Diamond_2$, 64 axioms of the form $\triangle\triangle \supset \triangle$, or of the form $\triangle \supset \triangle\triangle$, and 256 axioms of the form $\triangle\triangle \supset \triangle\triangle$.

For all possible extensions of $S5_2$ except two of them, obtained in the way described in the first paragraph, the resulting system remains sound, complete and decidable. We need however, in most cases, to impose additional constraints on the accessibility relations.

The first axiom we shall comment on characterizes the multi-agent logic that forces the knowledge of an agent, say agent 1, to be a subset of the knowledge of another agent, say agent 2. In other words, whatever is known by 1 is also known by 2 but not conversely. That logic captures the real life situation in which agent 2 is a central processing unit which receives messages sent from independent sources represented here by agent 1. The axiom reads as follows:

$$\Box_1 p \supset \Box_2 p \,. \tag{1}$$

System $S5_2$ extended with axiom 1 is proven to be sound and complete if and only if the equivalence relation which is the model-theoretic counterpart of $\Box_2$ is a subset of the equivalence relation which is the model-theoretic counterpart of $\Box_1$. Adding axiom 1 to $S5_2$ is a genuine extension. It validates frames which are richer in this sense: it is not enough to require of the accessibility relations that they be equivalence relations. An additional condition has to be fulfilled, namely $\sim_2 \,\subseteq\, \sim_1$.

This correspondence result (in the Sahlqvist-van Benthem sense of "correspondence") is easy to understand if we bear in mind the epistemic meaning of the accessibility relation. Intuitively, agent 2 knows more than agent 1 if fewer worlds are compatible with what he knows, or to say the same thing in a different way, if the worlds he considers to be indistinguishable are members of smaller classes (smaller equivalence classes) than those held indistinguishable by agent 1. To switch back to the formal terminology again, this amounts to saying that $[w]_{\sim_2} \subseteq [w]_{\sim_1}$.

The next formula we shall be considering is:

$$\Box_1 p \supset \Diamond_2 p\,. \tag{2}$$

This formula says that if agent 1 knows that p then agent 2 thinks p possible. As opposed to axiom 1, formula 2 is not an extension but merely a theorem of system $S5_2$. It follows from $\Box_1 p \supset p$ in conjunction with $p \supset \Diamond_2 p$.

In the third category ($\triangle \supset \triangle\triangle$), axiom 3 is of special interest:

$$\Box_1 \supset \Diamond_1 \Box_2 p \tag{3}$$

It says "if agent 1 knows p, he considers possible that agent 2 also knows p". Like axiom 1, axiom 3 is a proper extension of $S5_2$. It generates a new class of frames. A new constaint is imposed upon the accessibility relations. They must satisfy the following requirement: $\forall w \exists w' : [w]_{\sim_2} \subseteq [w']_{\sim_1}$.

Axiom 3 formally captures a principle of prudence. In situations in which agents have similar characteristics, it is reasonable to assume that the other agent could reach the same conclusions by acquiring the same information from the environment and by reasoning in the same way [Lomuscio, 1999].

3.5   A Dynamic Approach To Knowledge Sharing

However successful the static approach to knowledge sharing can be, it will not suffice to handle problems like the Muddy children puzzle. To handle this problem we are forced to take up a dynamic approach to knowledge sharing.

The setting is well known. Imagine that n children are sitting in a circle and that k of them have mud on their foreheads. Each can see the spot on the others, but not the one on his or her own forehead. The teacher comes along and says: "At least one of you has mud on your head", telling something known to each of the children before he or she spoke (if $k > 1$). The teacher then asks the following question over and over again: "Can any of you prove you have mud on your head?" It can be proved that the first $k - 1$ times that the teacher asks the question, the children will all say "no" but then the kth time the children with mud will say "yes" [Barwise, 1981, p. 382].

J. Barwise gave the following informal proof by induction on k: "For $k = 1$ the result is obvious: the dirty child sees that no one else is muddy, so he or she must be the muddy one. Let us do $k = 2$. So there are just two dirty children a and b. Each answers "no" the first time, because of the mud on the other. But when b says "no", a realizes that he or she must be muddy, for otherwise b would have known the mud was on his or her head and answered "yes" the first time. Thus a answers "yes" the second time. But b goes through the same reasoning [...]" [Barwise, 1981, pp. 382-383].

This is a brand of reasoning that typically belongs to epistemic logic. The language of epistemic logic is indispensable to stating the premises and the conclusion. The proof system of epistemic logic is needed to derive the conclusion

from the premises, namely axiom K and axiom T together with the monotonicity rule $\varphi \supset \psi \vdash \mathsf{K}\varphi \supset \mathsf{K}\psi$.

As J. Geanakoplos observes, "the story is surprising because aside from the apparently innocuous remark of the teacher, the students appear to learn from nothing except their own ignorance" [Geanakoplos, 1992, p. 257]. The teacher's remark appears innocuous insofar as $k > 1$, it seems to convey no new information to the children. Yet it is crucial. It turns knowledge into common knowledge. Before the teacher spoke, every student knew that at least one of them was muddy, but this was not yet common knowledge, i.e. they did not know that everybody knew. A second crucial feature emphasized by Geanakoplos is that the pronouncements of ignorance of the children are public: "[e]ach time a student maintains his ignorance, he knows that everyone else knows he said he didn't know, etc." [Geanakoplos, 1992, 257-258]. Thirdly everyone knows the reasoning of the other.

To fully appreciate the dynamic nature of the reasoning which enables the children to solve the puzzle, we have to look at the reconstruction of that reasoning offered by the model checking method [Halpern and Vardi, 1991; Fagin *et al.*, 1995]. Roughly speaking the method consists in this. Instead of describing the puzzle by a set of formulae and using the proof theory of epistemic logic to derive the statement which is the solution of the puzzle (in this case the statement "every agent who has mud on his or her forehead knows it"), a single Kripke model $\mathbf{M}$ is constructed in which the situation is codified in such a way that one can check whether $\mathbf{M}$ verifies the above-mentioned statement. Consider the case with 3 children (a, b, c). There are eight possible combinations: three children have mud, two of them, one of them and finally none has. We can construct a cube whose eight vertices are marked with one of the triples $\langle 111 \rangle, \langle 110 \rangle, \ldots, \langle 000 \rangle$. The n-th position in the triple represents the n-th agent; "1" means that agent n has mud on his or her forehead, "0" means that he or she has a clean forehead. We have a structure with 2^n nodes, each marked with a triple of 0 and 1, such that two adjacent nodes differ in one component. Before the father speaks, child a knows whether b and c have a spot on their foreheads since he or she sees them. He or she is only ignorant about his or her own forehead. Take for instance the following case: agent b is clean and agent c is dirty. In that case, child a who sees agents b and c considers two situations possible: namely $\langle 1, 0, 1 \rangle$ (the actual situation) and $\langle 0, 0, 1 \rangle$.

In general child i has the same information in two possible worlds exactly if these two worlds agree in all components except possibly the i component. To capture this we define an accessibility relation K_i such that $(s, k) \in \mathsf{K}_i$ if and only if the worlds s and t agree in all components except possibly the ith component. The formal expression $(s, k) \in \mathsf{K}_i$ can be read as: "agent i considers worlds s and k possible" or "agent i takes s and k to be indistinguishable". The definition of K_i makes it an equivalence relation.

Before the father says "at least one of you has a dirty forehead", child a in the world $\langle 101 \rangle$ considers the situation $\langle 001 \rangle$ possible and in that world child c considers $\langle 000 \rangle$ possible. More accurately, child a thinks it possible that child

c thinks it possible that $\langle 000 \rangle$ even if child c does not. These nested epistemic operators are represented by a sequence of edges on the cube: $\langle 101 \rangle$ —a— $\langle 001 \rangle$ —c— $\langle 000 \rangle$. This sequence is not possible any longer after the father's announcement "at least one of you is muddy". The children do not acquire first-order knowledge by listening to the father's announcement. The effect of the later is to change the nature of their knowledge that there is at least one child having a muddy forehead. It turns this piece of information into common knowledge. An information update takes place. From now on, no child can think that another child thinks that nobody is muddy.

There is a very intuitive way of representing the information update effected by the father's announcement. One should simply remove the node marked with $\langle 000 \rangle$ and, by the same token, remove the edges leading to it, namely: $\langle 100 \rangle$ —a— $\langle 000 \rangle$, $\langle 010 \rangle$ —b— $\langle 000 \rangle$ and $\langle 001 \rangle$ —c— $\langle 000 \rangle$. Removing links amounts to removing epistemic possibilities. Hence truncating the cube is a way of depicting the increase of knowledge. Lomuscio calls that operation "model refinement". Let us now examine the role of the first public reply of the children to the father's question "can any of you prove that you have mud on your forehead?". If child 1 saw two clean foreheads, he would reply "Yes". But he replies "No". The same applies to children 2 and 3. From the public announcement "No" made by all of them at the same time, the three children can conclude that all the worlds containing two occurrences of "0" have to be removed. The utterance of "No" produces a mutual update, it becomes common knowledge that at least two children are muddy. Having heard the unanimous "No", each child who would see one clean forehead could conclude that he belongs to the pair of muddy children and reply "Yes" to the father. If they unanimously reply "No" for the second time, a new mutual update occurs to the effect that none of the three children sees a clean forehead. Let us now turn to the proof-theoretical approach to the Muddy children puzzle.

3.6   Dynamic Epistemic Logic And The Muddy Children Puzzle

A proof-theoretic treatment can be found in several places [Konolige, 1986; Genesereth and Nilsson, 1988; Thayse, 1989]. Genesereth and Nilsson gave a very elegant proof of the statement that answers the puzzle limited to two agents. The proof which uses a variant of the resolution method rests on these three premises:

1. $\Box_a(\neg \text{White}_a \supset \Box_b \neg \text{White}_a)$,

2. $\Box_a \Box_b(\text{White}_a \lor \text{White}_b)$,

3. $\Box_a \neg \Box_b \text{White}_b$.

The set of premises 1–3, however, lies open to a serious objection. Premise 2 happens to be false when evaluated before the father's announcement. As Gerbrandy observes, "[i]f there are two children that are dirty, it is indeed the case that each of them knows that at least one of them is dirty (they can see the other child).

But they do not know of each other that they know this. For example, child a, not knowing whether she herself is dirty, cannot be sure that b can see a dirty child (if a is clean, b sees only clear foreheads) [Gerbrandy, 1999, p. 155]". In other words, formula $\Box_a \Box_b \exists x \, \text{White}_x$ is false, hence premise 2 is false.

The standard formalization offered by static epistemic logic cannot do justice to the role of the father's announcement, nor for that matter, to the role of the denial uttered by the children. Each of these announcements performs a crucial change. It turns private knowledge into common knowledge. Standard epistemic logic however has no operators designed to handle change.

This deficiency was observed more than thirty years ago: "[h]ere is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just *one* modal operator. The only way to have any philosophically significant results in deontic logic or epistemic logic is to combine those operators with tense operators (otherwise how can you formulate principles of change?) . . ." [Scott, 1970, p. 161].[17]

Many reasonings can be analyzed without taking time into account. The derivation of the law of associativity of addition from Peano's axioms is a case in point. Here the only order that matters is logical order: some propositions logically depend upon others. This order can be contemplated *sub specie aeterni*. Temporal order does not play any logical role. In the Muddy children puzzle however, time plays a role which is obfuscated in the formalization presented above.

A more refined account of the puzzle which formally represents time and common knowledge and meets Scott's demands appears in [Gochet *et al.*, 2000, pp. 97-101], but the new formalization describes the progression of the reasoning in a staccato way. The question naturally arises whether it would not be possible to formalize the dynamic character of the information flow itself. This problem has been addressed in [Gerbrandy and Groeneveld, 1997; Gerbrandy, 1999]. Building upon the update semantics [1996], Gerbrandy and Groeneveld spelled out a formal language, an axiomatic system and a new semantics which formally represent the epistemic operation of information updating which takes place when the father makes his announcement and when the children publicly reply "No" to the father's question.

3.7   An Axiomatic System For The Update Operator

As a first step towards a formalization of information updating, a new operator is introduced into the language of epistemic logic: $[\varphi]_a \psi$, which can be rendered in natural language by "an update of the agent a's information with φ results in a situation where ψ is true". There is an operator $[\varphi]_a$ for each agent a and each sentence φ in the language. This reflects the idea that any sentence can be learned by any agent [Gerbrandy and Groeneveld, 1997, p. 150].

---

[17]The quote from [Scott, 1970] has been reprinted with the permission of Springer, The Netherlands.

If we replace the index referring to a single agent a by an index referring to a group of agents B, and write $[\varphi]_B\psi$, we obtain an operator which represents updates turning private knowledge into common knowledge. An operator such as $[\varphi]_B$ can be added for each group which is a subset of the group of agents ($B \subseteq A$). The logical force of the new operator is given by five new axioms which are added to standard epistemic logic:

(a) Axiom (a) is the analogue of the familiar normality axiom.
$[\chi]_B(\varphi \supset \psi) \supset ([\chi]_B\varphi \supset [\chi]_B\psi)$.

(b) Axiom (b) says that if it is not the case that a certain sentence is true after an update with a certain sentence, then, since the update operation always returns a unique result (it is a function), then it must be the case that the negation of that sentence is true in the updated possibility [Gerbrandy and Groeneveld, 1997, p. 156].
$\neg[\varphi]_B\psi \equiv [\varphi]_B\neg\psi$.

(c) Epistemic actions do not change the world, i.e. the current state p.
$p \equiv [\varphi]_B p$, if p is an atom.

(d) Axiom (d) says that if a knows ψ after the announcement of φ to the group B, then he must have already known that if φ were true then ψ would be true after the announcement of φ to B and conversely.
$[\varphi]_B\square_a\psi \equiv \square_a(\varphi \supset [\varphi]_B\psi)$, if $a \in B$.

(e) Axiom (e) says that a public update does not increase the information of the outsiders who do not hear the announcement.
$\square_a\varphi \equiv [\psi]_B\square_a\varphi$, if $a \notin B$.

The necessitation rule $\varphi \vdash [\psi]_B\varphi$ is also used.

From the premises

1. Every child is muddy;

2. Every child sees all the other children (Vision);

3. It is common knowledge that every child sees all the other children (Common vision).

the new axiom system of dynamic epistemic logic allows to derive the conclusion that each child knows he or she is muddy [Gerbrandy and Groeneveld, 1997, pp. 161–162].

At this point it is worth examining how the new logic can show that the father's announcement provides the second premise of Genesereth's and Nilsson's proof. From the premise Vision, we obtain:

$$\text{White}_a \lor \text{White}_b.$$

The public announcement made by the father turns the above mentioned statement into common knowledge, formally:

$$[\text{White}_a \vee \text{White}_b]_B \vdash C_B(\text{White}_a \vee \text{White}_b).$$

In virtue of the definition of the common knowledge operator C_B, knowing that $B = \{a, b\}$, we get:

$$C_B(\text{White}_a \vee \text{White}_b) \vdash \Box_a \Box_b(\text{White}_a \vee \text{White}_b).$$

Hence we can formally derive the second premise whose use was not warranted in the formalization provided by static epistemic logic.

Similar considerations can be abduced concerning the utterance of "No" (i.e. "I do not know whether I am muddy") made by the children [Gerbrandy, 1999, p. 155]. It can be shown that just as the father's announcement, the children's utterance formally represented by the update operator $[\text{No}]_B$, serves to turn private knowledge into common knowledge. We have just seen that update with the father's announcement and update by the children's denial provide missing premises that are necessary to solve the Muddy children puzzle for n children; the number of times we have to update with "No" is $n - 1$.

3.8   A Formal Semantics For Epistemic Actions

In "The Logic of Public Announcements, Common Knowledge and Private Suspicion", A. Baltag, L. Moss and S. Solecki embarked upon the challenging task of providing a formal representation for epistemic updates of various kinds [1998]. Plaza, Gerbrandy and Groeneveld had already addressed this issue for public or semi-public announcements to mutually isolated groups. Baltag *et al.* broadened the field of inquiry and covered information-updating actions of various types:

(1) information-gathering (such a learning),

(2) information-exchange (such as making public or semi-public announcement),

(3) information-hiding (lying, sending encrypted messages)

(4) losing information or misinforming.

A major innovation of these authors lies in the use of a special kind of Kripke structures for modeling actions. Two reasons may be adduced for adopting that policy: (1) actions can be seen as transitions leading from an input state to an output state. Hence it is natural to construe actions as relations, (2) epistemic actions involve a *belief component*, the agent's views of beliefs about the very action that is taking place. Just as Kripke accessibility relations can profitably be used to capture the uncertainty of each agent concerning the current *state* of a distributed system, other accessibility relations can be used to represent each agent's uncertainty concerning the current *action* taking place [Baltag, 2002, p. 3]. This insight leads to the idea that *epistemic update* can be modeled by *combining two Kripke structures* [Baltag, 2002; Moss, 2002]:

(a) an *epistemic model* $\mathbf{W} = \langle W, (\rightarrow_a^W)_{a \in \mathrm{Ag}}, V^W \rangle$ where W is a set of worlds, $(\rightarrow_a^W)_{a \in \mathrm{Ag}}$ are finitely many accessibility relations and V^W is a valuation function which assigns a set of possible worlds to each propositional atom (the set of the worlds in which it is true).

(b) an *action structure* $\mathbf{K} = \langle K, (\rightarrow_a^K)_{a \in \mathrm{Ag}}, PRE^K \rangle$ where K is a set of possible action tokens, $(\rightarrow_a^K)_{a \in \mathrm{Ag}}$ are finitely many accessibility relations and PRE is a function which assigns presuppositions or preconditions to each action token.

We shall first illustrate the notion of epistemic model and action structure. We avail ourselves of an example due to L. Moss. Consider the following scenario. A box which is closed contains a coin which either lies Heads up (H) or Tails up (T). Let us denote these two possible states by "s: H", and "t: T" respectively. Two agents A and B are present and neither of them knows which of the two states is the real one. We represent the agent's uncertainty about which of the two states is the real one by an arrow ($\leftrightarrow$). Let us assume that the coin is lying Heads up, i.e. that the truth value of the atomic sentence "H" evaluated at state s is 1. The epistemic situation can be depicted by the following epistemic model $\mathbf{W}_1$, where the real world is inserted between double parentheses:

$$\mathrm{A, B} \circlearrowright ((\mathrm{s:\ H})) \xleftrightarrow{\mathrm{A,B}} (\mathrm{t:\ T}) \circlearrowright \mathrm{A, B}.$$

Let us now suppose that agent A, unknown to B, learns that H is the case. Learning is an epistemic action which does not change the state of the world, i.e. the position of the coin lying Head up. This action will be represented by "σ: H" . While A performs this secrete action, which can be described as a kind of cheating, B does nothing, or, to use a term borrowed from [Baltag, 1999], he performed a trivial action which can happen anywhere. The trivial action's happening will be denoted by "τ: true". Agent B believes that nothing but the trivial action is taking place, belief which is expressed by an arrow ($\rightarrow$). The epistemic action which takes place (A's learning that H is the case) and B's belief that no action is taking place can be depicted by the following action structure $\mathbf{K}$:

$$\mathrm{A} \circlearrowright ((\sigma\mathrm{:\ H})) \xrightarrow{\mathrm{B}} (\tau\mathrm{:\ true})_{\mathrm{A,B}}$$

The epistemic model $\mathbf{W}_1$ is the input of the action described by the action structure $\mathbf{K}$. The output is the epistemic model $\mathbf{W}_2$ which results from combining $\mathbf{W}_1$ with $\mathbf{K}$. How should the combination of $\mathbf{W}_1$ with $\mathbf{K}$ be conceived? The combination is conceived by A. Baltag, L. Moss and S. Solecki as a kind of *product*. As the structures $\mathbf{W}_1$ and $\mathbf{K}$ involve three components, we expect to have three operations to perform. First we have to multiply the domains of the two structures. Next we have to multiply their respective accessibility relations. Finally we have to consider the valuation function and bear in mind that we are dealing with epistemic actions which do not change the world.

Let us first consider the first operation, i.e. the product of the domains W_1 of $\mathbf{W}_1$ with the domain K of $\mathbf{K}$. In the 1998 paper, Baltag *et al.* spoke of a "restricted product". Later on both A. Baltag and L. Moss conceived of this restricted product as a subset of the Cartesian product of the set of states W_1 with the set of actions K.

The Cartesian product under consideration is a set of pairs whose first member is a state and whose second member is an action. Actions have preconditions. For instance one cannot perform the epistemic action of *learning* (getting to know) p unless p is true. A pair of a state s and an action σ is an impossible pair if s violates a precondition of σ. In the Cartesian product $W_1 \times K = \{(s,\sigma), (s,\tau), (t,\sigma), (t,\tau)\}$, the third pair is an "impossible pair" since performing action σ, i.e. learning that H is the case, is impossible in state t as "H" is false at t. Hence we have first to remove the impossible pair from the Cartesian product we started with. This gives us the subset we want, namely $W_1 \otimes K = \{(s,\sigma), (s,\tau), (t,\tau)\}$. Achieving this is only the first step toward forming the *update product* $\mathbf{W}_1 \otimes \mathbf{K}$.

Next we have to multiply the second components of the structures under consideration, i.e. the accessibility relations. Such a multiplication of accessibility relations amounts to turning a pair of arrows (such as $s \to s'$ and $\sigma \to \sigma'$) into an arrow linking pairs together (such as $s, s' \to \sigma, \sigma'$). This is permitted only if the two accessibility relations are independent from one another. In L. Moss' example, the condition is fulfilled. The uncertainty about which epistemic action is going on (learning that H is the case or performing the trivial action τ) is independent of the uncertainty about which of the two atomic propositions "H" or "T" describes the real world.

The relational statement "$s, s' \to \sigma, \sigma'$" is true if and only if "$s \to s'$" and "$\sigma \to \sigma'$" are true. In the example under consideration "$s \xrightarrow{\mathrm{B}} s$" is true (i.e. the state s is indistinguishable from itself by agent B) and the statement "$\sigma \xrightarrow{\mathrm{B}} \tau$" is true. As the action σ (learning) is performed by A secretly, B does not distinguish it from the trivial action τ. Hence the following statement is true: "$(s,\sigma) \xrightarrow{\mathrm{B}} (s,\tau)$".

The third change has to do with the valuation function. Initially sentences "H" and "T" were evaluated with respect to the set of world W_1. The set of states in which "H" is true was the singleton $\{s\}$ and that in which "T" is true was the singleton $\{t\}$. Now we have to evaluate "H" and "T" with respect to $W_1 \otimes K$. The interpretation is no longer made up of states. It is made up of pairs of states. The interpretation of "H" is $\{(s,\sigma), (s,\tau)\}$. That of "T" is (t,τ).

With this model-theoretic apparatus at our disposal we can construct the update product $\mathbf{W}_1 \otimes \mathbf{K}$. After agent A has performed his hidden act of learning that H is the case, the epistemic state resulting from his epistemic action can be represented by:

$$a = ((\,(s,\sigma) : \mathrm{H}\,))$$

Agent B who, unknown to him, is in the same state of affairs as A takes two other states of affairs (b and c) as possible and as mutually indistinguishable, namely the states $b = ((s, \tau) : \text{H})$ and $c = ((t, \tau) : \text{H})$.

The absence of B under the loop starting from vertex a is worth noting. It expresses that the action of cheating induces B to have false beliefs about the world (beliefs falsifying axiom T). The philosophical lesson that these examples teach us is that, as Ladrière contends, formalization gives access to areas of meaning to which no other access is available [Ladrière, 1975, p. 241].

A. Baltag [1999; 2002] has used this kind of setting to study modified versions of the Muddy Children puzzle in which some children cheat by sending messages to tell their friends that they are dirty. He also points out that in addition to its philosophical importance the product-semantics for update has simplified his own work on the completeness and decidability of various logics proposed by authors such as J. Gerbrandy and H. van Ditmarsch [2002]. These important technical developments cannot be described here.

4   DYNAMIC DOXASTIC LOGIC

4.1   Problems Connected With The Dynamics Of Belief

V. McGee found a baffling counter-example to *modus ponens* [1985] which has been reported in [Segerberg, 1998, p. 293]. The story takes place in California on the eve of the 1980 election. We are invited to consider the following sentences, bearing in mind that Anderson and Reagan are Republicans and that Carter is a Democrat:

  (a) Anderson will win.

  (b) Carter will win.

  (c) Reagan will win.

  (d) A Republican will win.

As K. Segerberg observes, it would have been rational for a well-informed, rational agent to believe (1) and (2):

  (1) A Republican will win,

  (2) If a Republican will win, then if Reagan does not win, then Anderson will win,

but not to believe the conclusion:

  (3) If Reagan will not win, then Anderson will win.

This failure of *modus ponens* with respect to the conditional is intriguing. K. Segerberg has worked out both a proof theory and a model theory for belief revision which we shall survey in this section. Without going into details here, let us say that when we recast McGee's example into the formalism of K. Segerberg's dynamic doxastic logic, the syllogism also fails in its new guise, and the cause of its failure is displayed. There is however an interpretation of the conditional (3) as a counterfactual. Under this interpretation Mc Gee's example ceases to be a genuine counterexample to *modus ponens*.[18]

Moore's sentence is another and more familiar puzzle which can serve as a test for the explanatory power of dynamic doxastic logic. Consider the following sentence: "p but I do not believe that p" [Moore, 1912 ed 1976, p. 125] which has the form "$p \wedge \neg B_a p$". It is clearly logically odd, but several features distinguish it from typical contradictions of the form $p \wedge \neg p$. Moore's sentence can be true as opposed to sentences of the form $p \wedge \neg p$, which cannot. Moreover, as Hintikka observed, a change of speaker removes the absurdity from Moore's sentence. The sentence "p but he does not believe that p" is a perfectly natural sentence. Similarly, a change of tense removes the absurdity as shown by the sentence "He was at home but I did not believe it" [Hintikka, 1962, p. 65]. According to Hintikka, Moore's sentence does not violate consistency. Rather it violates the general presumption that the speaker believes what he or she says. Yet there is something logically wrong in it. Though Moore's sentence is not necessarily false in virtue of its logical form, as standard inconsistent sentences are, it is nevertheless "necessarily unbelievable by the speaker" [Hintikka, 1962, p. 67]. In other words, "$p \wedge \neg B_a p$", as such, is not inconsistent, but it suddenly becomes inconsistent when its presumptions are made explicit i.e. when $p \wedge \neg B_a p$ is put within the scope of B_a: "$B_a(p \wedge \neg B_a p)$". This latter sentence is genuinely inconsistent. A plain contradiction, $p \wedge \neg p$, can be derived from it with the help of doxastic logic.

Hintikka's (static) doxastic logic can account for the effect of the change of speaker in Moore's sentence. If, however, our goal is to explain the logical effect of a shift in the tense of the belief verbs, static doxastic logic does not suffice. We need dynamic doxastic logic. Before presenting the basics of dynamic doxastic logic, let us first show what can be achieved with static doxastic logic.

4.2 *A Formalization Of Moore's Paradox In Static Doxastic Logic*

We shall write down an axiomatic proof showing that $B_a(p \wedge \neg B_a p)$ is inconsistent. It is carried out within the modal calculus $KD4$. We shall use the standard box and diamond operators of modal logic for ease of reading and switch to the doxastic operator only at the last line. Whenever a formula is obtained in applying a principle of the propositional calculus (such as, for instance, syllogism), we write PC on the right of the formula together with the number(s) of the lines from which the formula is derived. The principles used are familiar and we leave the task of identifying them to the reader. It will be clear that besides principles of

---

[18]We owe this observation to J. Halpern. See also [Levi, 1996, pp. 109–111].

the propositional calculus, we also make essential use of principles of the modal calculus K and of axiom 4 and axiom D.

1.	$\Box(\varphi \wedge \psi) \supset (\Box\varphi \wedge \Box\psi)$	K
2.	$\Box(p \wedge \neg\Box p) \supset (\Box p \wedge \Box\neg\Box p)$	$1, \mathrm{Subst.}$
3.	$\Box(p \wedge \neg\Box p) \supset \Box p$	$2, \mathrm{PC}$
4.	$\Box(p \wedge \neg\Box p) \supset \Box\neg\Box p$	$2, \mathrm{PC}$
5.	$\Box\varphi \supset \Box\Box\varphi$	$\mathrm{Ax.}\ 4$
6.	$\Box p \supset \Box\Box p$	$5, \mathrm{Subst.}$
7.	$\Box(p \wedge \neg\Box p) \supset \Box\Box p$	$3, 6, \mathrm{PC}$
8.	$\Box\varphi \supset \Diamond\varphi$	$\mathrm{Ax.}\ D$
9.	$\Box\Box p \supset \Diamond\Box p$	$8, \mathrm{Subst.}$
10.	$\Box(p \wedge \neg\Box p) \supset \Diamond\Box p$	$7, 9, \mathrm{PC}$
11.	$\Box(p \wedge \neg\Box p) \supset \neg\Diamond\Box p$	$4, \mathrm{Df.}\ \Diamond$
12.	$\Box(p \wedge \neg\Box p) \supset (\Diamond\Box p \wedge \neg\Diamond\Box p)$	$10, 11, \mathrm{PC}$
13.	$(\varphi \wedge \neg\varphi) \supset (\psi \wedge \neg\psi)$	PC
14.	$(\Diamond\Box p \wedge \neg\Diamond\Box p) \supset (p \wedge \neg p)$	$13, \mathrm{Subst.}$
15.	$\Box(p \wedge \neg\Box p) \supset (p \wedge \neg p)$	$12, 14, \mathrm{PC}$
16.	$\mathsf{B}_a(p \wedge \neg\mathsf{B}_a p) \supset \bot$	15

4.3   From The Theory Of Belief Revision To Dynamic Doxastic Logic

In 1985, Alchourrón, Gärdenfors and Makinson developed a theory of belief revision ("AGM") which has become classic [1985]; it investigates the rationality constraints that can be imposed upon belief changes. They describe belief changes with three basic operations performed on belief sets: expansion, contraction and revision. Expansion with formula φ consists in forming the union of the prior belief set X with $\{\varphi\}$ and taking the closure of $X \cup \{\varphi\}$ under the operation of classical consequence Cn. Contraction consists in deleting φ in X together with all formulas that imply φ so that the result is logically closed. Revision consists in adding φ to X in such a way that the resulting set $X * \varphi$ is consistent [Wassermann, 1999, p. 429]. In 1994, M. de Rijke showed that the axioms of expansion and revision of AGM can be translated into the object language of dynamic modal logic and be proved valid in well-founded DML-models [1994]. The same year, K. Segerberg showed how to recast the theory of belief change in terms of modal logic.

In 1995, B. van Linder, W. van der Hoek and J.-J. Ch. Meyer initiated a new approach to belief change. They focused on the actions which an agent performs to bring about belief changes. A new way of distinguishing knowledge and belief emerges: what distinguishes knowledge from belief is not only the static property of veridicality, but also the dynamic property of non-defeasibility. The knowledge of agents concerning propositional formulas is immune to change: "it persists under the execution of belief-changing actions [such as expansion, contraction and revision]" [1995, p. 111].

Also in 1995, K. Segerberg undertook to bridge the gap between Hintikka's style of (static) doxastic logic and the AGM style of belief revision theory by creating

a new kind of logic which he dubbed "dynamic doxastic logic"(DDL). The latter
was meant to serve the interest both of logic, whose coverage was increased, and
of belief revision itself. As Segerberg puts it, "[t]he great advantage of recasting
belief revision theory as dynamic doxastic logic is that it puts at our disposal
the rich meta-theory developed in the study of modal and dynamic logic" [1999,
p. 142].[19] We shall first examine how K. Segerberg brought together the language
of standard doxastic logic and that of [Alchourrón *et al.*, 1985] into the unified
language of DDL.

4.4   The Language Of Dynamic Doxastic Logic

Standard doxastic logic using Hintikka's operator B distinguishes between the
following statements:

$$\begin{array}{ll} \mathsf{B}\varphi & \text{the agent believes that } \varphi \\ \neg\mathsf{B}\varphi & \text{the agent does not believe that } \varphi \\ \mathsf{B}\neg\varphi & \text{the agent believes that not } \varphi \\ \neg\mathsf{B}\neg\varphi & \text{the agent does not believe that not } \varphi \end{array}$$

As K. Segerberg observes, AGM is able to capture the above combinations of belief
operators and negations. It does it however in a different way:

$$\begin{array}{ll} \varphi \in T & \varphi \text{ is in the agent's belief-set } T, \\ \varphi \notin T & \varphi \text{ is not in the agent's belief-set } T, \\ \neg\varphi \in T & \neg\varphi \text{ is in the agent's belief-set } T, \\ \neg\varphi \notin T & \neg\varphi \text{ is not in the agent's belief-set } T. \end{array}$$

The AGM language can also express operations on beliefs which have no equivalent
in standard doxastic logic: expansion, revision, contraction, respectively expressed
by the terms '+', '∗' and '−':

$$\begin{array}{ll} \chi \in T + \varphi & \chi \text{ is in the agent's belief set expanded by } \varphi, \\ \chi \in T * \varphi & \chi \text{ is in the agent's belief set revised by } \varphi, \\ \chi \in T - \varphi & \chi \text{ is in the agent's belief set contracted by } \varphi. \end{array}$$

If we aim at representing doxastic actions such as expansion, revision and con-
traction in doxastic logic, we need new operators. K. Segerberg borrows operators
$[\alpha]$ and $\langle\alpha\rangle$, familiar in dynamic logic, but he gives them a new interpretation.
Generally speaking, when a formula φ is prefixed by $[\alpha]$ or $\langle\alpha\rangle$, it means "after
every way (respectively some way) of performing α, it is the case that φ".

   These dynamic operators can be turned into dynamic doxastic operators if we
restrict actions to doxastic actions (actions on our beliefs as opposed to actions
on the world). Three new operators (and their duals) are obtained. The "box-
operators" capture the three basic operations of belief change: $[+]$, $[*]$ and $[-]$.

---

[19]The quote from [Segerberg, 1999] has been reprinted with the permission of Springer, The
Netherlands.

With these new dynamic operators available, the three basic statements of the theory of belief revision are rendered in this way:

$$[+\varphi]\mathsf{B}\chi\,;\ [*\varphi]\mathsf{B}\chi\,;\ [-\varphi]\mathsf{B}\chi\,.$$

To sum up, the language of doxastic dynamic logic contains three sets of logical constants:

1. Boolean connectives;

2. the static doxastic operator B;

3. the dynamic doxastic operators $[+], [*], [-]$.

B is a box-operator which takes Boolean formulas only as arguments. Its dual, the diamond operator b, can be introduced by definition:

$$b\chi\ =_{def}\ \neg\mathsf{B}\neg\chi\,.$$

4.5   A New Account Of Belief Change

Whoever intends to recast belief revision theory as dynamic doxastic logic is expected to begin by translating the axioms of belief revision theory into doxastic logic and then to build up an appropriate model for the interpretation of the new logic. K. Segerberg did not make that move immediately. The reason was that he wanted first to modify classical theory of belief revision on several crucial points.

Classical theory of belief revision describes belief changes (expansion, contraction, revision) as transitions between theories. The latter are conceived as belief sets closed under logical consequences. K. Segerberg objects that such a view can never do full justice to the agent's doxastic dispositions.

An agent confronted with a piece of information which clashes with his current beliefs usually has several ways of modifying his belief set (i.e. theory), or, to use Lindström's and Rabinowicz's terminology, several "fall back positions" to retreat back to by revising the initial theory. Knowing the doxastic state of an agent requires more than knowing his or her current belief set, it also requires knowing how he or she would respond to new information about the world.

What emerges out of this is a conception according to which belief changes are not moves taking us from one belief set to another, but moves taking us from one belief state to another. A belief state is a "belief set *cum* dispositions for belief change". If we want to accommodate doxastic dispositions, we have to complicate our conceptual apparatus. For that purpose, Segerberg introduces the technical concept of hypertheories into the semantics of doxastic logic. The word "hypertheory" was first used by Grove [1988]. Let us notice that the word is used in a set-theoretic sense.

Adopting Grove's picture, Segerberg sees hypertheories as concentric spheres. The central sphere represents the initial hypertheory, namely the current belief state of the agent. Each of the successive shells that we go through when we

move away from the central sphere toward the periphery represents a theory (a "fallback" in Lindström's and Rabinowicz's terminology) back to which the agent may retreat when forced to revise his or her initial theory by deleting propositions. Deletion is not a random process. We should first delete propositions which are less "entrenched".

Entrenchment is an epistemic notion which was initially defined for propositions: "Intuitively, a proposition α is at least as entrenched in the agent's belief set as another proposition β if and only if the following holds: provided the agent would have to revise his beliefs so as to falsify the conjunction $\alpha \wedge \beta$, he should do it in such a way as to allow for the falsity of β" [Lindström and Rabinowicz, 1999, p. 356]. It is important to keep in mind that the spheres are "nested". They are linearly ordered by the relation of inclusion. The ordering of spheres represents the entrenchment ordering of theories. The successive shells around the initial sphere represent fallbacks containing propositions which are less and less entrenched as we move away from the central sphere toward the periphery.

The first difference between AGM and Segerberg's account of belief revision lies in this: in AGM, doxastic actions are defined on theories. In Segerberg's formulation, they are defined on hypertheories.

A second difference between AGM and Segerberg's account of belief revision is that Segerberg, as opposed to Lewis and Grove, does not require the spheres to be linearly ordered by set inclusion. This important change in belief revision was first proposed by S. Lindström and W. Rabinowicz to account for the fact that there may be several equally reasonable revisions of a theory. A third difference between AGM and Segerberg's treatment, connected with the previous one, lies in the nature of contraction. In AGM, contraction, like expansion, is functional. Segerberg's model theory (like that of Lindström and Rabinowicz) allows contraction to be merely relational.

4.6   A Semantics For Dynamic Doxastic Logic

- $\|p\| = \mathbf{V}(p)$: the intension of p is the set of possible worlds at which atom p is true;

- $\|\neg P\| = \|\mathbf{U} - P\|$;

- $\|P \vee Q\| = \|P\| \cup \|Q\|$ and so on for the remaining connectives.

Next we list all the clauses of a recursive definition of truth in a model $\mathbf{M}$ for a given hypertheory $\mathbf{H}$, a given belief set $\mathbf{X}$, at a given state u.

- $\mathbf{H}, \mathbf{X}, u \models \varphi$ iff $u \in \|\varphi\|$, if φ is a purely Boolean formula.

- $\mathbf{H}, \mathbf{X}, u \models \varphi \wedge \psi$ iff $\mathbf{H}, \mathbf{X}, u \models \varphi$ and $\mathbf{H}, \mathbf{X}, u \models \psi$.

- $\mathbf{H}, \mathbf{X}, u \models \neg\varphi$ iff not $\mathbf{H}, \mathbf{X}, u \models \varphi$.

- $\mathbf{H}, \mathbf{X}, u \models B\varphi$ iff $\mathbf{X} \subseteq \|\varphi\|$.

The more interesting clause deals with doxastic actions (expansion, contraction):

- $\mathbf{H}, \mathbf{X}, u \models [\alpha]\varphi$ iff $\forall \mathbf{H}' \forall \mathbf{X}' \forall u'(\langle \mathbf{H}, \mathbf{X}, u \rangle R^{\alpha} \langle \mathbf{H}', \mathbf{X}', u' \rangle \supset \mathbf{H}', \mathbf{X}', u' \models \phi)$,
  where $[\alpha]$ can be either $[+]$ or $[-]$ and R^{α} can be either R^{+} or R^{-} respectively.

Expansion and contraction by a proposition P may be identified with the relations [Segerberg, 1997, p. 189]:

$$R^{+P} = \{(H, H') : H' \backslash P\}$$
$$R^{-P} = \{(H, H') : \exists Z \, (Z \text{ is minimal in } H \cap (\mathbf{U} - P) \, \& \, H' = H \mid Z\}$$

R^{+P} is the augmentation of hypertheory $\mathbf{H}$ by proposition P (i.e. by a set of possible worlds). Augmentation is functional. R^{-P} is restriction of hypertheory $\mathbf{H}$ by proposition P. Restriction is relational. A formula is valid in a frame for dynamic doxastic logic if it is true in all models on the frame relative to all hypertheories ($\mathbf{H}$), to all belief-sets ($\mathbf{X}$) and states (u) in the universe $\mathbf{U}$.

As we have seen, K. Segerberg disagrees with the identification of belief sets with belief states on the ground that two agents may hold identical beliefs about what the actual state of the world is like and yet react differently to new information. To drive the point home, it is worth examining a law of doxastic dynamic logic, namely the law of recovery, whose validity conditions differ depending on whether it applies to belief-sets or to belief-states. The conditions for the law of recovery to hold are more demanding in the second case.

Consider a given belief-set to which we successively apply the restriction (R^{-}) and augmentation (R^{+}). This amounts to taking the relative product of those two relations. Do we recover the initial belief set at the end of the process? The answer is "yes", if the belief sets ($\mathbf{X}$) to which we successively apply contraction ($[-]$) and expansion ($[+]$) are the same. Next consider a given belief-state to which we successively apply restriction and augmentation. To recover the initial belief state, two conditions instead of one must be fulfilled: the belief-sets($\mathbf{X}$) and also the hypertheories ($\mathbf{H}$) must be the same. There are cases, however, in which the first condition is fulfilled and the second is not. The following example is a case in point: We start with hyper-theory $\mathbf{H}$ and apply restriction. We get $\mathbf{H}'$. The fact that the result of restriction is not uniquely defined (restriction is not functional) does not prevent us from forming the relative product of restriction $\mathbf{H}R^{-}\mathbf{H}'$ and augmentation $\mathbf{H}'R^{+}\mathbf{H}$. The formal schema of relative product, i.e. $\exists z(xR^{-}z \wedge zR^{+}y)$ can be instantiated even if there are two z i.e. $\mathbf{H}'_1$ and $\mathbf{H}'_2$. What is ruled out when there are two $\mathbf{H}'$, however, is the possibility of recovery. With $\mathbf{H}R^{-}\mathbf{H}'_1$ and $\mathbf{H}'_2 R^{+}\mathbf{H}''$, there is no possibility of recovery since there is no "common middle term", due to our choice of different $\mathbf{H}'$.

4.7   Axiomatic Systems Of Dynamic Doxastic Logic

K. Segerberg spelled out several axiom systems [1995; 1998; 1999] and proved that they are sound and complete for the class of intended models. We shall restrict ourselves to presenting and commenting upon the most fundamental axioms of

the first axiomatic system of DDL [Segerberg, 1995], just hinting at what can be found in the others.

The language of the system presented in 1995 contains:

1. Boolean connectives;

2. the operator B which can take as arguments Boolean formulas only;

3. the operators [+] and [−].

The system of axioms contains nine fundamental axiom schemata and rules:

1. Three $K-$ schemas of the form $\bigcirc(\varphi \supset \psi) \supset (\bigcirc\varphi \supset \bigcirc\psi)$ in which $\bigcirc$ can be B, [+] and [−];

2. the rule of *modus ponens*;

3. Three necessitation rules: If $\vdash \varphi$ then $\vdash \bigcirc\varphi$ in which again $\bigcirc$ can be B, [+] and [−];

4. Two rules of congruence:
   If $\vdash \varphi \equiv \psi$ then $\vdash [+\varphi]\chi \equiv [+\psi]\chi$,
   If $\vdash \varphi \equiv \psi$ then $\vdash [-\varphi]\chi \equiv [-\psi]\chi$;

5. The replacement rule which states that the set of valid formulas is closed under replacement of provably equivalent valid formulas.

There are, however, restrictions to the law of uniform substitution. To the nine rules and axioms described above, sixteen others are added, most of which characterize the interplay between the static doxastic operator B and the dynamic doxastic operators [+] and [−]. Let us mention these two examples:

1. $\vdash B\varphi \supset (\chi \equiv [+\varphi]\chi)$,
   which says that if you already believe something, accepting it changes nothing;

2. $\vdash \neg B\varphi \supset (\chi \equiv [-\varphi]\chi)$,
   which says that if you do not believe something, removing it changes nothing.

4.8   From Basic To Full Dynamic Doxastic Logic

In basic DDL,

1. the belief operator B only takes Boolean operators as formulas;

2. the formulas which can be substituted for φ in the doxastic dynamic operators, $[+\varphi]$, $[-\varphi]$ and $[*\varphi]$, must also be Boolean.

These limitations have been removed by S. Lindström and W. Rabinowicz in the so-called "full dynamic doxastic logic". In this logic, no restriction is imposed upon the substituends of φ and ψ in formulas of the form $[\alpha\varphi]B\psi$. From now on, we are allowed to apply doxastic actions (expansion, contraction and revision) not only to propositions about the world, but also to introspective propositions.

This increase in expressive power has unexpected consequences which are worth examining. Consider the success lemma

$$[*(\varphi)]B\varphi$$

which says that after revising his belief set with φ the agent believes φ.

If we substitute an atomic proposition or a Boolean proposition for φ nothing strange happens. Let us however substitute Moore's sentence for φ. We get a sentence which has the logical form of formula F.

$$F : [*(\varphi \wedge \neg B\varphi)]B(\varphi \wedge \neg B\varphi)].$$

This says that after revising my beliefs with $\varphi \wedge \neg B\varphi$ which is a consistent proposition, I obtain $B(\varphi \wedge \neg B\varphi)$ which is not. Does this mean that the success lemma is not sound? Lindström and Rabinowicz suggest another answer. Formula F is *ambiguous*. It allows two readings:

1. After revising my beliefs with "p is the case and I do not believe it", I believe that (before the revision), p was the case and that I did not believe it.

2. After revising my beliefs with "p is the case and I do not believe it", I believe that (after the revision), p is the case and that I do not believe it.

To capture these two readings we have to draw a distinction familiar to tense logicians [Reichenbach, 1947; Gabbay, 1974; Cresswell, 1991], i.e. the distinction between the time at which a sentence is evaluated (evaluation time) and the time at which the event reported took place (reference time).

That important distinction should be expressed in our symbolism. This requires that we enrich the formal language of dynamic doxastic logic with an appropriate tense operator and that we work out a more refined semantics. Without going into details, let us note that a tensed dynamic doxastic logic is precisely what we need to account for the fact that a shift of tense — just as a shift of speaker — in Moore's sentence removes its paradoxicality. As opposed to "I believe that my neighbour is at home and that I do not believe it", the sentence "I believe that my neighbour was at home and that I did not believe it" is immune to the charge of inconsistency [Gochet, 2004].

5   FIRST ORDER EPISTEMIC LOGIC

5.1   *Problems Raised By Quine About Quantified Modal Logic*

The invention of formal systems of quantified modal logic goes back to 1946. Two foundational papers appeared in succession in the *Journal of Symbolic Logic*: "A

Functional Calculus of First order Based on Strict Implication" by Ruth C. Barcan [1946] and "Modalities and Quantification" by Rudolf Carnap [1946]. As G. Corsi notes [2001], even before functional calculi were born, in "Notes on existence and necessity" of 1943, W. V. O. Quine pointed out that modal contexts resist two classical laws of first-order logic with identity:

(a) the principle of substitutivity which states that "given a true statement of identity, one of its two terms may be substituted for the other in any true statement and the result will be true";

(b) the law of existential generalization which licenses the derivation of 'There is an x such that x is φ' from 't is φ'.

Quine pursued his criticism in "The problem of interpreting modal logic". His "Reference and Modality" grew out of these two papers and was published in his *From a Logical Point of View* in [1953; 2nd ed. 1961].

Consider the following inference, which is a modified version of a puzzle raised by Aristotle in *De Sophisticis Elenchis* 24 (179b 1–3) [Føllesdal, 1967, p. 4]:

Philip is unaware that Tully denounced Catiline,
Tully = Cicero,
Therefore Philip is unaware that Cicero denounced Catiline.

A state of affairs in which Philip has only a moderate acquaintance with Roman history would make the premises true and the conclusion false and shows that, in the scope of epistemic terms like 'is unaware', 'knows', 'believes', the principle of substitutivity breaks down.

Consider this application of existential generalization:

Philip is unaware that Tully denounced Catiline,
Therefore something is such that Philip is unaware that it denounced Catiline.

Here again there is a problem. As Quine asks: "[w]hat is this object, that denounced Catiline without Philip's having become aware of the fact? Tully, that is, Cicero? To suppose this would conflict with the fact that 'Philip is unaware that Cicero denounced Catiline' is false" [Quine, 1953; 2nd ed. 1961, p. 147].

5.2   Hintikka's Solution

Knowledge and Belief [1962] was the first systematic effort to provide both a proof theory and a semantics for the verbs of propositional attitudes 'a knows' and 'a believes' taken as logical operators on a par with the necessity operator familiar since C.I. Lewis' work. Hintikka introduced the notions of model sets, model systems and alternativeness relations.

A model set is a formal counterpart of the informal idea of a partial description of a possible state of affairs. It is incumbent upon the logician to lay down conditions

that sentences must fulfill to be admitted into any such a set. Consider a model set μ containing the sentence "$P_a p$" which says that it is possible, for all that the person referred to by the term "a" knows, that p. The last statement "can be true only if there is a possible state of affairs in which p would be true: but this state of affairs need not be identical with the one in which the statement was made. A description of such a state of affairs will be called an alternative to μ with respect to a" [Hintikka, 1962, p. 42]. In order to show that a given set of sentences is true in at least one world in which the agent follows the consequences of what he or she knows, we have to consider a set of model sets. Such sets of model sets are called "model systems".

In *Knowledge and Belief* and more extensively in subsequent works, Hintikka came to grips with Quine's objections. He agrees that the two inferences mentioned in section 5.1 are invalid as they stand. But he claims that validity can be restored. He sees those inferences as enthymemes.

To turn the inference:

> Philip knows (believes) that Cicero denounced Catiline.
> Cicero = Tully.
> Therefore Philip knows (believes) that Tully denounced Catiline.

into a valid inference we have to supply the auxiliary premise: Philip knows (believes) that (Cicero = Tully).

Let us now consider the problem of *existential* generalization across modal operators. Here again Hintikka has a solution to offer to Quine's problem. If one of the occurrences of a singular term is buried under n_i layers of modal operators, we are speaking of its several references in the possible worlds described by all the different alternatives, n_i times removed, to the description of the actual one. Taking this into account we should ceased to be puzzled by our inability to generalize with respect to such a singular term. To restore the laws of first-order logic, it suffices to bring in auxiliary premises: if there are no iterations of modal operators and if only one modal operator is present, then simple statements can be found to express explicitly that a term specifies a well-defined individual, namely: $(\exists x)\mathsf{K}_b(a = x)$, $(\exists x)\mathsf{B}_b(a = x)$, etc. [Hintikka, 1972, p. 403].

5.3   The Problem Of Hybrid Contexts And Nested Operators

To appreciate the explanatory power of Hintikka's auxiliary premises fully, we have to revisit the solution offered by Frege to the problem raised by the failure of the substitutivity principle in belief contexts and to examine a difficulty which passed unnoticed, but which can be solved with the help of Hintikka's techniques.

In *Über Sinn und Bedeutung*, Frege explained the semantic difference between the informative identity statement 'The morning star is the evening star' and the trivial identity statement 'The morning star is the morning star'. For that purpose he distinguished between sense and reference (or *nominatum*). The definite description (to use Russell's term) 'the morning star' has the same reference as,

but a different sense than, 'the evening star' and this is true also of the sentences which contain them.

Having firmly established the sense-reference distinction, Frege brought it to bear on a quite different issue: the problem raised by the failure of the principle of substitutivity of identity in belief contexts and indirect discourse. He claimed that when we move from direct to indirect discourse, a shift of reference and sense occurs. "In indirect (oblique) discourse", Frege says, "we speak of the sense, e.g. of the words of someone else. From this it becomes clear that also in indirect discourse words do not have their customary nominata; they name what customarily would be their sense" [1949, p. 87].

Frege's account of the sense and reference in indirect discourse (and belief contexts) aroused strong opposition. "If we could recover our pre-Fregean semantic innocence" Davidson writes, "it would seem to us plainly incredible that the words 'The earth moves', uttered after the words 'Galileo said that', mean anything different, or refer to anything else, than is their wont when they come in different environments" [Davidson, 1968 9, p. 144].[20]

Davidson's objection to Frege strikes us as well taken. In indirect discourse, the words of a reported speech do not differ in meaning from the same words in direct speech. Clearly the difference lies elsewhere. The speaker who reports somebody else's speech may not endorse the truth of the proposition he reports. He may refrain from accepting the existence presuppositions that the proposition carries. But as far as the meaning and reference are concerned, they are the same in both direct and indirect speech. That was not so clear for Frege because he conceived of sentences as names, an issue that does not concern us here.

We shall drive this point home by examining a kind of inference which cannot be accounted for at all if we adopt Frege's semantics of indirect discourse, but which raises no difficulty if we stick to the received view about indirect speech recalled by Davidson.

Consider the following inference:

> The morning star is a planet
> and it is known (believed, said) that the morning star is a planet.
> Hence there is an x such that x is a planet
> and it is known (believed, said) that x is a planet.

On Frege's construal, the use of two occurrences of the same variable x is illegitimate. In the first occurrence after 'such that', the variable x takes individuals as values, in the second, it takes individual concepts, i.e. senses, as values. There is a kind of equivocation here.

On Hintikka's construal, no hidden equivocation is in the offing. In the two occurrences, the variable takes individuals as values. Yet the inference is not unconditionally valid. It is valid with the proviso that the appropriate auxiliary premise be true. Taking a as proxy for the definite description 'the morning star',

---

[20]The quote from [Davidson, 1968 9] has been reprinted with the permission of Springer, The Netherlands.

the inference, strengthened by the appropriate auxiliary premise, can be formalized in this way:

Premise: $\quad\quad\quad\quad\quad P(a) \wedge \mathsf{B}_b(P(a))$

Auxiliary premise: $\quad \exists\, x\, (x = a \wedge \mathsf{B}_b(x = a))$

Conclusion: $\quad\quad\quad\, \exists\, x\, (P(x) \wedge \mathsf{B}_b(P(x)))$

The technique of auxiliary premises can be extended to specify the precise conditions under which a singular term a obeys the usual laws of instantiation and generalization. If we are considering (1) the actual world, (2) what b believes, (3) that a knows, we are considering not only the real world but also "epistemic d-alternatives to doxastic b-alternatives" and the requisite auxiliary premise for this complex case of nested modal operators is [Hintikka, 1972, p. 407][21]:

$$\exists\, x\, [(a = x) \wedge (b \text{ believes that } (a = x)) \wedge (b \text{ believes that } d \text{ knows that } (a = x))]\,.$$

5.4 Føllesdal's Defense Of Quantified Epistemic Logic

Quine has criticized first-order epistemic logic on two scores:

1. the law of substitutivity of identity fails;

2. the law of existential generalization fails.

J. Hintikka has shown how to vindicate first-order epistemic logic. He made some concessions however. The pure law of substitutivity of identity is abandoned and existential generalization is sound only under the proviso that auxiliary premises are introduced.

D. Føllesdal addressed the same issues but solves Quine's problems in a different manner. As opposed to Hintikka, Føllesdal leaves the core of quantification theory unaffected but (a) he imposes strong restrictions on the vocabulary of singular terms which can be used when we apply the principle of substitutivity of identity or the law of existential generalization, and (b), he modifies the modal part of the axiom system. We shall deal with (a) only. For (b) the reader is referred to [Føllesdal, 1967].

Føllesdal considers the situation in which to one and the same individual b in the actual world, there are two individuals b' and b'' which correspond in the world w' compatible with the agent's knowledge as shown in the diagram below:

w (Non actual world compatible with the agent's knowledge): b' $\quad\quad\quad\quad$ b''

w' (Actual world): $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ b

Such a situation, Føllesdal finds intolerable. He agrees with Quine that the existential quantifier cannot be used meaningfully in the situation just described where the bound variable does not refer to a unique individual in the actual world

---

[21]The quotes from [Hintikka, 1972] are reprinted with the permission of Springer, The Netherlands.

w and in its epistemic alternative w'. But, contrary to Quine, he does not disallow quantification into epistemic contexts once and for all. What he does instead is to restrict quantifying into epistemic contexts. Quantifying into epistemic context is permitted only when "the variables keep their reference as we pass from one epistemically possible world to the next [Føllesdal, 1967, p. 11]". In the same way, to insure substitutivity of identity, we have to restrict our vocabulary of singular terms to "genuine singular terms", i.e. to singular terms which keep their reference in all epistemically possible worlds [Føllesdal, 1967, p. 17]. Føllesdal's concept of genuine singular term is identical with the concept of rigid designator which occupies a central position in the causal theory of naming spelled out by Kripke in his essay *Naming and Necessity* [1980]. Føllesdal's policy is *sufficient* to remove the unwanted inferences. It may be objected however that it is too *restrictive*. Rigid designators pick up the same individuals across all possible worlds, including the actual one, but, as P. Jackson and H. Reichgelt observe, we wish to be able to apply the rules of quantified modal logic also to individuals that do not exist in the real world [1989, p. 197].

Rules of inference for quantified modal logic operating with non rigid designators (designators indexed on world's names) can be found in several proof systems such as the method of *suffixed tableaux* [Ramsay, 1988, pp. 140–143], the method of *resolution* and *unification* [Jackson and Reichgelt, 1989, pp. 197–208] and the method of *prefixed tableaux* [Fitting and Mendelsohn, 1998, pp. 118–121].

5.5   The Epistemic Readings Of The Barcan Formula And Its Converse

The distinction between *de re* and *de dicto* constructions in epistemic contexts goes back to Aristotle. In *Analytica Priora* (II, 21, 67 a 16ff), Aristotle observes that to know of every triangle that it has its angles equal to two right angles is ambiguous between two senses. These two senses can be captured in natural language by constructions (1) and (2):

(1) Agent a knows that every triangle has angles equal to two right angles.

(2) Of every triangle agent a knows that it has angles equal to two right angles.

Føllesdal renders these two senses in the formal language by (3) and (4):

(3) $\mathsf{K}_a \forall x \varphi(x)$;

(4) $\forall x \mathsf{K}_a \varphi(x)$.

If we insert the connective of implication between (4) and (3) we obtain the epistemic reading of the Barcan formula:

(5) $\forall x \mathsf{K}_a \varphi(x) \supset \mathsf{K}_a \forall x \varphi(x)$.

If we insert the connective of implication between (3) and (4) we obtain the converse of the Barcan formula already mentioned in section 1.2:

(6) $\mathsf{K}_a \forall x \varphi(x) \supset \forall x \mathsf{K}_a \varphi(x)$.

On Føllesdal's reading, which is the most natural one, both are invalid.

On Hintikka's reading, the epistemic reading of the converse of Barcan's formula, far from being invalid, is a truism. The reason for this is given by Føllesdal. On Hintikka's approach (4) does not read 'of each object' agent a knows that it is φ' but 'of each object which a knows', a knows that it is φ [Føllesdal, 1967, p. 23].

The question whether we should accept or dismiss the Barcan formula and its converse cannot be settled simply by considering natural language. Deeper issues are involved which come to the foreground only if we take the trouble of building a fully fledged formal semantics for modal and epistemic logic.

5.6 A Kripke's Semantics For Epistemic Logic

In 1963, Kripke developed a semantics for quantified modal logic of necessity which has become standard. We shall first sum it up and then recast Hintikka's semantics of first-order epistemic logic into the terminology of Kripke's semantics.

A quantificational model structure (today we say "frame") is a quadruple:

$$\langle \mathbf{W}, \mathbf{w}_0, \mathbf{R}, \psi(\mathbf{w}_i) \rangle$$

in which $\mathbf{W}$ is a set of possible worlds, $\mathbf{w}_0$ is a designated world (the real world), $\mathbf{R}$ is the accessibility relation defined on $\mathbf{W} \times \mathbf{W}$, $\psi(\mathbf{w}_i)$ is a function which assigns a domain to each world.

We obtain a quantificational model by adding a valuation function $\mathbf{V}(P^n, w_i)$ to the quantificational model structure. A technical point should be stressed here. In Kripke's quantificational model, the domain of variation of the free variables differs from that of the bound variables. Free individual variables take their values in the union of the domains ($\mathbf{U}$). On the contrary, the bound variables take their values in the domain of the particular world w_i at which the formula is evaluated. Hence $\mathbf{V}(\forall x A(x, y_1, \ldots, y_n), w_i) = \mathbf{T}$ relative to an assignment of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ to $y_1, \ldots, y_n$ (where the $\mathbf{b}_i$ are elements of $\mathbf{U}$) if $\mathbf{V}(A(x, y_1, \ldots, y_n), w_i) = \mathbf{T}$ for every assignment of $\mathbf{a}, \mathbf{b}_1, \ldots, \mathbf{b}_n$ to $x, y_1, \ldots, y_n$ respectively, where $\mathbf{a} \in \psi(\mathbf{w}_i)$; otherwise we have $\mathbf{V}(\forall x A(x, y_1, \ldots, y_n), w_i) = \mathbf{F}$ [Kripke, 1963, p. 85]. Furthermore, $\mathbf{V}(\mathsf{K}\varphi, w_i) = \mathbf{T}$ if and only if $\mathbf{V}(\varphi, w_j) = \mathbf{T}$ in all worlds w_j accessible from w_i. Reading "the possible world w_j is accessible from the possible world w_i" as "the model set μ_j is an epistemic alternative to the model set μ_i", one recovers Hintikka's terminology.

One should notice that Kripke's semantics does not only accommodate possible worlds, it also allows domains of possible worlds to include possible individuals. Kripke writes: "We must associate with each world a domain of individuals, the individuals that exist in that world... [i]n worlds other than the real one, some actually existing individuals may be absent, while new individuals like Pegasus may appear" [Kripke, 1963, p. 85]. This will raise problems which will be investigated in section 6.

Four kinds of Kripkean semantics are usually distinguished: semantics with constant domains, semantics with decreasing domains, semantics with increasing domains and semantics with variable domains. The first kind is too restrictive for it does not permit non existent individuals to inhabit possible worlds [Pietarinen, 2003]. The last ones are the more general and less committal. This fourfold distinction matters when we are concerned with the validity conditions of the Barcan formula (BF) and its converse (CBF). For the Barcan formula to be valid in a Kripke model, a special condition has to be fulfilled: the domains of the world related by the accessibility relation have to be decreasing. Formally: $w_i R w_j \supset (Dw_j \subseteq Dw_i)$.

A. Nerode and R. Shore are wary of this constraint. The limit of the decreasing domain is the empty domain. If all objects cease to exist, "we have entirely left the realm of classical predicate logic which is formulated only for nonempty domains" [Nerode and Shore, 1993, p. 211]. They are willing to subscribe however to the converse of the Barcan formula (CBF): $\Box \forall x P x \supset \forall x \Box P x$.

Nevertheless CBF, however uncontroversial it may look, is not valid in Kripke's semantics unless we assume that the domains are increasing: $w_i R w_j \supset Dw_i \subseteq Dw_j$. This interaction between properties of the domains and properties of the accessibility relations is an undesirable feature of Kripke's semantics which deprives it of full generality. An analogous complaint can be expressed concerning the failure of CBF in Kripke's semantics when no constraint is made on the accessibility relation. This failure is not intrinsic to CBF itself. As G. Corsi shows, it is due to a peculiar feature of Kripke's possible world semantics. In Kripke's semantics, the domain of variation of the quantifiers is, in general, a proper subset of the domain of variation of the free variables. Alternative semantics have been put forward to remove these unwanted interactions between domains and accessibility relations [Gillet, 2000; Corsi, 2001].

The universal quantifiers occurring in the Barcan formula and its converse can be understood as free of existential assumption. Let us call *free quantified* BF and CBF the Barcan formulas interpreted in this way.

5.7 *Ghilardi's And Routley's Formulas*

Besides Barcan formula and its converse, two very similar formulas have attracted the attention of modal logicians, namely Ghilardi's formula (GF) and its converse (CGF):

$$\exists x \Box P x \supset \Box \exists x P x \,,$$
$$\Box \exists x P x \supset \exists x \Box P x \,.$$

In his book of 1962, but not in later work, Hintikka [1966] upholds the epistemic version of the first formula:

$$\exists x \mathsf{K} P x \supset \mathsf{K} \exists x P x \,,$$

and supplies a formal proof of it within his system. Commenting on it he says:

"[i]ntuitively, the self sustenance of

$$\exists x \mathsf{K} Px \supset \mathsf{K} \exists x Px$$

is not surprising. What it says is that if you know *who* does something you *ipso facto* know *that* someone does it" [Hintikka, 1962, p. 160]. R. Moore argues that the difference between $\exists x \mathsf{K}_a Px$ and $\mathsf{K}_a \exists x Px$ amounts "to a difference in the relative scopes of an existential and a universal quantifier [the 'every' in 'every possible world compatible with . . . ']" [Moore, 1995, p. 41]. The difference, he claims, can be expressed as the difference between $\exists x \forall w S(x,w)$ and $\forall w \exists x S(x,w)$. As the first formula entails the second (but not conversely) in first-order logic, Moore concludes that "[t]he possible world analysis [. . .] implies that we should be able to infer 'Ralph knows that there is a spy' from 'There is someone Ralph knows to be a spy' as indeed we can (Moore, Ibid.)". W. Lenzen, however, derived the unacceptable consequence (4) from sentences (1–3) where (2) is an instance of $\exists x \mathsf{K}_a Px \supset \mathsf{K}_a \exists x Px$ [Lenzen, 1976, p. 59]:

(1) $\exists x \, \mathsf{K}_a (x = \text{Pegasus})$
(2) $\exists x \, \mathsf{K}_a (x = \text{Pegasus}) \supset \mathsf{K}_a (\exists x \, (x = \text{Pegasus}))$
(3) $\mathsf{K}_a \exists x \, (x = \text{Pegasus}) \supset \exists x \, (x = \text{Pegasus})$
(4)   Therefore $\exists x \, (x = \text{Pegasus})$.

If we share Hintikka's later view that "there seems to be a perfectly good sense of knowing who a certain person is which does not commit one to holding that the person in question is known to exist" ([Hintikka, 1966, p. 4] quoted by Lenzen [Lenzen, 1976, p. 59]), premise (1) in the above inference causes no problem. Applying the epistemic version of GF together with axiom T, we derive the false statement $\exists x \, (x = \text{Pegasus})$. Since T cannot be disallowed, the culprit must be (2), i.e. the principle GF. Lenzen's objection nothwistanding, the GF principle continues to be upheld by many authors, even in its doxastic version. There is an account of quantifying-in for which, assuming that a can do existential generalization, this is valid: "[i]f a believes there is some particular object satisfying φ, then it certainly believes there is *some* object satisfying φ" [Genesereth and Nilsson, 1988, p. 218]. How can we settle the issue?

New light is shed on this problem by counterpart semantics as introduced by [Corsi, 2001]. Within the framework of that semantics, which is briefly described in section 7, the GF formula is shown to follow from a formula which syntactically captures the semantic assumption that the counterpart relation (see section 10) is everywhere defined. The only logical axioms and rules used in the proof belong to classical first-order logic and to modal propositional logic K [Corsi, 2001, pp. 21, 26]. The very need for this hidden extra-premise, which is revealed by counterpart semantics, highlights the richness of first-order modal logic and shows that the interplay between quantifiers and modal operators brings about something radically new, which cannot be found either in non modal first-order logic or in propositional modal logic.

As opposed to GF formula which, as we have seen, is still defended by several authors, its converse — called "Routley's formula" by Slater — is generally discarded as invalid. Yet, as we shall see, it can be vindicated under particular conditions. Consider this instance of Routley formula ('[HP]' stands for 'Hercule Poirot knows' and 'M' for 'is a murderer'): $[HP]\exists\, x Mx \supset \exists\, x[HP]Mx$.

Slater holds that in so far as verbs of attitude ("to know", "to believe", "to think of") are construed as relations, the validity of the Routley's formula can be defended under some proviso. If the antecedent is true, there is an object of Hercule Poirot's knowledge, belief or thought, but the object can only be captured by a purely referential term such as "$\varepsilon x Mx$" [Slater, 1994, pp. 40–43]. The idea is this: if I know that there is a spy, whoever he or she may be, there is an entity which I apprehend under the highly neutral predicate "known to be a spy by me". The truth conditions of the two sentences are the same.

5.8   The Necessity Of Identity And Its Epistemic Analogue

In quantified modal logic, it is easy to prove: $\forall\, x \,\forall\, y\,(x = y \supset \Box x = y)$. The proof rests upon the two axioms of identity and the rules of uniform substitution, necessitation and *modus ponens*. (See the details in section 5.10.) No move seems questionable. If, however, we instantiate the bound variables by definite descriptions we obtain blatant counter examples:

The morning star = the evening star	$\supset$	$\Box$(The morning star = the evening star),
The morning star = the evening star	$\supset$	K(The morning star = the evening star).

Even if we prohibit instantiation, the conclusion that all identities are necessary as such might be felt unacceptable. It is certainly so if we identify "necessary" with "logically true", but we need not do that. All we have to accept at this stage is that (1) $x = x$ is logically true, (2) $\Box x = x$ is merely true (but not logically true), (3) that if $x = y$ is true, it is necessarily so, but (3) is not a logical truth [Fitting and Mendelsohn, 1998, p. 146]. What emerges from this is that the necessity at stake when we say that identities are necessary identities (formally that: $\forall\, x \,\forall\, y\,(x = y \supset \Box x = y)$) is not logical necessity.

A technical characterization of the crucial difference between truth, logical truth and necessary truth in the metaphysical sense of necessity has been given by [Cocchiarella, 1984] and [Rivenc and de Rouilhan, 1997]. The truth of a wff in a model (indexed by a language suitable to that wff) is, as usual, the satisfaction of the wff by every assignment in the universe of the model. Logical truth is then truth in every model (indexed by any appropriate language) [Cocchiarella, 1984, p. 312]. Logical necessity captured by Carnap in the 1946 paper should be carefully distinguished from the kind of necessity captured by Kripke semantics in 1963.

Kripke's semantics, as opposed to Carnap's semantics, allows the quantificational interpretation of necessity in the metalanguage to refer not to *all* the possible worlds (models) of a given logical space but only to those in a given non-empty set of such worlds [Cocchiarella, 1984, p. 315]. This restriction enabled Kripke to

succeed where Carnap had failed. Kripke managed to prove the completeness of his system of quantified modal logic, but this result was obtained at a price. Kripke succeeded in capturing various forms of metaphysical necessity, but not logical necessity as Carnap did in the incomplete (and incompletable) system that he laid down in his paper of [1946].

The formula $\forall x \forall y (x = y \supset \Box x = y)$ is a theorem of one of the first two systems of quantified modal logic [Marcus, 1946]. Does it hold for epistemic logic? The laws of epistemic modalities (such as K or B) sometimes diverge from those of metaphysical modalities like $\Box$. An example due to L. Carlson brings out the specificity of epistemic logic very well. As L. Carlson observes, the following informal sentence sounds like a description of a possible epistemic situation [1988, p. 237]: "There is someone who might be two different people as far as the police knows". Its simplest natural formalization is $\exists x \exists y \, (x = y \land \neg \mathsf{K}x = y)$ which is the negation of $\forall x \forall y \, (x = y \supset \mathsf{K}x = y)$, i.e. the negation of the epistemic reading of $\forall x \forall y \, (x = y \supset \Box x = y)$. To understand what is going on here, it is not enough to paraphrase sentences into logical formulas. We need a rigorous formal semantics. The latter, however, cannot get off the ground if we do not first remove some difficulties connected with the notion of possible individuals.

5.9   Possible Individuals

In a paper first published in 1948, Quine put forward a famous argument against the postulation of possible individuals: "Take, for instance, the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? (...) How many of them are alike?" [1953; 2nd ed. 1961, p. 4]. What Quine finds objectionable is that possible men lack a criterion of identification. This concern for identity criteria would later give rise to Quine's famous slogan "No entity without identity".

One might try to meet Quine's demands for identification criteria by affirming that the identification of individuals across the boundaries of possible worlds rests upon continuity properties similar to those enabling us to trace the continuous world lines of an individual in space-time.

Quine does not accept this reply: "These considerations cannot", Quine maintained, "be extended across the worlds, because you can change anything to anything by easy stages through some connecting series of possible worlds. The devastating difference is that the series of momentary cross-sections of our real world is uniquely imposed on us, for better or for worse, whereas all manner of paths of continuous gradation from one possible world to another are free for the thinking up" [1981, p. 127].

There is however another answer to Quine's objection. One might say that the denizens of possible worlds which we need to bring in for making sense of quantified modal logic are not possible individuals but real individuals considered in a possible scenario. For instance we could imagine a world in which Richard

Nixon existed, but never won presidential elections. If we take that line, the problem of providing criteria of identification does not arise. We have already identified our man and we hold him constant across alternative scenarii.

This way of tracing an individual across possible worlds requires that we "tag" — to use R. Barcan Marcus' word — the individual with a proper name, not with a definite description. Proper names (like "Franklin") as opposed to definite descriptions (like "the inventor of bifocals") designate the same object in all the worlds in which this object exists. Using a terminology introduced by Kripke we can say that the former are rigid designators and the latter accidental designators.

Kripke's solution is suitable for the counterfactuals. It does not seem to suit the situation of the police who mistakes one person for two different persons. Here it seems that we have to work with two worlds: the real world containing the individual sought by the police and a possible world compatible with what the police knows in which there are two different individuals that correspond to the same individual in the real world [See Føllesdal's diagram in section 5.4]. An alternative account of possible worlds is called for.

Hintikka has developed a conception of individual as a function or, to use his favorite metaphor, "a world-line" that picks out from several possible worlds a member of their respective domains as the referent of a singular term. Such a function may be partial. This happens when a well-defined individual existing in one world fails to exist in another. A partial individuating function may fail to have a value in the actual world. A function of that kind is what counts as a possible individual for Hintikka [1972, p. 403]. An individuating function can also be ill-defined. This happens when different individuals in different possible worlds are associated with the same singular term. Here we are getting close to a solution to Carlson's puzzle.

It remains to recast Hintikka's insights in formal terms. This has been done by Carlson [1988, pp. 244–245]. A Kripkean model $\mathbf{M}$ for Hintikka's epistemic logic is a quintuple $\langle \mathbf{W}, \mathbf{D}, \mathbf{F}, \mathbf{R}, \mathbf{V} \rangle$ such that $\mathbf{W}$ is a set of epistemic alternatives ("possible worlds"), $\mathbf{D}$ is a set of individuals belonging to the union of the domains, $\mathbf{D}_w$ of each world $\mathbf{w}$, $\mathbf{F}$ is a set of partial individuating functions $\mathbf{f}$ defined on $\mathbf{W}$ such that $\mathbf{f}(\mathbf{w})$ is in $\mathbf{D_w}$ whenever $\mathbf{f}$ is defined at $\mathbf{w}$ and these conditions are fulfilled: if the individual $\mathbf{d}$ is member of the domain $\mathbf{D}$, then $\mathbf{d} = \mathbf{f}(\mathbf{w})$ (a) for at least and (b) for at most one $\mathbf{f}$ in $\mathbf{F}$, i.e. the individuals that can be named are possible values of bound variables and neither split nor merge. $\mathbf{R_d}$ is a subset of $\mathbf{W} \times \mathbf{W}$. $\mathbf{V}$ is a valuation function.

With that semantical apparatus, it is possible to model the predicament of the police in the example given above. The interpretation of $\exists x \, \exists y \, (x = y \wedge \neg \mathsf{K} x = y)$ reads as follows: x is the same individual as y in the real world, but for some worlds compatible with what the police knows, x and y have different values. Formally speaking $\mathbf{V}(x, w_0) = \mathbf{V}(y, w_0)$, but for some w_1 such that $w_0 R w_1$, $\mathbf{V}(x, w_1) \neq \mathbf{V}(y, w_1)$. In other words, the function $\mathbf{f}$ is ill-defined since it takes as a value an individual which splits when we move from the actual world to at least one of its epistemic alternatives.

We can also explain the role of the auxiliary premise $\exists\, x\, \mathsf{K}_a(x = b)$, i.e. 'The agent knows who b is' inserted in the rule of existential generalization for epistemic logics mentioned in sections 5.2 and 5.3. These premises are added to guarantee that the individuals that are possible values of our bound variables are well-defined.

5.10   Counterpart Semantics

Quine criticized the notion of possible individuals from an epistemological point of view. They lack criteria of identification. R. Barcan Marcus criticized them from an ontological point of view. Possible individuals cannot be related by a relation, be the relation that of identity or another relation. The question as to whether a possible individual is identical or not to another one, she claims, does not make sense for in that case there are no individual objects, which are what is needed for an identity relation [Marcus, 1993]. She adds that the converse of Quine's slogan "No entity without identity" also holds: "No identity, no entity" [1993, p. 208].

A way out would be to renounce talking about cross-world identity and to embrace some version of counterpart theory. Counterpart theory was invented by D. Lewis in 1968. A. Hazen showed that a model theory could be extracted out of it [Hazen, 1979]. In a counterpart semantics the domains of the different worlds are disjoint, but an individual of one world may have a counterpart in the domain of another.

The definition of the counterpart relation reads as follows: "an individual a existing at world w satisfies at w the formula $\Box P(x)$ iff every counterpart $a*$ of a in any accessible world v, satisfies $P(x)$" [Corsi, 2001, p. 11]. Counterpart relations as opposed to the identity relation allow individuals to split or to merge when we move from one world to another. Hence counterparthood is more general that identity, although the former contains the latter as a particular case. Now we have a very natural solution to Carlson's puzzle which does not force us to have recourse to possible individuals. We can say that a real individual has two counterparts in the police's thought. The main achievement of counterpart semantics however lies elsewhere. It lies in its ability to get to the roots of the problem raised by the formula $\forall\, x\, \forall\, y\,(x = y \supset \Box x = y)$. Before we substantiate our claim, a distinction drawn by Ghilardi and Meloni has to be introduced. Ghilardi and Meloni have refined the syntax of modal language by introducing a distinction between the arity of predicates and the arity of formulas (which they call "type"). In the formula $\forall\, x\, \forall\, y\, \forall\, z\,((xRy \wedge yRz) \supset xRz)$, each predicate has arity 2, but the whole formula has type 3. Observe that a sequence which satisfies $(xRy \wedge yRz) \supset xRz$ needs three individuals at least. (Sequences with only one or two individuals also satisfy it, but make it trivial.) Their aim is to control the free variables which occur in the formulas and to bring to the fore the combination of the operation of substitution and modality which are responsible for the anomalies in first-order modal logic. Relying on their findings which were expressed in the language of category theory and developing their insights further, G. Corsi reformulated their diagnosis of the faulty step in the standard proof of $\forall\, x\, \forall\, y\,(x = y \supset \Box x = y)$. In

accordance with a common practice in logic for computer science, the following conventions are adopted. Let $\langle m : t_1, \ldots, t_n \rangle$ be a term of type m and A be a formula. By $\langle m : t_1, \ldots, t_n \rangle A$, we denote "the formula of type m obtained by applying the operation of substitution to the formula A of type n and the complex term $\langle m : t_1, \ldots, t_n \rangle$ of type $m \to n$" [Corsi, 2001, p. 15].

To start with, we take Leibniz' law

(1) $x = y \supset \varphi x \supset \varphi y$,

and we substitute $\square(x = \hat{u})$ for φ where $\hat{u}$ designates the free variable to the right of φ in (1), i.e. respectively x and y. We obtain:

(2) $x = y \supset [\square(x = x) \supset \square(x = y)]$;

Applying the law $[A \supset (B \supset C)] \supset [B \supset (A \supset C)]$, we get:

(3) $\square(x = x) \supset [(x = y) \supset \square(x = y)]$;

(4) $x = x$; (Axiom of identity)

(5) $\square(x = x)$; (Necessitation applied to (4))

The theorem follows by *modus ponens* on (3) and (5). Let us adopt G. Corsi's notation and redo the proof. We obtain (2′) by substituting $\langle 2 : x, x \rangle \square(x = y)$ for φx and $\langle 2 : x, y \rangle \square(x = y)$ for φy:

(2′) $x = y \supset \langle 2 : x, x \rangle \square(x = y) \supset \langle 2 : x, y \rangle \square(x = y)$.

By the law of permutation of the antecedents, we get:

(3′) $\langle 2 : x, x \rangle \square(x = y) \supset (x = y \supset \langle 2 : x, y \rangle \square(x = y))$.

Now using the axiom of identity, and then necessitation, we obtain:

(4′) $\langle 2 : x, x \rangle x = y$.

(5′) $\square \langle 2 : x, x \rangle x = y$.

We try to build a *modus ponens* with (3′) and (5′) as premises, but we fail. In (3′) the necessity operator was there before substituting x for y in $\square(x = y)$. In (5′) the necessity operator was applied after the substitution of x to y. The two premises instantiate the forms p' and $p \supset q$ instead of p and $p \supset q$. Hence, as G. Corsi observes, we do not have a *modus ponens*. Can we remove the p/p' equivocation? We cannot. As G. Corsi shows, substitution and modal operators do not commute as opposed to substitution and connectives or substitution and quantifiers. Here again we see that there is more to quantified modal logic than what can be found in first-order logic alone, or in propositional logic alone.

From G. Corsi's proof (implicit in Ghilardi and Meloni) that the derivation of formula $\forall x \forall y (x = y \supset \square x = y)$ is faulty we cannot, however, conclude that the

formula itself should be discarded. Although it has to be given up as a theorem, it is still eligible as an axiom. In the last but one section of this chapter we shall argue that there are independent reasons for taking that line. The recognition that identity statements between proper names are necessary is the key to the solution of several recalcitrant puzzles. Hence we have good grounds for accepting the controverted formula as an axiom with the proviso that only proper names be allowed as substituends.

5.11   Kripke's Puzzle About Belief

In 1979, S. Kripke examined afresh the problem raised by the apparent failure of co-designative names (like 'Cicero' and 'Tully') to be interchangeable in belief contexts. He spelled out a puzzle about co-designative terms in a belief context, which arises even if no principle of substitutivity is invoked.

Kripke imagines a young monolingual Frenchman, Pierre, who has heard and read about London's being pretty and who assents to the sentence "Londres est jolie". Later he emigrates to England, settles down in an unattractive part of London with uneducated inhabitants. As none of his neighbours know any French, he has to learn English by exposure. In particular everyone speaks of the city 'London' as being where they live. Pierre's surroundings being unattractive, he is inclined to assent to the English sentence: "London is not pretty".

It looks as though Pierre has contradictory beliefs. But, Kripke observes, "it is clear that Pierre, as long as he is unaware that the cities he calls 'London' and 'Londres' are one and the same, is in no position to see, by logic alone, that at least one of his beliefs must be false. He lacks information, not logical acumen. He cannot be convicted of inconsistency" [1979, p. 257].[22]

A candid reader would say that though Pierre's beliefs are implicitly inconsistent, they can be made consistent on the proviso that they become explicit and undergo revision. The problem is that we need a formal semantics which accounts for these many-sided facts. R. Parikh's approach satisfies these requirements.

As we have to deal with an impossible state of affairs (described by the bilingual sentence 'Londres est jolie $\wedge$ London is not pretty'), the standard possible world semantics does not suffice. Parikh introduces the word 'scenario' "to mean a complete theory in Pierre's new language, consistent with his beliefs, but which [...] might not be a possible world in our sense" [Parikh, 2001, p. 386]. Possible worlds are a special kind of scenarii, namely scenarii held possible by what R. Parikh calls "the community theory" or "our theory".[23]

The contrast between scenarii and possible worlds hinges on a difference of status between our beliefs and Pierre's beliefs. The reason why we call our scenarii

---

[22]The quote from [Kripke, 1979] has been reprinted with the permission of Springer, The Netherlands.

[23]The quote from [Parikh, 2001] appears with permission from CSLI Publications. Copyright 2001 by CSLI Publications, Stanford University, Stanford, CA 94305-4101.

possible worlds is that we are in a Moore's paradox situation: we cannot say that something is true, but that we do not believe it.

Taking advantage of the distinction between scenarii and possible worlds, R. Parikh distinguishes between two kinds of inconsistencies which he calls 1-inconsistency and 2-inconsistency. The terms "1-inconsistency" and "2-inconsistency" apply to theories, more precisely to "complete theories", and only derivatively to beliefs. A complete theory T is defined as a theory which for any closed formula A contains either A or its negation. Moreover T contains the axiom of first-order modal logic $S5$ and is closed under *modus ponens*, universal generalization and necessitation. It also contains $a = b \supset \Box(a = b)$ as an axiom.

An individual theory T_i is defined as 1-inconsistent if there are no scenarii among its complete extensions. This happens whenever all formulas are theorems of T_i, namely whenever the distinction between formulas and theorems collapses. The occurrence of $\varphi \wedge \neg\varphi$ in the theory is a sufficient (but not a necessary) condition to produce this effect in virtue of the theorem $\varphi \wedge \neg\varphi \supset \psi$ (*ex falso sequitur quodlibet*).

An individual theory T_i is defined as 2-inconsistent if there are no possible worlds among the extensions of $T_i \cup T_c^{\Box}$ where $T_c^{\Box}$ is the set of all formulas which are considered in the community theory T_c to be necessary.

The notions of 1-belief and 2-belief are introduced next. They denote respectively (1) somebody's belief seen from his or her own standpoint and (2) his or her belief seen from our standpoint. R. Parikh defines them in this way: Pierre 1-believes A if $A \in T_p$. Pierre 2-believes A if A can be proved in T_p together with appropriate formulae in $T_c^{\Box}$, where $T_c^{\Box} = \{A : \Box A \in T_c\}$. Applying these definitions we shall see that Pierre's beliefs are 2-inconsistent but 1-consistent. Pierre's complete theory T_p contains four scenarii described by the following sentences:

(1)   Jolie(Londres) $\wedge \neg$Pretty(London) ,
(2)   Jolie(Londres) $\wedge$ Pretty(London) ,
(3) $\neg$Jolie (Londres) $\wedge$ Pretty(London) ,
(4) $\neg$Jolie (Londres) $\wedge \neg$Pretty(London) .

Theory T_p is 1-consistent since some of its scenarii are possible worlds, namely (2) and (4). But at the same time it is 2-inconsistent. To show this we have first to consider what Pierre 2-believes. For that purpose we look at the necessary statements of the community theory T_c. It contains the statement 'London = Londres' from which we derive '$\Box$ (London = Londres)' by applying axiom $a = b \supset \Box(a = b)$, which encapsulates R. Barcan Marcus' and Kripke's idea that identities between names are necessary. '$\Box$ (London = Londres)' qualifies for membership in $T_c^{\Box}$. Hence to know what Pierre 2-believes we have to include 'London = Londres' in theory T_p. Since T_p is logically closed, we easily get the blatant inconsistency 'Jolie(Londres) $\wedge \neg$ Jolie(Londres)' modulo the translation of 'pretty' into 'jolie'.

R. Parikh's semantics explains how Pierre is inconsistent from our point of view but consistent from his. It also does something which other solutions of Kripke's puzzle fail to do, i.e. it provides a unified account of the puzzles about proper

names. For instance it can also deal with the problem raised by Philip's ignorance that Tully is Cicero. Quine's early puzzle about Philip differs from Kripke's puzzle about Pierre. The concepts defined in the theory of belief revision enable us to bring out the difference. To perform the inference on which he is stuck Philip has to expand his knowledge of Roman history. To circumvent the contradiction which threatens him, Pierre has to revise his beliefs about geography.

6   LOGICAL OMNISCIENCE AND EPISTEMIC LOGIC

6.1   Various Forms Of Logical Omniscience

The concepts of knowledge and belief analyzed in Hintikka's foundational book are highly idealized. Logical omniscience is built into his logic of the two notions. This immediately prompted the criticism that his "senses of 'knowledge' and 'belief' are much too strong [. . .] since most people do not know every proposition entailed by what they know; indeed many people do not even understand all deductions from premises they know to be true" [Castañeda, 1964, p. 134].

Castañeda's criticism notwithstanding, [Fagin *et al.*, 1995] still maintain that the system $KT45$, which is open to the same criticisms as Hintikka's logic, is one of the very best formalisms for epistemic logic, at least if the applications of epistemic logic are our main concern. System $KT45$ however displays the seven forms of logical omniscience listed below:

(1)   If $\vdash \varphi$ then $\vdash \mathsf{K}\varphi$ (closure under theoremhood);
(2)   If $\vdash \varphi \supset \psi$ then $\vdash \mathsf{K}\varphi \supset \mathsf{K}\psi$ (closure under logical implication);
(3)   If $\vdash \varphi \equiv \psi$ then $\vdash \mathsf{K}\varphi \equiv \mathsf{K}\psi$ (closure under logical equivalence);
(4) $\vdash \mathsf{K}(\varphi \supset \psi) \supset (\mathsf{K}\varphi \supset \mathsf{K}\psi)$ (closure under material implication);
(5) $\vdash \mathsf{K}(\varphi \equiv \psi) \supset (\mathsf{K}\varphi \equiv \mathsf{K}\psi)$ (closure under material equivalence);
(6) $\vdash (\mathsf{K}\varphi \wedge \mathsf{K}\psi) \supset \mathsf{K}(\varphi \wedge \psi)$ (closure under conjunction);
(7) $\vdash \mathsf{K}(\varphi \wedge \psi) \supset (\mathsf{K}\varphi \wedge \mathsf{K}\psi)$ (closure under simplification).

It is worth stressing that what is at stake in (1) is not factual omniscience, but logical omniscience. The necessitation rule does not say that from a proposition's being true we are entitled to derive that it is known. It says that from a proposition's being a theorem, or being valid, we are entitled to derive that it is known.

6.2   Belief, A Borderline Concept Between Logic And Psychology

Hintikka fully recognizes the idealized nature of his account of knowledge and belief, but he questions the very possibility of giving a characterization of human logical competence in purely logical terms. Considering the rules (2) and (3) he writes that what "causes the breakdown of these rules is broadly speaking the fact that one cannot usually see all the logical consequences of what one knows or believes". Hintikka adds that "it may seem completely impossible to draw a line between the implications one sees and those one does not see by means

of general logical considerations alone. A genius might readily see quite distant consequences while another man may almost literally 'fail to put two and two together'" [Hintikka, 1970, p. 36].

As Hintikka observes, the extent to which one follows the logical consequences of what one believes varies with one's mood, training and degree of concentration. But these limits of our logical insights are both ephemeral and idiosyncratic.They seem to fall outside of logic.

Later on, Levesque turned the distinction between a logical and a psychological account of knowledge into a distinction between implicit and explicit knowledge and addressed the issue raised by Hintikka in logical terms. Fagin and Halpern took a further step forward and developed a logic of awareness which was briefly described in section 2.9. (For other uses of the concept of awareness in connection with the problem of logical omniscience, see [Huang and Kwast, 1991; Thijsse, 1991; 1992].)

The conception of explicit knowledge proposed by Halpern and Fagin is open to at least three criticisms. First, it rests upon a "sentence storage model" of awareness which ignores the process of getting access to knowledge. Second, it is too fine-grained. It is possible for an agent to be aware of $\varphi \vee \psi$ without being aware of $\psi \vee \varphi$, i.e. he might have explicit knowledge of the first formula and lack explicit knowledge of the second. Hence we have a formalization of logical blindness rather than a logic of limited logical competence. Third, the system contains the following version of the necessitation law: "From $\vdash \varphi$ infer $\vdash \mathsf{A}\varphi \supset \mathsf{B}\varphi$" which formally captures the idea that as soon as an agent is aware of a tautology he or she believes it explicitly.

This is again unrealistic. We may be aware of a complicated tautology without knowing that it is a tautology. This leads Fagin *et al.* to admit that the axioms of the system under consideration "do not give us much insight into the properties of explicit knowledge" [Fagin *et al.*, 1995, p. 340]. Meanwhile E. Gillet put forward a characterization of the logical competence of rational agents in terms of their ability to uncover several layers in the logical structure of an argument. In his account, "analysis functions", as he calls them, play the role played by awareness in Fagin's and Halpern's formalism. They enable E. Gillet to circumvent the various forms of logical omniscience without attributing logical blindness to the agent [Gillet and Gochet, 1993].

6.3   A Logic For Occurrent Beliefs Free Of Logical Omniscience

Not all of the seven forms of logical omniscience listed in the first section are equally unacceptable. Three of them are really unwanted: closure under theoremhood, closure under logical implication and closure under logical equivalence. They are unacceptable insofar as they impute an infinite capacity to the agent. The other forms of logical omniscience, however, might be defended. Closure under conjunction captures the ability to put two things together. Closure under material implication captures the ability of an agent to practice *modus ponens*.

What we are after is a logic which avoids the unacceptable forms of logical omniscience, but which also accounts for the agent's rationality in a principled way. Not only *modus ponens*, conjunction and simplification should be safeguarded, but also other familiar rules and principles such as hypothetical syllogism, disjunctive syllogism, *modus tollens* etc. A. Wiśniewski tackled this problem in "Two logics for occurrent belief" [1998]. The first of these logics, called $S.0$, is a weak modal logic containing only one rule of inference: *modus ponens*. Necessitation cannot be used and the rule of extensionality (law of exchange of equivalents) is not allowed within the scope of the doxastic operator B. Although the author axiomatizes two doxastic concepts (believing and admitting), we shall restrict ourselves to examining the first one. For that purpose we shall slightly change the axiom system. The change is inconsequential. Axioms of $S.0$ are tautologies of the classical propositional language enriched with the following forms:

$$\text{Ax.}\,1: \quad \mathsf{B}(\varphi \supset \psi) \supset (\mathsf{B}\varphi \supset \mathsf{B}\psi)$$
$$\text{Ax.}\,2: \quad \mathsf{B}(\varphi \wedge \psi) \supset (\mathsf{B}\varphi \wedge \mathsf{B}\varphi)$$
$$\text{Ax.}\,3: \quad (\mathsf{B}\varphi \wedge \mathsf{B}\psi) \supset \mathsf{B}(\varphi \wedge \psi)$$
$$\text{Ax.}\,4: \quad \mathsf{B}\varphi \supset \neg\mathsf{B}\neg\varphi$$

A four-valued matrix is provided by A. Wiśniewski. It has been borrowed from Lukasiewicz's modal system L. Its values are 1, 2, 3, 0. The designated value is 1. The truth-functions that correspond to the operators $\neg, \wedge, \vee, \supset, \mathsf{B}$ are spelled out in tables. We only give the table for the doxastic operator B:

α	$\mathsf{B}\alpha$
1	2
2	2
3	0
0	0

Every axiom of the system receives the designated value 1 and *modus ponens* carries it from the premises to the conclusion. Hence every thesis of the system has value 1. From the truth-function associated with the operator B, it is clear that no formula of the form $\mathsf{B}\varphi$ has the value 1. We cannot infer $\vdash \mathsf{B}\varphi$ from $\vdash \varphi$. Hence the analogue of the rule of necessitation is not derivable in $S.0$ [Wiśniewski, 1998, p. 117].

To see the point of $S.0$, it is worth comparing the following two versions of the principle of hypothetical syllogism:

(1) $[(\mathsf{B}\varphi \supset \mathsf{B}\psi) \wedge (\mathsf{B}\psi \supset \mathsf{B}\chi)] \supset (\mathsf{B}\varphi \supset \mathsf{B}\chi)$.
(2) $[\mathsf{B}(\varphi \supset \psi) \wedge \mathsf{B}(\psi \supset \chi)] \supset (\mathsf{B}\varphi \supset \mathsf{B}\chi)$.

Theorem (1) is a trivial instance of hypothetical syllogism. It is external in so far as no analytic power is ascribed to the believer. We reason about the believer's beliefs but we do not capture the believer's own reasoning. Theorem (2), on the contrary, is not trivial. It is internal. We impersonate the believer and reconstruct his or her reasoning from within.

The derivation of (2) from (1) in $S.0$ is easy:

(1)	$\mathsf{B}(\varphi \supset \psi) \supset (\mathsf{B}\varphi \supset \mathsf{B}\psi)$	Ax. 4
(2)	$\mathsf{B}(\psi \supset \chi) \supset (\mathsf{B}\psi \supset \mathsf{B}\chi)$	Ax. 4
(3)	$[(p \supset q) \wedge (r \supset s)] \supset [(p \wedge r) \supset (q \wedge s)]$	Classical prop. calculus
(4)	$[\mathsf{B}(\varphi \supset \psi) \wedge \mathsf{B}(\psi \supset \chi)] \supset [(\mathsf{B}\varphi \supset \mathsf{B}\psi) \wedge (\mathsf{B}\psi \supset \mathsf{B}\chi)]$	
	Subst. in (3): $\mathsf{B}(\varphi \supset \psi)/p \ldots \mathsf{B}\psi \supset \mathsf{B}\chi/s$; *modus ponens* (1) $\wedge$ (2), (3).	
(5)	$[(\mathsf{B}\varphi \supset \mathsf{B}\psi) \wedge (\mathsf{B}\psi \supset \mathsf{B}\chi)] \supset (\mathsf{B}\varphi \supset \mathsf{B}\chi)$	Classical hyp. syl.
(6)	$[\mathsf{B}(\varphi \supset \psi) \wedge \mathsf{B}(\psi \supset \chi)] \supset (\mathsf{B}\varphi \supset \mathsf{B}\chi)$	(4) $\wedge$ (5) hyp. syl.

The agent's rationality also manifests itself in not believing patent contradictions. This is captured in $S.0$ by the theorem $\neg\mathsf{B}(\varphi \wedge \neg\varphi)$. The proof based on ax. 2, ax. 4 and the principle of non-contradiction is trivial. Observe that from the agent refraining from believing a contradiction, it does not follow that he believes the principle of contradiction itself. Formally $\neg\mathsf{B}(\varphi \wedge \neg\varphi) \nvdash \mathsf{B}\neg(\varphi \wedge \neg\varphi)$.

The idea of distinguishing between the agent's reasoning and the observer's reasoning was systematically used for the first time in Konolige's *Deduction Model of Belief* [1986, p. 55]. Konolige presents a tableau system which allows nested auxiliary tableaux to occur inside main tableaux. An auxiliary tableau represents the internal proof process of the agent, as opposed to the original tableau (the main tableau) which is the external observer's view of the agent.

Wiśniewski's system $S.1$ shares its axioms with $S.0$. It differs from it in so far as it contains the following rule of definitional replacement (R):

Rule R: From $\varphi =_{def} \psi$ and context C, derive $C[\varphi/\psi]$.

The definitions of $\wedge$ and $\vee$ are:

$$\varphi \wedge \psi =_{def\ 1} \neg(\varphi \supset \neg\psi),$$
$$\varphi \vee \psi =_{def\ 2} \neg\varphi \supset \psi.$$

Observe that the adoption of R does not commit us to closure under material equivalence. Replacement of materially equivalent formulas in the scope of belief operator remains prohibited.

System $S.1$ captures the agent's ability to use disjunctive syllogism:

(1)	$\mathsf{B}(\neg\varphi \supset \psi) \supset (\mathsf{B}\neg\varphi \supset \mathsf{B}\psi)$	Ax. 1
(2)	$[\mathsf{B}(\neg\varphi \supset \psi) \wedge \mathsf{B}\neg\varphi] \supset \mathsf{B}\psi$	(1), Importation
(3)	$[\mathsf{B}(\varphi \vee \psi) \wedge \mathsf{B}\neg\varphi] \supset \mathsf{B}\psi$	(2), Df. 2, Rule R

Systems $S.0$ and $S.1$ succeed in eradicating logical omniscience and in doing justice to the agent's limited rationality at the same time. The limited rationality is reflected by the incompleteness of the rules granted to the agents. One might, however, want to account for another kind of limitation. Consider an agent who has learned a complete set of rules of natural deduction (a set of rules powerful enough to secure the derivability of all validities). This agent may be said to

be potentially logically omniscient, but actually not logically omniscient due to limitation of time and memory. We shall describe a new epistemic logic which addresses this issue. Before we spell it out, we shall show that logicians who take computation time into account should not be blamed for committing some form of "psychologistic fallacy".

6.4   The Role Of Time In Logical Computation

As Lewis Carroll's paradox of inference shows, we cannot derive proposition q from p and $p \supset q$ by applying the proposition $[p \wedge (p \supset q)] \supset q$ on pain of generating an infinite regress [Toms, 1962, p. 44]. What is missing is the operation of detachment licensed by the inference rule of *modus ponens*. The drawback of the awareness system is its failure to take into account the procedural nature of inference.

The necessity of performing an operation cannot be bypassed by the recourse to truth-tables. Being aware of all the values of truth-functions expressed by the connectives in $p \wedge (p \supset q)$ does not provide an answer to the question: "What follows from p and $p \supset q$?" unless we cross out the lines in which either p or $p \supset q$ or both are assigned the value 0.

If we reckon with that operation of erasing lines, we free ourselves from the perspicuous criticism formulated by F. Lepage and S. Lapierre in this passage: "The notion of interpretation, from its Tarskian origins until to-day, is spoiled by a major original sin: the values of the expressions retain no trace of the way they have been computed or assigned" [2000, p. 179].

We shall now examine a new treatment of the problem of logical omniscience which duly takes computation into account. To understand the full significance of this new treatment and to show that it is immune to the criticism levelled against psychologism, one should bear in mind this well known truth about derivation: if (1) every formula has a finite length, (2) every proof is finite, and (3) the propositional calculus adopted is decidable, then every formula of the chosen propositional calculus which is derivable will be derived "in the long run". We have to enumerate longer and longer sequences of formulas and periodically check whether they are proofs of φ or of $\neg\varphi$. Two notions emerge in this uncontroversial statement: the notion of "after" and the notion of "proof".

6.5   A New Dynamic Epistemic Logic

Bringing together the two above-mentioned notions, Dr. Ho Ngoc Duc has worked out a new epistemic logic in which the tense operator $\diamond$ understood as "sometimes after using rule R" and its dual $\square$ understood as "always after using rule R" are added to the logical constants of standard epistemic logic. The knowledge operator K is given its usual interpretation which goes back to Hintikka: "$\mathsf{K}_i\varphi$" means "in all states compatible with (or accessible from) what agent i knows, it is the case that φ".

This new system called DEK_N (and its variants) is to be found in a Ph.D. dissertation defended in [2001] at the University of Leipzig .

The basic system DEK_N contains five sets of axioms:

(1)   Lukasiewicz's axioms for propositional logic;
(2)   Axioms K and 4 for the box operator ($\Box$) of temporal logic;
(3)   A weakened version of axiom K for the K-operator of epistemic logic:
      $\mathsf{K}_i(\varphi \supset \psi) \supset (\mathsf{K}_i\varphi \supset \Diamond\mathsf{K}_i\varphi)$;
(4)   A dynamic epistemic version of Lukasiewicz's axioms obtained
      by prefixing them with $\Diamond\mathsf{K}_i$;
(5)   The persistence axiom: $\mathsf{K}_i\varphi \supset \Box\mathsf{K}_i\varphi$,
      which states that premises that are true at a time remain true.

and two rules of inference: *modus ponens* and necessitation restricted to the box operator of tense logic. The necessitation rule for K is not admitted.

None of the seven forms of logical omniscience listed in § 1 is a theorem. As an example we shall show that $\mathsf{K}(\varphi \supset \psi) \supset (\mathsf{K}\varphi \supset \mathsf{K}\psi)$ is not a theorem. For that purpose we have to build a model in which $\mathsf{K}(\varphi \supset \psi) \wedge \mathsf{K}\varphi \wedge \neg\mathsf{K}\psi$ is satisfied. The following concrete model, inspired by J. van Benthem [private communication] will do. Fix a rule system. Let states be all finite sets of formulas, and let temporal steps add conclusions via the rules, but one by one.

Let **K** be a knowledge base, i.e. a finite set of formulas containing either Boolean formulas or epistemic formulas. $\mathsf{K}\varphi$ is interpreted as $\varphi \in \mathbf{K}$. Hence $\mathsf{K}\neg\varphi$ is interpreted as $\neg\varphi \in \mathbf{K}$ and $\neg\mathsf{K}\psi$ is interpreted as $\psi \notin \mathbf{K}$; $\Diamond\mathsf{K}\varphi$ is read as "Sometimes after using rule R, the agent knows φ". The application of a rule leads the agent from knowledge base **K** to knowledge base $\mathbf{K}'$. The move can be depicted by an arrow. In virtue of the persistence axiom, if $\varphi \in \mathbf{K}$ and if $\mathbf{K}'$ is directly reachable from **K** by an arrow, then $\varphi \in \mathbf{K}'$. In this structure, formula $\mathsf{K}(\varphi \supset \psi) \wedge \mathsf{K}\varphi \wedge \neg\mathsf{K}\psi$ is satisfiable. It says that $\varphi \supset \psi \in \mathbf{K}$, $\varphi \in \mathbf{K}$ and $\psi \notin \mathbf{K}$. This state of affairs is realizable. On the contrary $\mathsf{K}(\varphi \supset \psi) \wedge \mathsf{K}\varphi \wedge \neg\Diamond\mathsf{K}\psi$ is not satisfiable. The third conjunct $\neg\Diamond\mathsf{K}\psi$ is equivalent to $\Box\neg\mathsf{K}\psi$, which means "after all applications of a rule, ψ is still not a member of $\mathbf{K}'$". A counter-example is easy to find. Take $\mathbf{K}'$ as obtained from **K** by applying *modus ponens*.

6.6   *Interaction Between* B *And* R

Most systems of epistemic logic which succeed in avoiding all the forms of logical omniscience in propositional logic fail to give rise to new non trivial theorems. This is a serious defect if our goal is to give a formal account of the limited rationality of human or artificial agents. The system DEK_N and its variants are free from this objection. For all its simplicity the bimodal system DEK_N fits its intended interpretation remarkably well. Consider the formula $[\Diamond\mathsf{K}\varphi \wedge \Diamond\mathsf{K}(\varphi \supset \psi)] \supset \Diamond\mathsf{K}\psi$. It means: if after some course of thought the first premise φ is known, and if after some other course of thought the second premise $\varphi \supset \psi$ is known, then after some course of thought the conclusion ψ will be known.

The formula, as the author observes, is not a theorem and this is as it should be for the intended interpretation. The agent may fail to place side by side the two premises. If this happens we can say that the two premises diverge. The relation *later than* corresponding to the box operator allows for such a divergence. The only property imposed on time by the system K4 is transitivity. Linearity is not required. Hence time may be branching. Branching time captures the situation in which our courses of thought split off in such a way that we never reach the conclusion of our premises. To bring that out, we shall construct a model which invalidates the formula $(\Diamond \mathsf{K}\varphi \wedge \Diamond \mathsf{K}(\varphi \supset \psi)) \supset \Diamond \mathsf{K}\psi$ and satisfies its negation, i.e. $\Diamond \mathsf{K}_i\varphi \wedge \Diamond \mathsf{K}_i(\varphi \supset \psi) \wedge \Box \neg \mathsf{K}_i\psi$. This amounts to producing a structure fulfilling the following requirements : from the same point 0, two diverging arrows lead to two distinct knowledge bases $\mathbf{K}_1$ and $\mathbf{K}_2$. The first one only contains $\varphi \supset \psi$ and the second one only contains φ. Neither of them can lead to $\mathbf{K}_3$ by *modus ponens*. Hence formula $\Diamond \mathsf{K}_i\varphi \wedge \Diamond \mathsf{K}_i(\varphi \supset \psi) \wedge \Box \neg \mathsf{K}_i\psi$ is satisfied. If, however, we adopt the axiom of convergence which forces $\mathbf{K}_1$ and $\mathbf{K}_2$ to coalesce, then we can apply *modus ponens* and it is no longer possible to satisfy $\Diamond \mathsf{K}_i\varphi \wedge \Diamond \mathsf{K}_i(\varphi \supset \psi) \wedge \Box \neg \mathsf{K}_i\psi$. In other words, if time is assumed to be confluent we recover the validity of $(\Diamond \mathsf{K}\varphi \wedge \Diamond \mathsf{K}(\varphi \supset \psi)) \supset \Diamond \mathsf{K}\psi$. Dr. Ho Ngoc Duc showed indeed how $(\Diamond \mathsf{K}\varphi \wedge \Diamond \mathsf{K}(\varphi \supset \psi)) \supset \Diamond \mathsf{K}\psi$ can be derived from the axioms of DEK_N augmented with the well known axiom of convergence G, namely $\Diamond \Box \varphi \supset \Box \Diamond \varphi$. This is as it should be since, as we saw in Section 2.5, this modal axiom corresponds to the first-order property of confluence.

6.7   *Substructural Logic As A Remedy To Logical Omniscience*

M. Cozic approached the problem of logical omniscience in a new way. According to the received view, perfect rationality in epistemic reasoning represented by S5 should be limited from outside by limitations added *a posteriori* (such as limited awareness, limited analytic power, absent-mindedness, memory shortage). M. Cozic holds that this view is misguided. It does not do justice to a crucial difference between alethic logic and epistemic logic.

In alethic logic we require the preservation of truth only. In epistemic logic we expect that the rules of logic carry truth from the premises $\varphi_1, \ldots, \varphi_n$ to the conclusion ψ. But we also expect that they carry epistemic access over and above truth. More precisely, rules of epistemic logic must comply with the following principle of epistemic preservation: the justification of the premises $\varphi_1, \ldots, \varphi_n$ must imply that of the conclusion ψ.

This principle (due to J. Dubucs, [1997]) acquires a new content when we cease to look at logic as a set of truths of logic and look at it as a set of inference methods, especially if we adopt the format of Gentzen's Sequent Calculus in which a distinction is drawn between operational rules which govern the use of connectives, quantifiers and operators on one side and structural rules on the other. Let us consider the structural law of contraction which reads as follows (φ and ψ stand for formulas, Γ and Δ stand for sets of formulas):

$$\frac{\Gamma, \varphi, \varphi, \Delta \vdash \psi}{\Gamma, \varphi, \Delta \vdash \psi}$$

Epistemically interpreted the rule of contraction allows us to infer from an agent's being able to construct a justification of ψ on the basis of several justifications of φ_i that the same agent can spare some justifications of φ_i in constructing a justification of ψ. Treating premises like "resources" in the proof process helps us see that the rule is questionable. As M. Cozic observes, our ability to generate ψ by using the resource φ_i twice does not prove that we could do it by using it only once. Hence a realistic epistemic logic should be wary of granting structural rules like the rule of contraction too liberally. The law of contraction is one of the hidden roots of logical omniscience hence if we drop it we reduce logical omniscience.

Dropping structural rules in the sequent calculus is a well-known policy which is an essential ingredient of Substructural logics such as Lambek calculus, relevant logic and linear logic. Linear logic which does without contraction rule provides a formalism which is sensitive to resources. M. Cozic suggests that it could be used to build an epistemic logic which captures the bounded character of the agent's reasoning *from within*, i.e. without bringing extraneous and psychological considerations into the picture. More precisely M. Cozic proposes a sequent calculus for the implicational fragment of epistemic logic which takes the form of a sequent calculus made up of three components:

(1) Structural rules: exchange in the antecedent of the sequent, cut and the identity axiom $\varphi \vdash \varphi$,

(2) The standard sequent rule of introduction and elimination of $\supset$, namely:

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \supset \psi}$$

and

$$\frac{\Gamma \vdash \varphi \supset \psi}{\Gamma, \varphi \vdash \psi}$$

(3) A doxastic monotony rule for 'B' which is the box operator of doxastic logic:

$$\frac{\varphi \vdash \psi}{\mathsf{B}\varphi \vdash \mathsf{B}\psi}$$

A doxastic monotony rule for 'P' which is its diamond operator:

$$\frac{\varphi \vdash \psi}{\mathsf{P}\varphi \vdash \mathsf{P}\psi}$$

The D axiom: $\mathsf{B}\varphi \vdash \mathsf{P}\varphi$.

The rules of doxastic monotony mentioned under the heading (3) generate omniscience but the withdrawal of some structural rules (let us just mention the contraction rule) imposes inner bounds to the inferential apparatus and to that extent it accounts for the lack of logical omniscience of the agent.

6.8 How To Provide Linear Doxastic Logic With A Unified Formal Semantics

Whoever intends to combine linear with epistemic logic has to face a major challenge. He has to provide a unified semantics for a system which combines linear implication with doxastic modalities. The challenge was taken up in 1997 by M. D'Agostino, D. Gabbay and A. Russo. A striking innovation of their semantics lies in the crucial role played by the concept of information in the new semantics.

In standard semantics for modal logic, the modal formulas $\Box\varphi$ or $\Diamond\varphi$ are verified in world w_i if φ is verified in all, respectively some, worlds w_j accessible from w_i. The intuitive idea which underlies the new semantics is that the verification of a proposition of the form $\Box\varphi$ or $\Diamond\varphi$ by means of a given information token or resource x, depends on what is verified by other information tokens or resources accessible from x. That emphasis put on the notion of information is characteristic of the "move away from still reflection of abstract truth to a concern with the structure of information and the mechanism of its processing", a tendency of recent logical research to which J. van Benthem drew attention in [1991, p. 185].

The formal semantics needed to interpret substructural logic rests upon frames which have an algebraic structure richer than the frames used in the usual Kripke semantics for modal logic. Information frames, called "quantale frames", are introduced. A quantale frame is a structure $\mathbf{F} = \langle Q, \circ, 1, \sqsubseteq \rangle$ such that:

1. Q is a non-empty set of elements called information tokens.

2. $\sqsubseteq$ is a partial ordering which makes Q into a complete lattice; "$x \sqsubseteq y$" can be read as "y contains at least the same information as x".

3. $\circ$ is a binary operation on Q which is associative and distributive over the lattice join.

4. $1 \in Q$ is a unit element for Q.

Different classes of quantale frames can be defined by imposing additional conditions on the ordering relation $\sqsubseteq$. Let us just mention the contractive constraint: $x \circ x \sqsubseteq x$. A quantale frame validates the structural rule of contraction iff it satisfies the contractive constraint.

To get a fully-fledged semantics which enables us to map information tokens onto formulas of the language L of our substructural logic, we need to turn our quantale frame into a model, i.e. we have to define , as expected, a *valuation function*. Here again, the construction becomes a little harder. The valuation has to satisfy additional conditions (one of them is the *heredity condition* which says that if a formula evaluated with respect to an information token x is true, it remains true when evaluated with respect to an information token y which contains at least the same information as x). To turn a *quantale frame* for substructural logic into a *modal quantale frame* for substructural modal logic, we have to enrich the quantale frame with an accessibility relation which captures the meaning of

the modal (doxastic) operator. Here again a few refinements are needed. Firstly we have to require that the accessibility relation be closed under arbitrary join and meet of the lattice under consideration. Secondly we have to impose conditions which ensure that the hereditary property of valuation are preserved when we add the accessibility relation to the quantale frame. As shown by M. D'Agostino, D. Gabbay and A. Russo [1997] this amounts to satisfying the following conditions:

1. If $x \sqsubseteq y$ and xRz then $\exists z'(yRz' \wedge z \sqsubseteq z')$.

2. If $x \sqsubseteq y$ and yRz then $\exists z'(xRz' \wedge z' \sqsubseteq z)$.

Moreover the definition of what it means for a formula φ to be satisfied in a modal implication model differs from the classical definition: a formula φ is verified in a modal implication model $\mathbf{M}$ if it is verified at the identity point 1 of $\mathbf{M}$. It is verified in a frame $\mathbf{F}$ if it is verified in all modal implication models based on $\mathbf{F}$.

M. Cozic's linear doxastic logic couched in the formalism of Sequent Calculus can be proved to be sound and complete for the semantics sketched above. In the 1997 paper M. D'Agostino *et al.* provided a detailed soundness and completeness proof for several modal substructural logics built on the pattern of the labelled tableau proof method described in [1994; 2000].

6.9 *Logical Omniscience And Belief Revision*

We have just described a very innovative set up that eliminates all the forms of logical omniscience and takes into account the dynamic character of inference. We shall now examine another approach which also brings into focus the dynamics of inference, i.e. a new version of the AGM theory of belief revision described in section 4.

As we have seen, classical AGM theory of belief revision describes belief states as belief sets $\mathbf{S}$ upon which three basic operations are defined, namely expansion, contraction and revision. R. Wassermann observes that the AGM paradigm is a theory of highly idealized reasoners. The closure of beliefs under logical inference is built into the very definition of the basic operations in the terms of which belief change is described [Wassermann, 1999, p. 429]. Moreover one of the postulates characterizing contraction stipulates that contraction is closed under logical equivalence. The sixth postulate reads as follows: If $\vdash \varphi \equiv \psi$ then $\mathbf{K} \setminus \varphi = \mathbf{K} \setminus \psi$ [Wassermann, 2000, p. 20]. The question arises whether the theory of belief revision could be freed of this idealization.

R. Wassermann has worked out a theory of belief revision for resource bounded agents that precisely achieves that goal. Her new theory is expressive enough to distinguish different statuses of beliefs according as they are implicit, explicit, provisional or active and to represent the agent as a reasoner operating with bounded resources and finite memory. Instead of defining belief states in terms of belief sets as it is the case in the standard AGM theory, Fuhrmann and others define them in terms of belief bases. As opposed to belief sets, belief bases are finite sets

of formulas which are not closed under logical consequence. Moreover, as belief bases are sensitive to syntactic differences, they provide us with the possibility to treat hidden inconsistencies and patent inconsistencies in a different way.

However significant the replacement of belief sets by belief bases may be, it does not suffice to eliminate the shortcomings of the classical theory of belief revision. The source of the trouble lies in the operations of belief change defined for belief bases. As R. Wassermann observes, generally they "still make use of the operation of logical closure" [2000, p. 35]. An additional innovation is needed to get around the logical closure predicament. Such an innovation will now be described.

6.10   Compartments And Local Inference

Psychologists have drawn a distinction between long-term and short-term memory. R. Wassermann's formalism captures this distinction. In her framework, an agent's long-term memory is represented as a belief base, i.e. as a set of formulas which is not closed under logical consequence. Short-term memory is the place where belief changes occur. The operations of expansion, contraction and revision are no longer defined over the whole belief base but only over compartments of the belief base. The compartment of the belief base $\mathbf{B}$ around the formula φ is the set of formulas of $\mathbf{B}$ that are logically relevant to φ, namely that contribute to prove or disprove φ (where φ is neither a tautology nor a contradiction).

Using the notion of compartment $c(\varphi, \mathbf{B})$, R. Wassermann defines a localized consequence operation. It turns out that the consequence of $\mathbf{B}$ localized to formula φ (or to the set of formulas $\mathbf{R}$) = the classical consequence of the compartment of $\mathbf{B}$ around φ (or around $\mathbf{R}$). Hence the consequence operator Cn_φ (respectively $Cn_\mathbf{R}$) behaves classically inside the compartment around φ, but not outside. For instance if an inconsistency occurs inside the compartment around φ, the whole compartment is spoiled in virtue of the principle *ex falso sequitur quodlibet*. An inconsistency located outside the compartment however does not trivialize the whole belief base. This captures the difference between inconsistencies which fly in your face and inconsistencies which remain hidden.

At a later stage, the local consequence operation is used to define local versions of the standard operations of belief revision. This is a striking result. Given that the local notion of consequence only shares a few properties with the standard consequence operation (compactness and monotony) one would not expect to obtain local versions of the operation of revision that are very similar to the standard ones. The local partial meet contraction offers a good illustration of how this can be done. Let $\mathbf{B} \perp \varphi$ denote the maximal subsets of belief base $\mathbf{B}$ that fail to imply φ ($\perp$ is here the remainder operator, not the symbol *falsum*). Let γ be a function that selects some elements of $\mathbf{B} \perp \varphi$. Take the intersection $\bigcap \gamma(\mathbf{B} \perp \varphi)$, you get partial meet contraction: $\mathbf{B} \setminus \varphi$. If you want to contract a belief base $\mathbf{B}$ by the formula φ with respect to a set of formulas $\mathbf{R}$, all you have to do is to take the compartment $c(\mathbf{R}, \mathbf{B})$ rather than $\mathbf{B}$ as first argument of the operator $\perp$. This amounts to saying that if you want to construct the local partial meet contraction

$\mathbf{B} \setminus_R \varphi$, the beliefs to be discarded are those in the R-compartment of $\mathbf{B}$ that are not contained in all the selected φ-remainders of the compartment [Wassermann, 1999, p. 439].

6.11   The Dynamics Of Inference

A belief state is a triple of the form $\langle \mathbf{E}, \mathbf{Inf}, \mathbf{A} \rangle$; $\mathbf{E}$ represents the set of the agent's explicit beliefs, $\mathbf{Inf}$ represents the functions which return the agent's inferred beliefs when they are applied to a set of beliefs of an agent; $\mathbf{A}$ represents the agent's set of active beliefs. An inference can be seen as a sequence of steps from an initial belief state to a terminal belief state. Each of these steps are micro-operations (observation retrieval, deletion and so on) which underlie the macro-operations of expansion, contraction and revision.

R. Wassermann's formalism succeeds in capturing not only the reasoning structure, but also the reasoning process. It should be stressed that she achieves this goal without falling prey to psychologism, i.e. without blurring the distinction between logic and psychology. Let $\mathbf{E} = \{\neg a, \neg b, a \vee b, q, q \supset p\}$ be the set of explicit beliefs of an agent contained in his or her long-term memory. Let $\mathbf{A} = \{\neg p\}$ be the set of active beliefs contained in his or her short-term memory. The occurrence of $\neg p$ will lead the agent to retrieve q and $q \supset p$ and to infer p. The set of active beliefs is enlarged. We now have $\mathbf{A}' = \{q, q \supset p, \neg q, p\}$. An inconsistency has occurred which can be eliminated by local partial meet consolidation (a variant of contraction in which *falsum* plays the role of φ in $\mathbf{A} \perp \varphi$) [Wassermann, 1999, p. 442].

Let us observe that the inconsistency contained in $\mathbf{E} \setminus \mathbf{A}$ is innocuous, as opposed to the new inconsistency which arose in $\mathbf{A}'$. This difference of treatment is by no means arbitrary. The inconsistency of $\{\neg a, \neg b, a \vee b\}$ is innocuous in so far as it does not belong to the active part of the belief base. Is $\mathbf{E}$ closed under *modus ponens*? R. Wassermann realizes that the question should not be given a Yes-No answer. As long as q and $q \supset p$ remain in $\mathbf{E} \cap \mathbf{A}$, the answer is negative. As soon as the premises enter into $\mathbf{Inf}(\mathbf{E} \cap \mathbf{A})$, the answer is positive. This twofold answer does justice to the dynamic character of inference: some of the explicit beliefs of an agent (but not all) are retrieved in the set of active sentences. This change however is not a merely psychological event. It is rationally motivated in so far as the construction of compartments which shows which beliefs should move from $\mathbf{E}$ to $\mathbf{A}$ rests upon logical considerations. (See the definition of "compartment around φ".) The logical character of compartments distinguishes them from clusters which also serve to accommodate local reasoning (see section 2.6). Even considerations of computational efficiency have been taken into account in the study of how to structure belief bases [Wassermann, 2001].

7 COMMON KNOWLEDGE AND COMMON BELIEF

7.1 A Challenging Combination Of Infinity And Effectivity

Formula $\mathsf{E}\varphi$ is true if every agent i knows φ. Hence it is quite natural to take as an axiom:

$$\vdash \mathsf{E}\varphi \;\equiv\; \bigwedge_i \mathsf{K}_i\varphi\,.$$

To capture common knowledge $\mathsf{C}\varphi$, Fagin *et al.* used the fixed point axiom:

$$\vdash \mathsf{C}\varphi \;\equiv\; \mathsf{E}(\varphi \wedge \mathsf{C}\varphi)\,.$$

In their account of common belief (also represented by the symbol C), L. Lismont and Ph. Mongin use the definition:

$$\mathsf{E}\varphi \;=_{def}\; \bigwedge_i \mathsf{B}_i\varphi\,,$$

and the fixed point axiom (*FP*):

$$\vdash \mathsf{C}\varphi \;\supset\; \mathsf{E}(\varphi \wedge \mathsf{C}\varphi)\,.$$

The fixed point axiom is needed because "the commonsense definition of C through an infinite conjunction of higher-order belief sentences could not be expressed directly in the formal language [of classical, i.e. finitary, logic]" [Lismont and Mongin, 1994b, p. 79]. Moreover the fixed point axiom in conjunction with suitable monotonicity requirements on E and C generate $\mathsf{C}\varphi \supset \mathsf{E}^k\varphi$ for any finite number k larger than 1. We shall see later that the fixed point axiom can be replaced by an infinite conjunction in an infinitary logic.

Independently, Fagin *et al.*, on the one hand, and L. Lismont and Ph. Mongin, on the other, managed to prove the soundness and completeness of a finite axiomatization of common knowledge and common belief and to establish an even more startling result: the decidability of propositional logic of common knowledge and common belief. This came as a surprise: "given the semantic force of the CB operator [Common Belief Operator] [...] one would have expected that properties of this operator could *not* be falsified by referring to finite models only" [Lismont and Mongin, 1994b, p. 99].

7.2 A Weak Axiomatization Of Common Belief

As we saw in section 5, standard epistemic and doxastic logic such as $KT45$ and $KD45$ respectively are affected by the problem of logical omniscience. This is true also of the logic of common belief. Three axioms or rules are specially damaging: (1) the necessitation rule which discards models in which agents do not believe anything at all; (2) the axiom of closure under conjunction which is incompatible with probabilistic belief except for the limiting case of events having probability 1;

(3) the monotonicity rule for doxastic logic $\varphi \supset \psi \vdash \mathsf{B}\varphi \supset \mathsf{B}\psi$ which involves a questionable commitment to logical omniscience. L. Lismont and Ph. Mongin built an axiomatic system with two components, called (1) the Individual Belief Axiom Block and (2) the Common Belief Axiom Block respectively, in which necessitation and conjunctiveness are dropped while a restricted monotonicity rule is added to the standard one. They proved that this axiomatic system for individual and common belief is sound and complete for a special class of neighbourhood structures, the C-restricted Monotonic Structures.

The new monotonicity rule for individual belief (B) is: $\varphi \supset \psi \vdash \mathsf{C}\varphi \supset \mathsf{B}\psi$ which can be rendered in this way: if a statement implies another then if the former is common belief then the latter is private belief. This captures the idea that we believe the consequences of commonplaces (Aristotelian *topoi*) that everybody believes. The first block contains the standard monotonicity rule for individual beliefs, the innocuous definition of "everybody believes" for a finite number of agents and the new rule of restricted monotonicity. The second block contains the fixed point axiom for common belief, the standard monotonicity rule for common belief, i.e. $\varphi \supset \psi \vdash \mathsf{C}\varphi \supset \mathsf{C}\psi$ and the induction rule $\varphi \supset \mathsf{E}\varphi \vdash \mathsf{E}\varphi \supset \mathsf{C}\varphi$.

The induction rule establishes a connection between public and common belief which has been highlighted in economic literature. It says that if a statement φ is inherently public — if it is a theorem that φ cannot happen without everybody's believing it — then φ is inherently common belief [Lismont and Mongin, 1994a, p. 367].

As far as decidability is concerned, Halpern and Moses proved that the language of common knowledge (or common belief) is decidable [1985; 1992 first version in 1985] by extending the completeness and complexity results for PDL due to Fischer and Ladner (for complexity) and those due to Kozen and Parikh (for completeness).

Cognitive philosophers have complained that the epistemic states of common knowledge or common belief can only be reached after the agent has performed infinitely many steps. If this were the case, the efforts deployed to get rid of necessitation and conjunctiveness would have little significance. L. Lismont and Ph. Mongin reply that the sort of infinity they are concerned with here is merely potential infinity. For reasons of elegance, the iteration of the knowledge operator allowed by the FP axiom is unbounded, but real iteration involved in particular application is always finite. A case in point is the muddy children puzzle. The actual inference steps which the children must perform to answer the query of their teacher is finite. Hence "[a]ny particular model should involve a finite sequence of shared belief operators $\mathsf{E}^1, \ldots, \mathsf{E}^k$, but it is easier and more elegant to encompass all particular models at once by introducing C" [Lismont and Mongin, 1994b, p. 100].

7.3   A Neighbourhood Semantics For The Logic Of Common Beliefs

As we saw in section 4, standard Kripke semantics for modal logic involves relational structures ("frames") of this form: $\mathbf{F} = \langle W, \mathbf{R}_1, \ldots, \mathbf{R}_n \rangle$. There is, however, a more general kind of structures used for the interpretation of modal logics, the neighbourhood structures, also called Scott-Montague structures. In these structures, instead of working with a point to point relation $\mathbf{R} \subseteq W \times W$, we work with a function $\mathbf{F} \to \mathbf{2}^W$ or point to set relation [Gabbay, 1976, p. 2] and [Chellas, 1980, ch. 7].

Those structures are closed under logical equivalence, but free of the other forms of logical omniscience [Fagin *et al.*, 1995, p. 318]. This is one of the reasons why L. Lismont and Ph. Mongin find them philosophically more appealing when an epistemic interpretation of the formal system is intended.

The C-restricted Monotonic Structures introduced by L. Lismont and Ph. Mongin are a variant of the neighbourhood structures described by B. Chellas under the name of "minimal models". Let A denote a finite set of agents and $|A|$ its cardinality. A C-Restricted Monotonic Structure is any $(|A| + 2)$-tuple:

$$m = \langle W, (\mathbf{N}_a)_{a \in A}, \mathbf{v} \rangle$$

where:

- W is a nonempty set (the set of possible worlds),

- For all $a \in A$, N_a is a mapping from the set W into the power set of the power set of W. The elements of $N_a(w)$ are the *neighbourhoods* of world w for agent a. We also define $N_{\mathsf{E}}(w)$ as $\bigcap_{a \in A} N_a(w)$. By $P \in N_{\mathsf{E}}(w)$ we mean that P is a neighbourhood of w for all agents.
  For each subset P of W, let $i(P)$ denote the set of all worlds for which P is a neighbourhood for all agents, i.e. $i(P) =_{def} \{w \in W : P \in N_{\mathsf{E}}(w)\}$.
  The following C-*Restricted Monotonic Closure* condition must be satisfied:

  If $w \in i(P)$ and if $P \subset i(P)$,
  then every superset of P is a neighbourhood of world w for all agents.

  This rather involved condition replaces the simpler (and stronger) condition of *Monotonic Closure* which reads as follows:

  If $w \in i(P)$
  then every superset of P is a neighbourhood of world w for all agents.

- $\mathbf{v}$ is a mapping such that $\mathbf{v}(w, p)$ is the truth-value of propositional variable p at world w.

What comes next is a recursive definition of truth at world w in the model structure m. Before stating the clauses for modal formulas $\mathsf{B}_a\psi$, $\mathsf{E}\psi$ and $\mathsf{C}\psi$, it is useful to give the intuitive interpretation of the key notions. If ψ is a formula, $\|\psi\|^m$ is the set of worlds which "forces" ψ (i.e. which supports the truth of ψ) at world w. This set of worlds is often called the proposition expressed by ψ. The class $N_a(w)$ of subsets of W, called a neighbourhood system for a at w, is just a system of beliefs for agent a at world w. The pair $\langle m, w \rangle$ forces $\mathsf{B}_a\psi$ iff the proposition $\|\psi\|^m$ is a neighbourhood of w for a, formally:

- $\langle m, w \rangle \models \mathsf{B}_a\psi$ iff $\|\psi\|^m \in N_a(w)$.

The clause for $\mathsf{E}_a\psi$ is the same, except that the system of beliefs considered is that of everybody, rather than that of a single agent a, formally:

- $\langle m, w \rangle \models \mathsf{E}\psi$ iff $\|\psi\|^m \in N_\mathsf{E}(w)$.

The clause for common belief is as follows:

- $\langle m, w \rangle \models \mathsf{C}\psi$ iff there is a subset P of W such that

    - $P \subseteq \|\psi\|^m$;
    - $P \in N_\mathsf{E}(w)$;
    - $P \subseteq i(P)$.

Observe that L. Lismont and Ph. Mongin allow the minimal amount of monotonicity and logical omniscience required by the semantics of the operator of common belief C [Lismont and Mongin, 1994a].

7.4   Two Styles Of Logic For Common Knowledge

We surveyed papers in the literature of epistemic logic in which the concept of common knowledge is treated as a part of logic and in which a fixed point operator is employed. This is not the only way of handling the problem. M. Kaneko and T. Nagashima advocate another approach. As common knowledge is an infinitary concept, they chose "a framework in which infinitary conjunctions and disjunctions are allowed to express common knowledge explicitly as a logical formula" [Kaneko and Nagashima, 1996, p. 326]. This new framework enabled them to treat common knowledge as an object instead of a part of their logic.

However different they might be, the two approaches are not unrelated. M. Kaneko constructed a faithful embedding of the propositional common knowledge logics into infinitary ones [1999]. Both approaches have the same power in propositional logic of common knowledge. But we need first-order logic of common knowledge whenever the application domain is infinite, or when we want to represent individual or common knowledge about a finite domain whose cardinality is not known in advance. In first-order logic, however, an important difference between the fixed point approach and the infinitary approach was discovered in 2000 which forces us to reconsider the matter. F. Wolter has proved that finitary first-order logic of common knowledge is incomplete [2000].

7.5 Finitary First-Order Logic Of Common Knowledge

As J. H. Halpern and Y. Moses have shown, propositional logics for common knowledge which are very close to the standard epistemic logic for n agent logic can be obtained by extending the familiar system $KD4^n$ with one axiom, i.e. the axiom CA designed to capture the fixed point property, and one rule of inference, i.e. the induction rule CI. Below we review the propositional logic of common knowledge presented by Kaneko *et al.* under the name HM (Halpern Moses) and we focus on the transition from propositional to first-order logic of common knowledge [Kaneko *et al.*, 2002]. The language of HM logic contains: (1) free variables $a_0, a_1, \ldots$; (2) bound variables $x_0, x_1, \ldots$; (3) the usual connectives; (4) the usual quantifiers; (5) function symbols: $f_0, f_1, \ldots$; (6) predicate symbols $P_0, P_1, \ldots$; (7) unary belief operator symbols: $\mathsf{B}_1, \ldots, \mathsf{B}_n$; (8) Unary common knowledge symbol C; (9) parentheses.

The axioms are those of $KD4^n$ and

$$CA: \quad \mathsf{C}A \supset (A \wedge \mathsf{B}_1\mathsf{C}A \wedge \ldots \wedge \mathsf{B}_n\mathsf{C}A);$$
$$CI: \quad D \supset (A \wedge \mathsf{B}_1 D \wedge \ldots \wedge \mathsf{B}_n D) \vdash D \supset \mathsf{C}A.$$

A predicate logic of common knowledge called QHM by [Kaneko *et al.*, 2002] can be obtained by adding CA and CI to $QKD4^n$ which is the quantified version of $KD4^n$. $QKD4^n$ contains $KD4^n$ axioms and rules together with these new axioms:

$$\forall x A(x) \supset A(t),$$
$$A(t) \supset \exists x A(x),$$

and these new rules:

$$A \supset B(a) \vdash A \supset \forall x B(x),$$
$$A(a) \supset B \vdash \exists x A(x) \supset B,$$

and the Barcan formula (BF):

$$\forall x \, \mathsf{B}_i A(x) \supset \mathsf{B}_i \, \forall x A(x).$$

7.6 A Semantics for First-order Epistemic Logic With Common Knowledge And Common Belief

A Kripke frame $\mathbf{F}$ for the logic under consideration is a tuple $(W; \mathbf{R}_1, \ldots, \mathbf{R}_n; D)$. The domain of individuals D is the same for all worlds (a sufficient condition for satisfying the Barcan formula).

We turn the frame $\mathbf{F}$ into a model $\mathbf{M}$ by adding an interpretation I, namely a function that assigns a function from D^k to D to each k-ary function symbol. The interpretation of each function symbol remains constant over W. Hence proper names, i.e. function symbols with arity 0, are rigid designators. The interpretation function I also assigns to each k-ary predicate a subset of D^k which may vary from world to world.

The accessibility relations are serial and transitive (to satisfy D and 4). Free variable are interpreted independently of possible worlds. For that purpose an assignment function is used: $\sigma : V \to D$.

Next the notion of reachability is introduced: $u \in W$ is 1-reachable from w in the Kripke frame $\mathbf{F}$ iff $(u, w) \in \bigcup_{i=1}^{n} \mathbf{R}_i$; the reachability relation is the transitive closure of the 1-reachability relation.

The recursive definition of truth relative to model $\mathbf{M}$, assignment function σ and world w is standard except for the clause concerning common knowledge which reads as follows:

$$(\mathbf{M}, \sigma, w) \models \mathsf{C}A \quad \text{iff} \quad (\mathbf{M}, \sigma, w) \models A \text{ for all } u \text{ reachable from } w.$$

Kaneko *et al.* state a lemma that allows us to replace the above-mentioned clause for Common knowledge by the following one:

$$(\mathbf{M}, \sigma, w) \models \mathsf{C}A \quad \text{iff} \quad (\mathbf{M}, \sigma, w) \models A \wedge \mathsf{B}_{i_1} \ldots \mathsf{B}_{i_n} A$$
$$\text{for each sequence } i_1, \ldots, i_n \text{ of agents.}$$

This second version captures the intuitive meaning of *common knowledge* of formula A, i.e. that A is true, that each player believes A, that each player believes that each player believes A and so on. *Common belief* of A is defined as follows:

$$(\mathbf{M}, \sigma, w) \models \mathsf{C}A \quad \text{iff} \quad (\mathbf{M}, \sigma, w) \models \mathsf{B}_{i_1} \ldots \mathsf{B}_{i_n} A$$
$$\text{for each sequence } i_1, \ldots, i_n \text{ of agents.}$$

Completeness proofs of $KD4^n$ and $QKD4^n$ are available [Hughes and Cresswell, 1984].

7.7 The Incompleteness Of First-order Epistemic Logic With Common Knowledge And Common Belief

In Sections 1.2, 3.5–3.7, the concept of common knowledge was formally represented by a special operator and put to use in the treatment of the Muddy Children puzzle. In this section it will be studied for itself, from the standpoint of proof theory and from that of model theory.

In 2000, F. Wolter showed that weak fragments of first-order common knowledge are not recursively axiomatizable. This is the case, for instance, for fragments the first-order part of which involves names and the equality symbol only.

As H. Sturm *et al.* observe, the status of first-order logic of common knowledge is similar to that of arithmetic and second-order logic. It is impossible to characterize a semantically defined first-order logic of common knowledge by means of an effective proof system [Sturm *et al.*, 2002].

After establishing his negative results, F. Wolter raised the question as to whether there are well-behaved fragments of first-order logic of common knowledge and came to a positive conclusion. He defined the fragment of "monodic" (not to be confused with "monadic") formulas which can be shown to have good properties. The program was systematically carried out in a later work [Sturm *et al.*, 2002] examined in the next section.

7.8 A Logically Well-Behaved Fragment Of First-Order Logic Of Common Knowledge And Common Belief

By QCL_1, Sturm *et al.* [2002] denote the set of all formulas φ belonging to first-order common knowledge logic such that any subformula of φ of the form $\mathsf{K}_i\psi$ or $\mathsf{C}\psi$ has at most one free variable. Such formulas are called *monodic*.

The monodic fragment is more expressive than propositional logic. Any closed wff of first-order logic can occur within the scope of epistemic operators. Monodic formulas can formalize the *de re – de dicto* contrast between "It is commonly known that someone is taller than John" $(\mathsf{C}\exists x\, L(x,j))$ and "there is someone of whom it is commonly known that he or she is taller than John" $(\exists x\, \mathsf{C}L(x,j))$.

In the monodic fragment, we cannot formalize the difference between common knowledge that a Nash-Equilibrium for a game of $n > 1$ players exists and common knowledge of a specific Nash-Equilibrium. The reason is that in the formula $\exists x_1 \ldots \exists x_n\, \mathsf{C}\mathrm{Nash}(x_1, \ldots, x_n)$, the sub-formula $\mathrm{Nash}(x_1, \ldots, x_n)$, which falls in the scope of C, contains more than one free variable. Hence the monodic fragment is less expressive than full first-order logic of common knowledge.

There is an unavoidable trade-off between expressivity and other good logical properties such as axiomatizability and decidability. H. Sturm *et al.* [2002] have proved that the monodic fragments of first-order logics of common knowledge (QKT_n^C) and common belief (QKD_n^C) defined by the standard Kripke structures can be axiomatized. All the valid formulas are provable from the following axioms and rules: the axioms and inference rules of the propositional part, those of classical first-order logic and the Barcan formula.

The logic of common knowledge is not compact. There are infinite sets of formulas of that logic that are not satisfiable although all their finite subsets are. This is the case of the set $\zeta = \{\mathsf{E}^n p : n \geq 0\} \cup \{\neg\mathsf{C}p\}$. Hence the method of canonical models does not work for proving completeness here. New concepts had to be built (such as the concept of quasi-model) to come over this obstacle [Sturm *et al.*, 2002].

7.9 How To Recover The Completeness Of The Full First-Order Logic Of Common Knowledge

The first kind of logic to restore the completeness of the first-order logic of common knowledge admits infinitary proofs, but no infinitary formulas. It is called QCY [Kaneko *et al.*, 2002, p. 73]. The proper part of QCY, designed to handle the operator C, involves two axiom schemata $(CA^*$ and $CB)$ and one inference rule (CI^*) which is a strengthened version of rule CI_0^*.

$CA^* :\ \mathsf{C}A \supset \mathsf{B}_{i_1} \ldots \mathsf{B}_{i_n} A\,,$
        for each sequence $i_1, \ldots, i_n$ of agents;

$CB :\ \mathsf{C}A \supset \mathsf{B}_i\mathsf{C}A\,,$
        for each agent i.

The second axiom schema captures the Barcan properties of common knowledge, i.e. $\models \mathsf{C}A \supset \mathsf{B}_i\mathsf{C}A$ [Kaneko $et\ al.$, 2002, p. 11]. Remember that at this stage infinitary proofs are allowed, but infinitary formulas are not. Kaneko $et\ al.$ solved the problem by strengthening the rule CI_0^* into the rule CI^*. Both rules have an infinite set of premises, one for each finite sequence $(i_1, \ldots, i_n)$ of agents. The former is

$$CI_0^* : \{D \supset \mathsf{B}_{i_1} \ldots \mathsf{B}_{i_n} A\} \vdash D \supset \mathsf{C}A.$$

The latter, which provides completeness, is

$$CI_1^* : \{D \supset \mathcal{T}(\mathsf{B}_{i_1} \ldots \mathsf{B}_{i_n} A)\} \vdash D \supset \mathcal{T}(\mathsf{C}A).$$

The operator $\mathcal{T}$ is any substitution which maps a formula φ onto a formula of the kind $\mathsf{B}_{j_k}(D_k \supset \ldots \mathsf{B}j_2(D_2 \supset \mathsf{B}j_1(D_1 \supset \varphi))\ldots)$, where $(j_1, \ldots, j_k)$ is any finite sequence of agents and $D_1, \ldots, D_k$ are arbitrary formulas.

An alternative to QCY is the infinitary epistemic logic QGL_ω, which does not use the C operator at all, but brings in infinite conjunctions and infinite disjunctions [Kaneko and Nagashima, 1996; 1997].

An infinitary logic for common belief (as opposed to common knowledge) can be obtained in this way:

- The common belief operator $\mathsf{C_B}$ is introduced as a primitive;

- Axiom C_BA and rule C_BI replace CA and CI stated above:

  $C_BA :$ $\mathsf{C_B}(A) \supset \mathsf{B}_1(A \wedge \mathsf{C_B}(A)) \wedge \ldots \wedge \mathsf{B}_n(A \wedge \mathsf{C_B}(A))$,
  $C_BI :$ $D \supset \mathsf{B}_1(A \wedge D) \wedge \ldots \wedge \mathsf{B}_n(A \wedge D) \vdash D \subset \mathsf{C_B}(A)$.

A general method for proving the completeness of both minimum predicate and infinitary extensions of modal propositional logic was invented by Y. Tanaka and H. Ono [2001]. It sits in the algebraic approach initiated by Jónsson and Tarski in 1951 [Blackburn $et\ al.$, 2001]. The method has been applied later to infinitary first-order logic of common knowledge by Kaneko $et\ al.$ [2002] Taking stock of what has been achieved so far, the following statement made by F. Wolter is an appropriate $mot\ de\ la\ fin$: there is no finite (alias effective) way to axiomatize first-order common knowledge logics, so a non-effective axiomatization in infinitary logic is an interesting alternative [1999].

8   SOME APPLICATIONS TO SOFTWARE ENGINEERING

8.1   *Knowledge In Concurrent Programs*

Increased software reliability has been one of the most important and elusive goals of the computer scientists for more than thirty years. Testing has always been, and still is, an appealing way to assess the reliability of a program. It is based on the idea that computers themselves usually are far more reliable than programs

functioning in a deterministic way. If a program has provided correct results for some given data at a given time, it will do so again at a later time, for the same data. An obvious problem is that it is usually not possible to test a program for all possible data sets, but nevertheless it is usual to assume that, if the program behaves correctly for a (well-chosen) sample of data sets, it will behave correctly for all.

The optimistic view of the matter is at best doubtful, and at times not acceptable at all. The most important illustration of this is to be found in concurrent programming. First, concurrent programs are usually non-deterministic. Full knowledge of the initial conditions of a computation is not enough to predict the whole computation, which may depend on external factors, including the variable speed of the processes involved in the computation. As a result, a concurrent system cannot in principle be fully tested, even for some fixed set of initial data. Secondly, the behaviour of distant concurrent processes depends not only on the reliability of the computers used by these processes, but also on communication devices between computers, which are usually less reliable, or at least less deterministic, and might confuse the testing procedure. Thirdly, concurrent programming frequently occurs in critical applications, for instance the kernel of the computer operating system, or the network operating system, where a strong reliability assessment is needed.

For these reasons and for some others, formal, mostly logic-based methods, have been devised to deal with the reliability problem in concurrent programming. We outline one of them in this section and show how epistemic logic can be used to improve on it. First, we show how a specific view of knowledge underlies the design of a concurrent system. An elementary but important problem in concurrent programming is the reliable transmission of a stream X of messages through an unreliable transmission medium. Let $X = X[1], X[2], \ldots$ be the sequence of messages to be transmitted from a Sender to a (distant) Receiver (i is the rank of message $X[i]$). The Receiver collects incoming messages and, in spite of possible loss, corruption, duplication and delay, has to reconstruct the stream. If $Y = Y[1], Y[2], \ldots$ is the reconstructed stream, the concurrent system comprising the Sender and the Receiver behaves correctly if $Y = X$.

At a rather abstract level, the transmission process is modelled by the following *transition*:

$$((HS, Y[HS + 1]) := (HS + 1, X[HS + 1])) \tag{4}$$

This transition reduces to an *assignment*. Assignments are used to specify a modification of the values of one or more variables. For instance, the assignment $(x, y) := (x + y, y + 2)$ leads from a system state[24] where $x = 5$ and $y = 4$ to a state where $x = 9$ and $y = 6$. This fact is formalized into a *Hoare triple*:

$$\{x = 5 \ \wedge \ y = 4\} \ (x, y) := (x + y, y + 2) \ \{x = 9 \ \wedge \ y = 6\} \, .$$

---

[24]A *(concurrent) system state* is a function that assigns values to (computer) variables.

A transition is a binary relation on the set of all accessible system states. A concurrent system can be represented as a set of transitions. Transition t_1 represents the successful transmission of a single message. The variable HS records the rank of the *h*ighest *s*ent message. The transmission system behaves correctly if the computation consists in repeatedly executing the transition, which can be described by a Hoare triple:

$$\{HS = n \wedge Y[1:n] = X[1:n]\}\ \tau\ \{HS = n+1 \wedge Y[1:n+1] = X[1:n+1]\}\,.$$

This specification has an operational look but is not a realistic implementation of the system, since no provision has been made for dealing with possible loss or corruption of the transmitted message. To keep the problem quite elementary we assume that any corruption or duplication is detected and discarded by the Receiver. So, no distinction is needed between loss, corruption and duplication of a message. We introduce a new variable LR (for *l*ast *r*eceived); if the highest sent message has been correctly received, then the equality $HS = LR$ holds and the next message can be transmitted. Otherwise, the current message has to be transmitted again. This simple transmission policy is formalized into a set of four transitions:

$$
\begin{aligned}
&1 : (LR = HS \ \longrightarrow\ (HS, LR, Y[HS+1]) := (HS+1, LR+1, X[HS+1]))\,,\\
&2 : (LR = HS \ \longrightarrow\ HS := HS+1)\,,\\
&3 : (LR < HS \ \longrightarrow\ (LR, Y[HS]) := (LR+1, X[HS]))\,,\\
&4 : (LR < HS \ \longrightarrow\ skip)\,.
\end{aligned}
\tag{5}
$$

Transition 5.1 models the correct transmission of a new message and transition 5.2 models the failed transmission of a new message; transition 5.3 models the correct retransmission of a message and transition 5.4 models a failed retransmission. A computation of the system consists of a sequence of transition executions. At a system state where the identity $LR = HS$ holds, the next transition to be executed is 5.1 or 5.2; at a state when the equality does not hold, it is 5.3 or 5.4. The choice is made non-deterministically, which means that from a given initial state (say, a state satisfying $HS = LR = 0$), several computations are possible.[25] A *trace* is a finite or infinite sequence $(\sigma_0, \tau_1, \sigma_1, \tau_2, \ldots)$ where $(\sigma_{i-1}, \sigma_i) \in \tau_i$, i.e. transition τ_i leads from system state σ_{i-1} to state σ_i. If the trace is finite, then its last element is a terminal state, that is, a state that does not have a successor for any transition of the system. The sequence $(\sigma_0, \sigma_1, \ldots)$ of states occurring in a trace is a *computation*, or a *run*.[26] The set of computations of a system can be viewed as its semantics.

Now we have a correct operational specification of the system. This can be proved with an *invariant* I, that is an *inductive* assertion. If some system state σ satisfies I, then all its successors also satisfy I. In the formalism of Hoare triples,

---

[25]In fact, infinitely many.

[26]In practice, the distinction between "computation" and "trace" is not respected and both words are taken as synonyms.

this can be stated as

$$\{I\}\,\tau\,\{I\} \quad \text{for each transition } \tau \text{ of the system.}$$

An appropriate invariant here is

$$LR \leq HS \leq LR + 1 \ \wedge \ Y[1:LR] = X[1:LR]\,.$$

It is easy to check that, for each transition τ listed in (5), a τ-successor of a state satisfying I also satisfies I. A state where $LR = HS = 0$ satisfies I. So all states of all computations whose initial state is σ_0 also satisfy I. As HS and LR are increasing, this guarantees the correctness of the system.[27]

Classical logic, either propositional or first-order, is adequate for expressing properties of system states, but not for whole computations. A temporal logic can be used for this purpose. It is convenient to adopt a discrete time, so that time steps are computation steps. Temporal logic can be viewed as a modal logic, where the accessibility relation between worlds is in fact the time relation between states. This relation is an ordering, that is, a reflexive, anti-symmetric and transitive relation. Time can be seen as branching, in order to reflect the fact that, due to nondeterminism, some state may have several successors. It is often more convenient to adopt linearly ordered time, which models the fact that, in any computation, every state has one successor. Statements in linear temporal logic should be true for every possible computation. The temporal operator $\Box$ (without subscript; read "always" or "henceforth") is similar to the knowledge operator K_i: $(\Sigma, n) \models \Box p$ means $(\Sigma, n+i) \models p$ for all $i = 0, 1, 2, \ldots$, that is, p is true and remains true forever. The dual operator $\Diamond$ (read "sometimes" or "eventually") is often used: $(\Sigma, n) \models \Diamond p$ means $(\Sigma, n+i) \models p$ for some $i = 0, 1, 2, \ldots$, that is, p will become true at least once, sooner or later; $\Diamond p$ is equivalent to $\neg\Box\neg p$. Another useful operator is $\bigcirc$ (read "next"): $(\Sigma, n) \models \bigcirc p$ means $(\Sigma, n+1) \models p$. There is also the binary operator U (read "until"): $(\Sigma, n) \models p\,U q$ means that $i \geq 0$ exists such that $(\Sigma, n+i) \models q$ and $(\Sigma, n+j) \models p$ for all j such that $0 \leq j < i$. Observe that $\Diamond p$ is equivalent to $true\,U p$. It is easy to see that $\Box\Diamond p$ means "infinitely often p", that is, $(\Sigma, n) \models \Box\Diamond p$ means that infinitely many $i \geq 0$ exist such that $(\Sigma, n+i) \models p$. Similarly $\Diamond\Box p$ means "almost everywhere p", that is, $(\Sigma, n) \models \Diamond\Box p$ means that $(\Sigma, n+i) \models p$ holds for all but finitely many $i \geq 0$. The formula $\Diamond\Box p \supset \Box\Diamond p$ is valid.[28]

We do not comment further about the correctness of system 5, since it cannot be implemented as such. Transitions represent shared actions between the Sender and the Receiver. For instance, the values of both HS and LR are needed to decide which transition can be the next to be executed but, obviously, only the

---

[27]If transmissions often fail, that is, if transitions 2 and 4 are executed more often than transitions 1 and 3, progress will be slow and may even stop. We assume that this will not happen (*fairness* hypothesis).

[28]This is for linear time temporal logic; branching time temporal logic is rather different, although it is also used to specify and verify concurrent systems in a formal way.

Sender knows *HS* and only the receiver knows *LR*. Epistemic logic can be used first to formalize what is known by each process taking part in the computation and secondly to express formal requirements about the whole system and prove whether they are respected or not.

From the epistemic point of view, a Kripke structure can be attached to a concurrent system. The agents are the processes, and also the environment when, as is usually the case, it plays a role in the computation. The states of the structure are the states of the system but it is sometimes necessary to consider not only the system state itself, but also the corresponding computation. Many useful properties of concurrent systems are about computations instead of isolated states. For instance, a statement like "Every message sent by the Sender is correctly received by the Receiver, sooner or later" means that, in every computation $\Sigma = (\sigma_0, \sigma_1, \ldots)$, if at some state σ_i message m has been sent, then there exists $j \geq i$ such that, at state σ_j, the message has been received. States are noted (Σ, i) instead of σ_i or $\Sigma(i)$ when some computation Σ is considered.

Comment. It is not mandatory for Kripke states and system states to be exactly the same. Kripke states can be viewed as valuations for a set of propositions, whose truth values are determined by system states. In this example, the set of propositions could be

$$\{HS = n \,, \ LR = n \,, \ Y[n] = X[n] : n = 0, 1, 2, \ldots\} \,.$$

This set is infinite. Another possibility is to use first-order Kripke structures.

The accessibility relations formalize the fact that each process has access to the value of some variables, but (usually) not to all of them. The *local state* of a process is the part of the system state known to the process. If p is an agent, that is, a process,[29] the ordered pair of states $((\Sigma, j), (\Sigma', j'))$ belongs to the accessibility relation $\mathbf{K}_p$ if the local state of process p is the same in both states. This means that, from the point of view of process p, it is not possible to distinguish between these two states. With this definition accessibility relations are equivalence relations, so the appropriate logic here is $S5_n$.

Let us again consider system (5). It cannot be implemented as such since neither the Sender nor the Receiver can evaluate the truth value of the guard; this can be stated formally:

$$\neg \exists n \exists m [\Box_S (HS = n \ \wedge \ LR = m) \ \vee \ \Box_R (HS = n \ \wedge \ LR = m)] \,.$$

From the intuitive point of view, the Sender has to be able to know which message to send next, so the Receiver has to tell it which message it received last, that is, messages should be acknowledged by the Receiver. Another, equivalent way to state this is, the Sender has to maintain at least an approximate copy of variable *LR*, say *LA* (for *l*ast *a*cknowledged message). A first attempt to use this idea leads to System (6)

---

[29]It is not always necessary to associate an accessibility relation with the environment.

$$
\begin{aligned}
&1 : (LA = HS \;\longrightarrow\; (HS, Y[HS+1]) := (HS+1, X[HS+1]))\,, \\
&2 : (LA = HS \;\longrightarrow\; HS := HS+1)\,, \\
&3 : (LA < HS \;\longrightarrow\; Y[HS] := X[HS])\,, \\
&4 : (LA < HS \;\longrightarrow\; skip)\,, \\
&5 : (Y[LR+1] \neq NIL \;\longrightarrow\; (LA, LR) := (LR+1, LR+1))\,, \\
&6 : (Y[LR+1] \neq NIL \;\longrightarrow\; LR := LR+1)\,, \\
&7 : (Y[LR+1] = NIL \;\longrightarrow\; LA := LR)\,, \\
&8 : (Y[LR+1] = NIL \;\longrightarrow\; skip))\,.
\end{aligned}
\tag{6}
$$

Variable LR has been replaced by the Sender's "local copy" LA in transitions 6.1–4. So the guard can now be evaluated by the Sender. Transitions 6.5–8 deal with reception and acknowledgment. The guard $Y[LR+1] \neq NIL$ indicates correct reception of message number $LR+1$, whereas its negation indicates bad reception or loss of the same message. In case of correct reception, LR is updated. In any case, an acknowledgment has to be sent by the Receiver, but this acknowledgment may be (correctly) received by the Sender (transitions 6.5 and 7) or be corrupted or lost (transitions 6.6 and 8). The invariant of the system is updated into

$$
\begin{aligned}
I : \quad & (LA \leq LR \leq HS \leq LA + 1) \;\wedge \\
& \forall s\,(1 \leq s \leq LR \supset Y[s] = X[s]) \;\wedge \\
& (Y[HS] = X[HS] \;\vee\; Y[HS] = NIL) \;\wedge \\
& \forall s\,(HS < s \supset Y[s] = NIL])\,.
\end{aligned}
$$

This formula has to be satisfied by every reachable state, that is, by every state of every computation starting from an acceptable initial state. An initial state is acceptable if it satisfies $LA = LR = HS = 0$. Furthermore, X denotes an arbitrary stream of messages and Y an empty stream of messages, that is, $Y[n] = NIL$ for all n. Besides, formula I is inductive, that is, the formula $\Box(I \supset \bigcirc I)$ is true for all states. The temporal inference rule

$$
\frac{I \qquad \Box(I \supset \bigcirc I)}{\Box I}
$$

can be used to conclude that I is a safety property of the system, that is, I is true in all reachable states of all computations. This safety property is not enough to ensure a satisfactory behaviour of the system. A liveness property is needed too. Safety properties guarantee that nothing wrong (such as assigning $Y[n]$ an incorrect value) ever happens, whereas liveness properties assert that something good (such as assigning the correct value to $Y[n]$) eventually happens. Insofar as the invariant already expresses that the value of $Y[1 : LR]$ is correct, the only liveness requirement is that LR reach its final value (or grow forever if the stream of messages is not bounded). The *proof graph* represented in Fig. 4 allows us to analyze the liveness requirement. The nodes of the graph are sets of states. Each set is defined by an assertion (in classical logic). The arcs represent moves between sets, which occur when transitions labelling the arcs are executed. A

computation is simply a path in the proof graph, and computations satisfying the liveness requirement correspond exactly to paths leading from A_n to A_{n+1} in finitely many steps, for all n. Otherwise stated, any incorrect computation is stuck forever at some node. For instance, a computation may stay forever in C_n if endless repeated execution of transitions 3 and 4 prevents execution of transition 5 or 6 and therefore prevents further progress in the transmission.

$$A_n : LA = LR = HS = n$$

$$B_n : LA = LR = n \ \wedge \ HS = n+1 \ \wedge \ Y[n+1] = NIL$$

$$C_n : LA = LR = n \ \wedge \ HS = n+1 \ \wedge \ Y[n+1] = X[n+1]$$

$$D_n : LA = n \ \wedge \ LR = HS = n+1$$
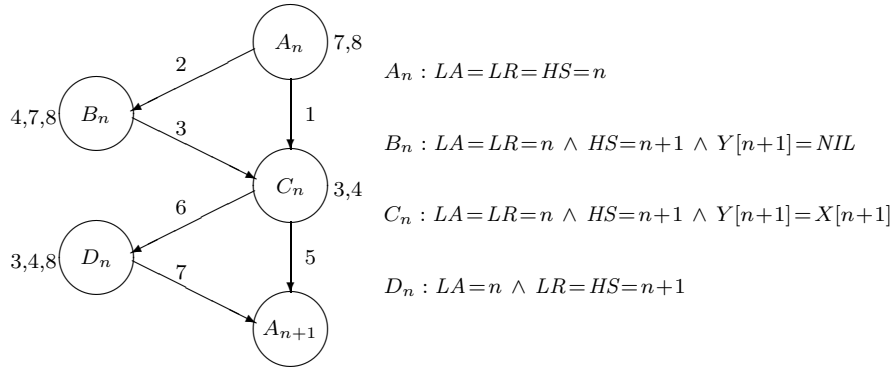
Figure 4. A proof graph

The system is now correct but assumes that successful message transmission from the Sender to the Receiver (transitions 6.1 and 6.3) is synchronous, that is, induces no delay, which is not quite realistic. The same is true for acknowledgment transmission (transitions 6.5 and 6.7). In fact, in order to model asynchronous transmission, these transitions should be broken into a sending part and a receiving part. This leads to a new version of the system:

$$
\begin{aligned}
&1(S): && LA = HS \ \longrightarrow \ (HS, MB) := (HS + 1, (HS + 1, X[HS + 1])), \\
&2(S): && LA \neq HS \ \longrightarrow \ MB := (HS, X[HS]), \\
&3(S): && AB \neq NIL \ \longrightarrow \ (LA, AB) := (AB, NIL), \\
&4(R): && MB \neq NIL \ \longrightarrow \ (Y[MB.1], MB) := (MB.2, NIL), \\
&5(R): && Y[LR + 1] \neq NIL \ \longrightarrow \ (AB, LR) := (LR + 1, LR + 1), \\
&6(R): && Y[LR + 1] = NIL \ \longrightarrow \ AB := LR, \\
&7(E): && MB := NIL, \\
&8(E): && AB := NIL.
\end{aligned}
\tag{7}
$$

It is now possible to assign each transition to the Sender (S), to the Receiver (R) or to the Environment (E). Writing MB (Message Buffer) means transmitting a message (transitions 7.1 and 7.2) and reading it means receiving the message

(transition 7.4). Transition 7.7 models message corruption or loss, by the environment. Acknowledgment is modelled in a similar way (AB is the Acknowledgment Buffer). The invariant is updated into:

$$(LA \leq LR \leq HS \leq LA + 1) \wedge$$
$$\forall s \,(1 \leq s \leq LR \supset Y[s] = X[s]) \wedge$$
$$(Y[HS] = X[HS] \ \vee \ Y[HS] = NIL) \wedge$$
$$(MB = NIL \ \vee \ MB = (HS, X[HS])) \wedge$$
$$(AB = NIL \ \vee \ AB = LR) \wedge$$
$$\forall s \,(HS < s \supset Y[s] = NIL).$$

Knowledge formulas can be used to express that the local state of each process allows transition execution:

$$\Box_S(LA = HS) \ \vee \ \Box_S(LA \neq HS),$$
$$\Box_S(AB = NIL) \ \vee \ \Box_S(AB \neq NIL),$$
$$\Box_R(Y[LR + 1] = NIL) \ \vee \ \Box_R(Y[LR + 1] \neq NIL),$$
$$\Box_R(MB = (HS, X[HS])) \ \vee \ \Box_R(MB = NIL).$$

Comment. System variables, like LA, MB and Y, are not (rigid) logical variables, but (nonrigid) logical constants or functions, so the values attributed to them by state interpretation may vary.

Comment. This program is a variant of the *alternating bit protocol*. Indeed, as only equality or inequality between LA, LR and HS is tested, it is sufficient to record only the last bit of these variables.

Comment. The accessibility relations are now easily characterized. Two states are S-equivalent if they assign the same value to the tuple (LA, HS, AB); they are R-equivalent if they assign the same value to the tuple (MB, LR, Y).

8.2   Knowledge In Asynchronous Message Passing Systems

The epistemic point of view can usefully supplement temporal logic and other formal systems to specify and verify concurrent systems, but it can also give rise to more general results about whole classes of concurrent systems, like those communicating by asynchronous messages. Such a system consists of a finite set of processes.[30] Each process performs three kinds of action: internal actions, that alter only their local state, message sending to another process, message receiving from another process. The local state of a process will be its history, that is, the initial state of the process followed by the list of all actions performed by this process. A process performs at most one action at a time, and it is not a restriction to suppose that only one process at a time actually performs an action. We may also assume that, in every computation, all actions are distinct.[31] There is also a

---

[30]The results to be presented here about knowledge change in asynchronous systems are due to Chandy and Misra [1986].

[31]Internal actions to be repeated become distinct if their occurrences are numbered; for instance, $int(k, \mathbf{a}, i)$ would denote the kth occurrence of internal action $\mathbf{a}$ by process i. Similarly, messages from process j to process i can be numbered.

consistency requirement: if process i receives message μ from process j at time k, that is, if state (Σ, k) is $(\Sigma, k-1)(i).receive(\mu, j, i)$,[32] then for some $\ell < k$ the performed action was the corresponding message transmission from j to i, that is, $send(\mu, i, j)$. The converse is assumed only for reliable message passing system, for which every sent message is eventually received.

Processes perform actions according to their own local state, but through communications receive partial knowledge of the state of the other processes. So an action performed by one process may be the cause of a (later) action performed by another process, provided that these actions are separated by a "message chain". We can speak here of *potential causality*, borrowing the notion from a well known paper due to Lamport [1978]. This induces a partial ordering relation between the actions of any computation c. This potential causality relation is defined in an inductive way. The basic cases are

- $a \xrightarrow{c} a$;

- $a \xrightarrow{c} a'$ if a precedes a' in the history of some process;

- $a \xrightarrow{c} a'$ if a is $send(\mu, j, i)$ and a' is the corresponding $receive(\mu, i, j)$.

The inductive case is just transitive closure, that is

- $a \xrightarrow{c} a'$ if $a \xrightarrow{c} a''$ and $a'' \xrightarrow{c} a'$ for some a''.

As actions occur at most once in a trace c, this relation is antisymmetric. A list of actions $(a_1, \ldots, a_n)$ is an *action chain* for computation c if $a_i \xrightarrow{c} a_{i+1}$ holds for all $i = 1, \ldots, n-1$. If action a_i is performed by process p_i, then the list $(p_1, \ldots, p_n)$ is a *process chain* for computation c; it is a *proper process chain* if $p_i \neq p_{i+1}$ for all $i = 1, \ldots, n-1$. Observe the *message chain theorem*: a proper process chain of length ℓ indicates a message chain whose length is at least $\ell - 1$; indeed, $\ldots, p_i, p_{i+1}, \ldots$ indicates either a direct communication from process p_i to process p_{i+1}, or an indirect one, involving one or more intermediate processes.

It is quite clear that messages convey knowledge and induce state change. Suppose that a computation (or trace) c involves the action chain $(a_1, \ldots, a_n)$, leading from state c_{k_0} to state c_{k_n}. If a process p is not involved in the transition from c_{k_0} to c_{k_n}, then both states are equivalent for this process and the ordered pair (c_{k_0}, c_{k_n}) will be a member of the accessibility relation of process p. In fact, process chains and message transmission are the only means to cut paths in the Kripke structure. If $(p_1, \ldots, p_n)$ is not a process chain, it is always possible to find intermediate states such that the ordered pair $(c_{k_{i-1}}, c_{k_i})$ belongs to the accessibility relation associated with process p_i. Knowledge is the ability to distinguish between system states, and processes gain knowledge only by receiving messages. In a dual way, they lose knowledge by sending message.[33] This *knowledge theorem*

---

[32] The dot denotes concatenation.

[33] This point might seem counter-intuitive, but we already observed it with the alternating bit protocol. For instance, suppose the Sender is in a state where $LA = HS = n$, and therefore $LR = n$; it can send message $X[n+1]$, which leads to a state where $LA = n$, $HS = n+1$ and LR is either n or $n+1$, the Sender will not know the exact value before receiving the acknowledgment.

for asynchronous message passing systems can be stated as follows:

- If $(c, k) \models \neg K_{p_n} \varphi$ and $(c, k') \models K_{p_1} \dots K_{p_n} \varphi$, with $k < k'$, then $(p_k, \dots, p_1)$ is a process chain in computation c.

- If $(c, k) \models K_{p_1} \dots K_{p_n} \varphi$ and $(c, k') \models \neg K_{p_n} \varphi$, with $k < k'$, then $(p_1, \dots, p_k)$ is a process chain in computation c.

It can also be proved that common knowledge cannot be gained or lost in a system communicating by asynchronous message passing, which leads to interesting impossibility results. One of them is about mutual exclusion algorithms, which are of most prominent importance in concurrent programming. The mutual exclusion problem among a family of processes $p_1, p_2, \dots$ occurs when these processes have to share a common resource that can be used by only one process at a time.[34] If we assume that cs_i holds when process p_i is in its critical section, that is, owns the shared resource, the mutual exclusion property is formalized into the assertion

$$\Box \forall i \, \forall j \, [(cs_i \wedge cs_j) \supset i = j].$$

A process knows when it is is in its critical section, so

$$cs_i \supset K_{p_i} cs_i \quad \text{and} \quad \neg cs_i \supset K_{p_i} \neg cs_i$$

are valid formulas. Let us suppose, in some computation c, that process p_i enters its critical section at time τ_i and that process p_j $(j \neq i)$ enters its critical section at time τ_j, with $\tau_j > \tau_i$. We have

$$(c, \tau_i) \models cs_i \wedge K_{p_i} K_{p_j} \neg cs_j$$

since process p_i enters its critical section only when it knows no other process is in its own. Furthermore,

$$(c, \tau_j) \models cs_j \wedge \neg K_{p_j} \neg cs_j$$

since cs_j holds and $\neg K_{p_j} \neg cs_j$ is a logical consequence of cs_j. Due to the knowledge theorem for asynchronous message passing systems, p_i, p_j is a process chain. This generalizes to sequences of n processes accessing their critical section; the message chain theorem shows that such sequences involve at least $n - 1$ messages.

8.3   Some Applications Of Common Knowledge

The fact that common knowledge cannot be obtained or increased with asynchronous message passing seems paradoxical, especially since examples such as the alternating bit protocol suggests otherwise. However, hypotheses about communication reliability are needed to obtain positive results. In the case of the

---

[34] A classical example is a printer used by the user processes in a computer network.

alternating bit protocol, the proof graph indicates how transmission unreliability might prevent progress. It should be emphasized that, when communication happens to have been reliable whereas such reliability was not known (i.e. guaranteed) beforehand, no common knowledge has been obtained. This is illustrated by the example of two generals who know that only a simultaneous attack of the enemy will win the battle. We assume that a general will attack if and only if he knows the other general will attack at the same time, and also that generals do not lie to each other. General A might send a message to general B saying "I will attack at dawn if you do the same"; both generals know that messages can fail to be delivered, so general A will not attack without acknowledgment. But if and when general A receives acknowledgment, general B does not know that his acknowledgment has been delivered, so he does not know that general A will attack, and he will not attack either. Further messages will also fail to provide an agreement based on common knowledge, so a coordinated attack is impossible, unless the generals accept to take some risk, or if some communication reliability is assumed.

The "coordinated attack" problem is mainly concerned with the reliability of communication; in particular, the processes (the generals) are supposed to be reliable. It is also useful to consider the case where communication is reliable (each message sent is received within a finite delay, say one computation step, or even immediately), but processes may fail. Upon failure, a process may omit some or all actions it is supposed to take; in case of "Byzantine" failure, a process may omit actions but also take arbitrary actions.

A standard problem is the agreement problem. All processes have a bit of information, 0 or 1, which is not necessarily the same for all processes. Reliable synchronous communication is available between any pair of processes. A protocol has to be found such that, after finitely many steps, all processes that have not failed decide simultaneously to adopt a common bit, with the restriction that the choice cannot be 1 (resp. 0) if the initial bit of all processes was 0 (resp. 1). This prevents the trivial protocol which would make all processes take an immediate decision, independent from the initial condition.

Suppose first the favourable case, where the initial bit is the same (say 0) for all processes. At the first step of the computation, each process transmits "0" to all other processes. Now, if process p receives "0" from all other processes, it knows that "0" will be the decision. Process p also knows that no process has failed ... before sending its message to p. However, process p cannot exclude the possibility that, say, process q has failed, after sending its message to p, but before sending it to process r. Therefore, process r would not know that the decision should be "0" and would not commit itself (at that time) to a decision. So, process p (nor any other process) will not decide immediately. In the previous problem, guaranteed reliability of communication was needed, otherwise, coordinated attack was impossible, even if, in some computation, all communications succeed. This is the same situation here: even though all processes have correctly sent and received all messages, they do not know it for sure. More specifically, the fact that

all processes have sent and received the bit "0" from each other is not common
knowledge between them. In fact, it can be proved that, if the number of failing
processes is bounded by t, protocols for agreement require $t + 1$ "communication
rounds"; if Byzantine failure is possible, a further limitation is that the number
of reliable processes must be more than double of the number of unreliable ones,
otherwise no solution may exist.

8.4   Knowledge Bases

A knowledge base is a set of facts about the external world. Queries can be asked
to the knowledge base; the answer to the query is "yes" if the query is a logical
consequence. Knowledge bases can be used for several purposes and give rise to
various interesting questions. An important kind of knowledge base is the logic
program. Let us consider briefly the classical example of list concatenation. The
notion of list can be defined in an inductive way: first, `[]` is a list (the empty
list) and, second, if `X` is an object and `Xs` is a list, then `[X|Xs]` is the list whose
first element is `X` and the other elements are those of `Xs`. For instance, the list
whose elements are `a`, `b` and `c` is `[a|[b|[c|[]]]]` which is more conveniently
written as `[a,b,c]`. List concatenation can also be defined in an inductive way.
First, the concatenation of the empty list and any list `Xs` is `Xs` and, second, if
the concatenation of `Xs` and `Ys` is `Zs`, then the concatenation of `[X|Xs]` and `Ys`
is `[X|Zs]`. This can be formalized as (the universal closure of) two Horn clauses,
which will be written in Prolog as the following program:

```
append([],Xs,Xs).
append([X|Xs],Ys,[X|Zs]) :- append(Xs,Ys,Zs).
```

The predicate `append(Xs,Ys,Zs)` is intended to be true when `Xs` and `Ys` are lists
whose concatenation is `Zs`. A specific procedure allows queries to be answered
automatically, in a rather efficient way:

```
    append([a],[b,c],[a,b,c]) ?        yes.
    append([a],[b,c],[b,a,c]) ?        no.
    append([a],[b,c],Xs) ?             Xs = [a,b,c].
    append(Xs,Ys,[a,b,c]) ?            Xs = [], Ys = [a,b,c] ;
                                       Xs = [a], Ys = [b,c] ;
                                       Xs = [a,b], Ys = [c] ;
                                       Xs = [a,b,c], Ys = [] .
```

"Logic programming" and its implementation Prolog are mainly an application of
classical first-order logic, but more epistemic questions do arise. For instance, it
is clear that Prolog has to answer "yes" and give appropriate values to the vari-
ables of the query (if any), if this query is a logical consequence of the program.
For instance, the query `append(Xs,Ys,[a,b,c])` is a logical consequence of the

program if and only if the variables `Xs` and `Ys` are instantiated with appropriate values; exactly four suitable valuations exist, which are the answers given by Prolog. However, it is not easy to decide when Prolog should answer "no". In classical databases, the "Closed World Assumption" (CWA) is frequently used: facts recorded in the database are true, all other facts are false. It is not always desirable to assume that a knowledge base knows everything. Besides, the CWA policy cannot be implemented since first-order logic is undecidable; a weaker policy (the "negation as failure" rule) is implemented instead.

From the epistemic point of view, a logic program is an elementary kind of knowledge base, since it satisfies several restrictions:

1. The language for recording facts is (a fragment of) classical logic;

2. Facts (Prolog clauses) recorded in the knowledge base are about a stable world, that does not change with time;

3. The writer of a logic program knows all the relevant facts;

4. Logic programs contain nothing about their own knowledge; the knowledge operator does not occur in the program, nor in the queries;

5. Only true facts are included in a logic program;

6. No implicit knowledge is assumed about a logic program.

Using epistemic logic (system $S5$) becomes natural when some of these restrictions are relaxed. Let us assume that a knowledge base KB contains propositional facts only, but that the queries may contain the operator K_{KB}. It is easy to associate semantics with this kind of knowledge base such that the answer to the propositional[35] query φ is "yes" if and only if $\mathsf{K}_{KB}\varphi$ holds; otherwise stated, $KB \models \varphi$, that is, φ is the logical consequence of the set of facts contained in KB, if and only if $\mathsf{K}_{KB}\varphi$ holds. The answer to the propositional query φ is "yes" if $KB \models \mathsf{K}_{KB}\varphi$, "no" if $KB \models \mathsf{K}_{KB}\neg\varphi$ and "I don't know" otherwise. So, the semantics tells us when formulas such as $\mathsf{K}_{KB}\varphi$, where φ is propositional, are true or not. An interesting fact about system $S5_1$ is that every formula is logically equivalent to a Boolean combination of formulas of this form, with the consequence that the knowledge base will also handle arbitrary queries, including those about its own knowledge.

Suppose that the (conjunctive) set of facts contained in KB is logically equivalent to the propositional formula φ. It is possible to define a specific $S5_1$-Kripke structure M^{φ} such that the answer to an arbitrary query is "yes" if and only if this query is true in this structure.

In the field of knowledge representation it is of central importance to have a logic which can express "it is only known that φ". Constructing such a logic is

---

[35]In this paragraph, *arbitrary* queries may contain the operator K_{KB}, *propositional* queries may not.

less easy than it might seem. Combining "it is *only* known that φ" with the axiom of negative introspection leads to counter-inuitive results. Consider the inference below:

(1) Only p is known;

(4) hence q is not known;

(5) hence it is known that q is not known [in virtue of the negative introspection axiom].

There is a problem here. We cannot derive $\mathsf{K}\neg\mathsf{K}q$ from only knowing $p \wedge q$ though the latter conjunction intuitively represents more knowledge than only knowing p [van der Hoek *et al.*, 1999, p. 26]. The first formalization of "only knows" is due to J. Halpern and Y. Moses [1985]. The formalization was designed for system $S5$. It takes the notion of minimal model as a primitive notion. In 1999, W. van der Hoek *et al.* offered a general approach to the representation of minimal information for arbitrary normal modal logic which only uses concepts borrowed from standard Kripke possible world semantics. They start by introducing a structural information order over possible worlds and use it to define the notion of *minimal model*. The counter-intuitive result mentioned above is avoided by excluding formulas which represent ignorance [formulas of the form $\neg\mathsf{K}\varphi$]. This is not however an *ad hoc* prohibition. A formal and independent justification is given for preferring a positive information order which preserves positive knowledge [*Ibid.* , pp. 40–45].

8.5 Knowledge-based Programming

Problems in concurrent programming often originate from the need to transmit information from a process to another. In the case of the alternating bit protocol, delays and possible corruption or loss of messages induce a lack of knowledge for the Sender and the Receiver and, quite obviously, they use their partial knowledge in order to select the action to be executed next. This can be expressed in a rather direct way; for instance, the "sending policy" of the Sender is summarized into:

$$1(S): \quad \mathsf{K}_S Y[HS] = X[HS] \; \longrightarrow \; (HS, MB) := (HS + 1, (HS + 1, X[HS + 1])) ,$$
$$2(S): \quad \neg\mathsf{K}_S Y[HS] = X[HS] \; \longrightarrow \; MB := (HS, X[HS]) ,$$

$$(8)$$

Either the Sender knows that the last sent message has been received or it does not. In the first case, it sends the next message; otherwise, it sends the latest message again. This description of the Sender's behaviour is rather abstract since there is no indication about the way the Sender might gain knowledge. The whole development of the alternating bit protocol consists in specifying knowledge gain, with a policy of acknowledgment of the Receiver. The implementation of this policy induces the introduction of LR, which can be seen as a copy of HS, local to the Receiver; similarly, LA is a copy of LR, local to the Sender. It is easy to get an intuitive idea of the meaning of such knowledge-based programs, in which

the guards of the transitions executed by process p may involve the operator K_p.[36] However, the definition of a formal semantics is more difficult. Knowledge-based programs may be ambiguous or inconsistent, even in the case of a single process. For instance, suppose that process p repeatedly executes the transitions

$$1. \quad \mathsf{K}_p(init \vee r) \longrightarrow (init, r) := (false, true),$$
$$2. \quad \neg\mathsf{K}_p(init \vee r) \longrightarrow init := false.$$

If the initial state satisfies $init \wedge \neg r$, then the program may exhibit two very different behaviours; either transition 1 is executed first, and then every subsequent state satisfies r, or transition 2 is executed first, and then every subsequent state satisfies $\neg r$. As far as the guards are mutually exclusive, this nondeterministic behaviour seems puzzling and probably indesirable. The program

$$1. \quad \mathsf{K}_p \square \neg r \longrightarrow r := true,$$
$$2. \quad \neg\mathsf{K}_p \square \neg r \longrightarrow r := r,$$

is even worse, since it cannot be executed from an initial state where r is false. Knowledge-based programs are not really programs, but specifications, which can be ambiguous or inconsistent.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

[Alchourrón et al., 1985] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functors. *The Journal of Symbolic Logic*, 50:510–530, 1985.
[Aumann, 1976] R.J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4:1236–1239, 1976.
[Ayer, 1956] Alfred Jules Ayer. *The Problem of Knowledge*. Macmillan, London, 1956.
[Baltag et al., 1998] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions (extended abstract). In *Proceedings of the int. conference TARK 1998*, pages 43–56. Morgan Kaufmann Publishers, 1998.

---

[36]As we are in $S5_n$, only Boolean combinations of propositions and formulas like $\mathsf{K}_i\varphi$, where φ is propositional, need to be considered.

[Baltag, 1999] Alexandru Baltag. A logic of epistemic actions. In W. van der Hoek, J.J. Meyer, and C. Witteveen, editors, *Foundations and Applications of Collective Agent Based Systems, Lecture notes of the international Summer School ESSLLI 1999*, pages 1–14, Utrecht, 1999.

[Baltag, 2002] Alexandru Baltag. A logic for suspicious players: epistemic actions and belief-updates in games. *Bulletin of Economic Research*, 54:1–45, 2002.

[Barwise, 1981] Jon Barwise. Scenes and other situations. *The Journal of Philosophy*, 78:369–397, 1981.

[Batens *et al.*, 2000] Diderik Batens, Chris Mortensen, Graham Priest, and Jean-Paul van Bendegem (Ed.). *Frontiers of Paraconsistent Logics*. Research Studies Press Ltd, Baldock, 2000.

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

[Boh, 1993] Ivan Boh. *Epistemic Logic in the Later Middle Ages*. Routledge, London, 1993.

[Carlson, 1988] Lauri Carlson. Quantified Hintikka-style epistemic logic. *Synthese*, 74:223–262, 1988.

[Carnap, 1946] Rudolf Carnap. Modalities and quantification. *Journal of Symbolic Logic*, 11:33–64, 1946.

[Castañeda, 1964] Hector-Neri Castañeda. Review of 'knowledge and belief'. *Journal of Symbolic Logic*, 29:132–134, 1964.

[Chandy and Misra, 1986] K.M. Chandy and J. Misra. How processes learn. *Distributed Computing*, 1, 1986.

[Chellas, 1980] Brian F. Chellas. *Modal Logic. An introduction*. Cambridge University Press, Cambridge, 1980.

[Cocchiarella, 1984] Nino B. Cocchiarella. Philosophical perspectives on quantifications in tense and modal logic. In Dov Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic Volume II: Extensions of Classical Logic*, pages 309–353. D. Reidel Publishing Company, Dordrecht, 1984.

[Corsi, 2001] Giovanna Corsi. Counterparts and possible worlds. A study on quantified modal logics. Technical report, Bologna, 2001.

[Cresswell, 1991] M.J. Cresswell. *Entities and Indices*. Kluwer Academic Publishers, Dordrecht, 1991.

[D'Agostino and Gabbay, 1994] M. D'Agostino and D.M. Gabbay. A generalization of analytic deduction via labeled deductive systems. part I: Basic substructural logics. *Journal of Automated Reasoning*, 13:242–281, 1994.

[D'Agostino *et al.*, 1997] M. D'Agostino, D.M. Gabbay, and A. Russo. Grafting modalities onto substructural implication systems. *Studia logica*, VII:1–40, 1997.

[Davidson, 1968 9] Donald Davidson. On saying that. *Synthese*, 19:130–146, 1968–9.

[de Rijke, 1994] Maarten de Rijke. Meeting some neighbours. In Jan van Eijck and Albert Visser, editors, *Logic and Information Flow*, pages 170–195. The MIT Press, Cambridge, Mass., 1994.

[Dretske, 1981] Fred I. Dretske. *Knowledge and the Flow of Information*. Basil Blackwell, Oxford, 1981.

[Dubucs, 1991] Jacques Dubucs. On logical omniscience. *Logique et Analyse, Numéro spécial. International Symposium on Epistemic Logic*, 133-134:41–55, 1991.

[Dubucs, 1997] Jacques Dubucs. Logique, effectivité et faisabilité. *Dialogue*, XXXVI:45–68, 1997.

[Duc, 2001] Ho Ngoc Duc. *Resource-Bounded Reasoning about Knowledge*. PhD thesis, 2001.

[Fagin and Halpern, 1988] Ronald Fagin and Joseph Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[Fagin *et al.*, 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. The MIT Press, Cambridge, Mass., 1995.

[Fitting and Mendelsohn, 1998] Melvin Fitting and Richard L. Mendelsohn. *First-Order Modal Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[Føllesdal, 1967] Dagfinn Føllesdal. Knowledge, identity and existence. *Theoria*, XXXIII:1–27, 1967.

[Frege, 1949] Gottlob Frege. Sense and nominatum. In Herbert Feigl and Wilfrid Sellars, editors, *Readings in Philosophical Analysis*, pages 85–102. Appleton-Century-Crofts, New York, 1949. Translated from "Ueber Sinn and Bedeutung", *Zeitschrift für Philos. und Philos. Kritik*, vol. 100, 1892.

[Gabbay, 1974] Dov M. Gabbay. Tense logics and the tenses of english. In J.M.E. Moravcsik, editor, *Logic and Philosophy for Linguists: A Book of Readings*, pages 177–186. Mouton, The Hague, 1974.

[Gabbay, 1976] Dov Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. D. Reidel Publishing Company, Dordrecht, 1976.

[Geanakoplos, 1992] John Geanakoplos. Common knowledge (paper covering tutorial given in tark 1990). In Yoram Moses, editor, *Proceedings of the fourth conference TARK 1992*, pages 254–315, San Mateo, 1992. Morgan Kaufmann Publishers.

[Genesereth and Nilsson, 1988] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, Palo Alto, 1988.

[Gerbrandy and Groeneveld, 1997] J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, 6:147–169, 1997.

[Gerbrandy, 1999] J. Gerbrandy. *Bisimulations on Planet Kripke*. 1999.

[Gettier, 1963] Edmund L. Gettier. Is justified true belief knowledge? *Analysis*, XXV:121–123, 1963.

[Gillet and Gochet, 1993] Eric Gillet and Paul Gochet. La logique de la connaissance. le problème de l'omniscience logique. *Dialectica*, 47:143–171, 1993.

[Gillet, 2000] Eric Gillet. Essentialisme et identité contingente. In François Beets and Eric Gillet, editors, *Logique en perspective. Mélanges offerts à Paul Gochet*, pages 481–500. Ousia, Bruxelles, 2000.

[Gochet *et al.*, 2000] P. Gochet, P. Gribomont, and A. Thayse. *Logique, Méthodes pour l'intelligence artificielle, tome 3*. Hermès, Paris, 2000.

[Gochet, 2004] Paul Gochet. The dynamic turn in epistemic logic. In Winfried Löffler and Paul Weingartner, editors, *26th International Wittgenstein Symposium 2003*, pages 120–125. Verlag öbv et hpt, Vienna, 2004.

[Grove, 1988] A. Grove. Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–70, 1988.

[Haack, 1978] Susan Haack. *Philosophy of Logics*. Cambridge University Press, Cambridge, 1978.

[Halpern and Moses, 1985] Joseph Halpern and Yoram Moses. Towards a theory of knowledge and ignorance: preliminary report. In K.Apt, editor, *Models of Concurrent Systems*, pages 459–476. Springer-Verlag, Berlin, 1985.

[Halpern and Moses, 1992 first version in 1985] Joseph Y. Halpern and Yoram Moses. A guide to completeness and complexity for propositional logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992 (first version in 1985).

[Halpern and Vardi, 1991] J.Y. Halpern and M.Y. Vardi. Model checking vs. theorem proving: a manifesto. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation [Papers in Honor of John McCarthy]*, pages 151–176. Academic Presss, San Diego, 1991.

[Halpern, 1996] J.Y. Halpern. Should knoweldge entail belief. *Journal of Philosophical Logic*, 25:483–494, 1996.

[Hazen, 1979] Allen Hazen. Counterpart-theoretic semantics for modal logic. *The Journal of Philosophy*, 76:319–338, 1979.

[Hintikka, 1962] Jaakko Hintikka. *Knowledge and Belief. An Introduction to the Logic of the two Notions*. Cornell University Press, Ithaca, New York, 1962. Reissue prepared by V. Hendricks and J. Symons, King's College London Publications, 2005.

[Hintikka, 1966] Jaakko Hintikka. Knowing oneself and other problems in epistemic logic. *Theoria*, 32:1–13, 1966.

[Hintikka, 1970] Jaakko Hintikka. Logic, belief and logical consequence. *Ajatus*, 32:32–47, 1970.

[Hintikka, 1972] Jaakko Hintikka. The semantics of modal notions and the indeterminacy of ontology. In Donald Davidson and Gilbert Harman, editors, *Semantics of Natural language*, pages 398–414. D. Reidel Publishing Company, Dordrecht, 1972.

[Huang and Kwast, 1991] Zhisheng Huang and Karen Kwast. Awareness, negation and logical omniscience. In J. van Eijck, editor, *Lecture Notes in Artificial Intelligence, 478*, pages 282–300, Berlin, 1991. Springer Verlag.

[Hughes and Cresswell, 1984] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic*. Methuen, London, 1984.

[Hughes and Cresswell, 1996] G.E. Hughes and M.J. Cresswell. *A New Introduction to Modal Logic*. Routledge, London, 1996.

[Jackson and Reichgelt, 1989] Peter Jackson and Han Reichgelt. A general proof method for modal predicate logic. In Peter Jackson, Han Reichgelt, and Frank van Harmelen, editors, *Logic-Based Knowledge Representation*, pages 177–218. The MIT Press, Cambridge, Massachusetts, 1989.

[Kaneko and Nagashima, 1996] Mamoru Kaneko and Takashi Nagashima. Game logic and its applications I. *Studia Logica*, 57:325–354, 1996.

[Kaneko and Nagashima, 1997] Mamoru Kaneko and Takashi Nagashima. Game logic and its applications II. *Studia Logica*, 58:273–303, 1997.

[Kaneko *et al.*, 2002] M. Kaneko, T. Nagashima, N.-Y. Suzuki, and Y. Tanaka. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[Kaneko, 1999] Mamoru Kaneko. Common knowledge logic and game logic. *Journal of Symbolic Logic*, 64:685–700, 1999.

[Konolige, 1986] Kurt Konolige. *A Deduction Model of Belief*. Pitman, London, 1986.

[Kraus and Lehmann, 1988] Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. *Theoretical Computer Science*, 58:155–174, 1988.

[Kripke, 1963] Saul A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.

[Kripke, 1979] Saul A. Kripke. A puzzle about belief. In Avishai Margalit, editor, *Meaning and Use*, pages 239–283. D. Reidel Publishing Company, Dordrecht, 1979.

[Kripke, 1980] Saul A. Kripke. *Naming and Necessity*. Basil Blackwell, Oxford, 1980. Revised and enlarged edition. First published in 1972 in G. Harmann and D. Davidson (editors), Semantics of natural Language.

[Ladrière, 1975] Jean Ladrière. Le formalisme et le sens. In Noel Mouloud, editor, *Les langages, le sens et l'histoire (Colloque international de l'Université de Lille III, 1972)*, pages 241–292, Lille, 1975. Publications de l'Université de Lille III.

[Lamport, 1978] Leslie Lamport. Time, clocks and the ordering of events in a distributed system. *Comm. of the ACM*, 21:558–565, 1978.

[Lenzen, 1976] Wolfgang Lenzen. Knowledge, belief, existence, and quantifiers – a note on Hintikka. *Grazer Philosophische Studien*, 2:55–65, 1976.

[Lenzen, 1978] Wolfgang Lenzen. Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:5–219, 1978.

[Lenzen, 1979] Wolfgang Lenzen. Epistemologische Betractungen zu [S4,S5]. *Erkenntnis*, 14:33–56, 1979.

[Lenzen, 1980] Wolfgang Lenzen. *Glauben, Wissen und Wahrscheinlichkeit. Systeme der epistemische Logik*. Springer-Verlag, Wien, 1980.

[Lepage and Lapierre, 2000] François Lepage and Serge Lapierre. *Logique partielle et savoir. Essai de philosophie formelle*. Bellarmin Vrin, Montréal Paris, 2000.

[Levesque, 1984] Hector J. Levesque. A logic of implicit and explicit beliefs. In *Proceedings of the national conference on artificial intelligence*, pages 198–202, Austin Texas, 1984.

[Levi, 1996] Isaac Levi. *For the Sake of the Argument*. Cambridge University Press, Cambridge, 1996.

[Lindström and Rabinowicz, 1999] Sten Lindström and Wlodek Rabinowicz. Unlimited dynamic doxastic logic for introspective agents. *Erkenntnis*, 50:353–385, 1999.

[Lismont and Mongin, 1994a] Luc Lismont and Philippe Mongin. A non-minimal but very weak axiomatization of common belief. *Artificial Intelligence*, 70:363–74, 1994.

[Lismont and Mongin, 1994b] Luc Lismont and Philippe Mongin. On the logic of common belief and common knowledge. *Theory and Decision*, 37:75–106, 1994.

[Lomuscio, 1999] Alessio Lomuscio. *Knowledge Sharing among Ideal Agents*. PhD thesis, 1999.

[Marcus, 1946] Ruth Barcan Marcus. A functional calculus of first order based on strict implication. *The Journal of Symbolic Logic*, 11:1–16, 1946.

[Marcus, 1993] Ruth Barcan Marcus. *Modalities. Philosophical Essays*. Oxford University Press, Oxford, 1993.

[McGee, 1985] V. McGee. A counterexample to modus ponens. *Journal of Philosophy*, 82:462–71, 1985.

[Meyer and van der Hoek, 1995] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995.

[Moore, 1912 ed 1976] G.E. Moore. *Ethics*. Williams and Norgate, London, 1912.

[Moore, 1995] Robert C. Moore. *Logic and Representation*. CSLI Publications, Stanford, 1995.

[Moss, 2002] Larry Moss. Dynamic epistemic logic. Technical report, Indiana, 2002.

[Nerode and Shore, 1993] Anil Nerode and Richard A. Shore. *Logic for Applications*. Springer-Verlag, Berlin, 1993.

[Orilia, 2000] Francesco Orilia. Belief revision and the alethic paradoxes. In A. Chapuis and A. Gupta, editors, *Circularity,Definition and Truth*, pages 273–296. Indian Council of Philosophical Research, New Delhi, 2000.

[Parikh, 2001] Rohit Parikh. Propositions, propositional attitudes and belief revision. In Michael Zakharyaschev, Krister Segerberg, Maarten de Rijke, and Heinrich Wansing, editors, *Advances in Modal Logic, Vol. 2*, pages 381–400. CSLI Publications, Stanford, 2001.

[Pietarinen, 2003] Ahti Pietarinen. Technical report, 2003.

[Quine, 1953; 2nd ed. 1961] Willard Van Orman Quine. *From a Logical Point of View*. Harper and Row, New York, 1953, 2nd ed. 1961.

[Quine, 1981] Willard Van Orman Quine. *Theories and Things*. The Belknap Press of Harvard University Press, Cambridge, Mass., 1981.

[Ramsay, 1988] Allan Ramsay. *Formal Methods in Artificial Intelligence*. Cambridge University Press, Cambridge, 1988.

[Reichenbach, 1947] Hans Reichenbach. *Elements of Symbolic Logic*. McMillan, New York, 1947.

[Rivenc and de Rouilhan, 1997] François Rivenc and Philippe de Rouilhan. *Introduction des Traducteurs (See Carnap, Signification et Nécessité)*. Gallimard, Cambridge, Mass., 1997.

[Ryle, 1967] Gilbert Ryle. Plato. In Paul Edwards, editor, *The Encyclopedia of Philosophy*, pages 314–333. Collier Macmillan Publishers, London, 1967.

[Sahlqvist, 1975] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. In Stig Kanger, editor, *Proceedings of the the Third Scandinavian Logic Symposium. Uppsala 1973*. North Holland, Amsterdam, 1975.

[Scott, 1970] Dana Scott. Advice on modal logic. In Karel Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. D. Reidel Publishing Company, Dordrecht, 1970.

[Segerberg, 1995] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, 3:535–553, 1995.

[Segerberg, 1997] Krister Segerberg. Proposal for a theory of belief revision along the lines of lindström and rabinowicz. *Fundamenta Informaticae*, 32:183–191, 1997.

[Segerberg, 1998] Krister Segerberg. Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39:287–306, 1998.

[Segerberg, 1999] Krister Segerberg. Two traditions in the logic of belief: Bringing them together. In Hans Jürgens Ohlbach and Uwe Reyle, editors, *Logic , Language and Reasoning. Essays in Honour of Dov Gabbay*, pages 135–147. Kluwer Academic Publishers, Dordrecht, 1999.

[Segerberg, 1999] Krister Segerberg. The basic dynamic doxastic logic of AGM. In Mary-Ann Williams and Hans Rott, editors, *Frontiers in Belief Revision*, pages 57–84. Kluwer Academic Publishers, Dordrecht, 2001.

[Simon, 1998] Laurent Simon. *Résolution d'un paradoxe en logique de croyance et de connaissance. MA thesis*. Facultés universitaires Notre Dame de la Paix, Institut d'informatique, 1998.

[Slater, 1994] B.H. Slater. *Intensional Logic*. Aldershot, Avesbury, 1994.

[Spaan, 1993a] E. Spaan. *Complexity of Modal Logics*. PhD thesis, 1993.

[Spaan, 1993b] E. Spaan. The complexity of propositional tense logics. In Maarten de Rijke, editor, *Diamonds and Defaults*, pages 239–252. Kluwer Academic Publishers, Dordrecht, 1993.

[Sturm et al., 2002] Holger Sturm, Frank Wolter, and Michael Zakharyaschev. Common knowledge and quantification. *Economic Theory*, 19:157–186, 2002.

[Tanaka and Ono, 2001] Yoshihito Tanaka and Hiroakira Ono. Rasiowa-Sikorski lemma and Kripke completeness of predicate and infinitary modal logics. In Michael Zakharyaschev, Krister Segerberg, Maarten de Rijke, and Heinrich Wansing, editors, *Advances in Modal Logic, volume 2*, pages 401–419. CSLI Publications, Stanford, 2001.

[Tarski, 1949] Alfred Tarski. The semantic conception of truth. pages 52–84. Appleton-Century-Crofts, New York, 1949. reprinted from *Philosophy and Phenomenological Research*, vol. 6, 1994.

[Thayse, 1989] A. Thayse, editor. *From Modal Logic to Deductive Databases*. Wiley, 1989.

[Thijsse, 1991] Elias C.C. Thijsse. Logics of consciousness explained and compared: Partial approaches to actual belief. *Logique et Analyse*, 133-134, Numéro spécial International Symposium on Epistemic Logic:221–250, 1991.

[Thijsse, 1992] Elias Thijsse. *Partial Logic and Knowledge Representation*. PhD thesis, 1992.

[Toms, 1962] Eric Toms. *Being, Negation and Logic*. Basil Blackwell, Oxford, 1962.

[van Benthem, 1976] Johan van Benthem. *Modal Correspondance Theory*. PhD thesis, 1976.

[van Benthem, 1983] Johan van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.

[van Benthem, 1991] Johan van Benthem. *Languages in Action*. North-Holland, Amsterdam, 1991.

[van der Hoek and Meyer, 1989] W. van der Hoek and J.-J. Ch. Meyer. Possible logics for belief. *Logique et Analyse*, 127-128:177–194, 1989.

[van der Hoek *et al.*, 1999] Wiebe van der Hoek, Jan Jaspars, and Elias Thijsse. Persistence and minimality in epistemic logic. *Annals of Mathematics and Artificial Intelligence*, 27:25–47, 1999.

[van der Hoek, 1993] Wiebe van der Hoek. Systems for knowledge and belief. *Journal of Logic and Computation*, 3:173–195, 1993.

[van Ditmarsch, 2002] Hans P. van Ditmarsch. Descriptions of game actions. *Jl. of Logic, Language and Information*, 11:349–365, 2002.

[van Linder *et al.*, 1995] Bernardus van Linder, Wiebe van der Hoek, and John-Jules Meyer. Actions that make you change your mind. In Armin Laux and Heinrich Wansing, editors, *Knowledge and Belief in Philosophy and Artificial Intelligence*, pages 103–146. Akademie Verlag, Berlin, 1995.

[Veltman, 1996] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 26:221–261, 1996.

[Viganò, 2000] Luca Viganò. *Labelled Non-Classical Logics*. Kluwer Academic Publishers, Dordrecht, 2000.

[Voorbraak, 1992] Frans Voorbraak. Generalized Kripke models for epistemic logic. pages 214–228, San Mateo, 1992. Morgan Kaufmann Publishers.

[Voorbraak, 1993] Frans Voorbraak. *As Far as I Know. Epistemic Logic and Uncertainty*. PhD thesis, 1993.

[Wassermann, 1999] Renata Wassermann. Resource-bounded belief revision. *Erkenntnis*, 50:429–446, 1999.

[Wassermann, 2000] Renata Wassermann. *Resource-Bounded Belief Revision*. PhD thesis, 2000.

[Wassermann, 2001] Renata Wassermann. On structured belief bases. In Mary-Anne Williams and Hans Rott, editors, *Frontiers in Belief Revision*, pages 349–367. Kluwer Academics Publishers, Dordrecht, 2001.

[Williamson, 2000] Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

[Williamson, 2001] Timothy Williamson. Some philosophical aspects of reasoning about knowledge. In Johan van Benthem, editor, *Proceedings of the 8th conference TARK 2001*, pages 97–97. Morgan Kaufmann Publishers, San Francisco, 2001.

[Wiśniewski, 1998] Andrzej Wiśniewski. Two logics of occurrent belief. *Acta Universitatis Wratislaviensis No 2023*, 18:115–121, 1998.

[Wolter, 2000] Frank Wolter. Fragments of first-order common knowledge logics. *Studia logica*, 65:249–271, 2000.

This Page is Intentionally Left Blank

DEONTIC LOGIC*

Paul McNamara

INTRODUCTION

Introductory Note: Items boxed off in the text can be skipped without loss of continuity. Similarly for the four appendices to which the reader is optionally directed at appropriate places in the main essay.

Deontic logic[1] is that branch of symbolic logic that has been the most concerned with the contribution that the following notions make to what follows from what:

permissible (permitted)	must
impermissible (forbidden, prohibited)	supererogatory (beyond the call of duty)
obligatory (duty, required)	indifferent / significant
gratuitous (non-obligatory)	the least one can do
optional	better than / best / good / bad
ought	claim / liberty / power / immunity.

indexobligatory

To be sure, some of these notions have received more attention in deontic logic than others. However, virtually everyone working in this area would see systems designed to model the logical contributions of these notions as part of deontic logic proper.

As a branch of symbolic logic, deontic logic is of *theoretical* interest for some of the same reasons that modal logic is of theoretical interest. However, despite the fact that we need to be cautious about making too easy a link between deontic logic and practicality, many of the notions listed *are* typically employed in attempting to regulate and coordinate our lives together (but also to evaluate states of affairs). For these reasons, deontic logics often directly involve topics of considerable practical significance such as morality, law, social and business organizations (their

---

* At the invitation of the editors of this series, this essay is a minor adaptation of McNamara [2005]. Thus it is primarily systematic, but with historical information weaved in throughout, especially in notes on the literature associated with a problem or a development.

[1]The term "*deontic* logic" appears to have arisen in English as the result of C. D. Broad's suggestion to von Wright [1951]; Mally used "Deontik" earlier to describe his work [Mally, 1926]. Both terms derive from the Greek term, $\delta\varepsilon o\nu$, for 'that which is binding', and $\iota\kappa$, a common Greek adjective-forming suffix for 'after the manner of', 'of the nature of', 'pertaining to', 'of', thus suggesting roughly the idea of a logic of duty. (The intervening τ in $\delta\varepsilon o\nu\tau\iota\kappa$ is inserted for phonetic reasons.)

norms, as well as their normative constitution), and security systems. To that extent, studying the logic of notions with such *practical* significance perhaps adds some practical significance to deontic logic itself.

On Defining Deontic Logic: Defining a discipline or area within one is often difficult. Deontic logic is no exception. Standard characterizations of deontic logic are arguably either too narrow or too wide. Deontic logic is often glossed as the logic of obligation, permission, and prohibition, but this is too narrow. For example, it would exclude a logic of supererogation as well as any non-reductive logic for legal notions like claims, liberties, powers, and immunities from falling within deontic logic. On the other hand, we might say that deontic logic is that branch of symbolic logic concerned with the logic of normative expressions: a systematic study of the contribution these expressions make to what follows from what. This is better in that it does not appear to be too exclusive, but it is arguably too broad, since deontic logic is not traditionally concerned with the contribution of every sort of normative expression. For example, "credible" and "dubious" are normative expressions, as are "rational" and "prudent" but these two pairs are not normally construed as within the purview of deontic logic (as opposed to say epistemic logic, and rational choice theory, respectively). Nor would it be enough to simply say that the normative notions of deontic logic are always practical, since the operator "it ought to be the case that", perhaps the most studied operator in deontic logic, appears to have no greater intrinsic link to practicality than does "credible" or "dubious". The following seem to be without practical import: "It ought to be the case that early humans did not exterminate Neanderthals."[2] Perhaps a more refined link to practicality is what separates deontic logic from epistemic logic, but this doesn't help distinguish it from rational choice theory, the latter being concerned with collective practical issues as well as individual ones. Perhaps there is no non-ad hoc or principled division between deontic logic and distinct formal disciplines focused on the logic of other normative expressions, such as epistemic logic and rational choice theory. These are interesting and largely unstudied meta-philosophical issues that we cannot settle here. Instead we have defined deontic logic contextually and provisionally.

This essay is divided into four main parts. The first provides preliminary background. The next two parts provide an introduction to the most standard monadic systems of deontic logic The fourth, and by far the largest, section is dedicated to

---

[2]Although this example has no practical significance for us, it is still true that without such capacities for counterfactual *evaluation*, we would have no capacity for such deeply human traits as a sense of tragedy and misfortune, and of course some judgments about what ought to be the case do and should guide our actions, but the link is not simple, and it is not clear that such evaluations of states of affairs are any less a part of deontic logic than evaluations of the future courses of action of agents.

various problems and challenges faced by the standard systems. This reflects the fact that the challenges posed to these standard systems are numerous.

1 INFORMAL PRELIMINARIES AND BACKGROUND

Deontic logic has been strongly influenced by ideas in modal logic. Analogies with alethic modal notions and deontic notions were noticed as far back as the fourteenth century, where we might say that the rudiments of modern deontic logic began [Knuuttila, 1981]. Although informal interest in what can be arguably called aspects of deontic logic continued, the trend toward studying logic using the *symbolic and exact* techniques of mathematics became dominant in the twentieth century, and logic became largely, *symbolic* logic. Work in twentieth century symbolic modal logic provided the explicit impetus for von Wright [1951], the central early figure in the emergence of deontic logic as a full-fledged branch of *symbolic* logic in the twentieth century. So we will begin by gently noting a few folk-logical features of alethic modal notions, and giving an impressionistic sense of how natural it was for early developments of deontic logic to mimic those of modal logic. We will then turn to a more direct exploration of deontic logic as a branch of symbolic logic.

However, before turning to von Wright, and the launching of deontic logic as an on-going active academic area of study, we need to note that there was a significant earlier episode, Mally [1926], that did not have the influence on symbolic deontic logic that it might have, due at least in part, to serious technical problems. The most notable of these problems was the provable equivalence of what ought to be the case (his main deontic notion) with what is the case, which is plainly self-defeating for a deontic logic. Despite the problems with the system he found, Mally was an impressive pioneer of deontic logic. He was apparently uninfluenced by, and thus did not benefit from, early developments of alethic modal logic. This is quite opposed to the later trend in the 1950s when deontic logic reemerged, this time as a full-fledged discipline, deeply influenced by earlier developments in alethic modal logic. Mally was the first to found deontic logic on the syntax of propositional calculus explicitly, a strategy that others quickly returned to after a deviation from this strategy in the very first work of von Wright. Mally was the first to employ deontic constants in deontic logic (reminiscent of Kanger and Anderson's later use of deontic constants, but without their "reduction"; more below). He was also the first to attempt to provide an integrated account of non-conditional and conditional ought statements, one that provided an *analysis* of conditional 'ought's via a monadic deontic operator coupled with a material conditional (reminiscent of similar failed attempts in von Wright [1951] to analyze the dyadic notion of commitment), and that allowed for a form of factual detachment (more below). All in all, this seems to be a remarkable achievement in retrospect. For more information on Mally's system, including a diagnosis of the source of his main technical problem, and a sketch of one way he might have avoided it, see the easily accessible Lokhorst [2004].

1.1   Some Informal Rudiments of Alethic Modal Logic

Alethic modal logic is roughly the logic of necessary truth and related notions. Consider five basic alethic modal statuses, expressed as sentential operators — constructions that, when applied to a sentence, yield a sentence (as does "it is not the case that"):

> it is necessary (necessarily true) that ($\Box$)
> it is possible that ($\Diamond$)
> it is impossible that
> it is non-necessary that
> it is contingent that.[3]

Although all of the above operators are generally deemed definable in terms of any one of the first four, the necessity operator is typically taken as basic and the rest defined accordingly:

> It is possible that $p(\Diamond p) =_{df} \sim\Box \sim p$
> It is impossible that $p =_{df} \Box \sim p$
> It is non-necessary that $p =_{df} \sim\Box p$
> It is contingent that $p =_{df} \sim\Box p \& \sim\Box \sim p$.

It is routinely assumed that the following threefold partition of propositions holds:
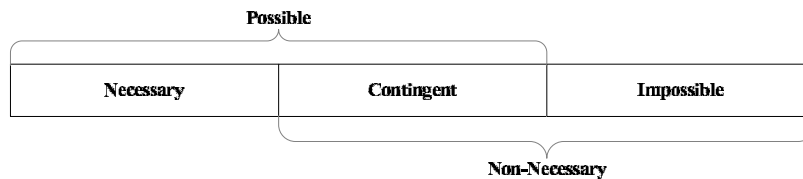
The three rectangular cells are jointly exhaustive and mutually exclusive: every proposition is either necessary, contingent, or impossible, but no proposition is more than one of these. The possible propositions are those that are either necessary or contingent, and the non-necessary propositions are those that are either impossible or contingent.

Another piece of folk logic for these notions is the following modal square of opposition:

---

[3]In keeping with very wide trends in logic over the past century or so, we will treat both modal notions and deontic notions as sentential (or propositional) operators unless otherwise stated. Although it is controversial whether the most fundamental (if there are such) modal and deontic notions have the logical form of propositional operators, focusing on these forms allowed for essentially seamless integration of these logics with propositional logics.
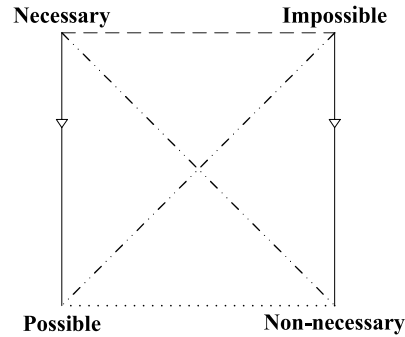
Arrowed Lines: represent implications   **Dotted Line**: connects sub-contraries.
Dashed Line: connects contraries.   **Dotted-Dashed Lines**: connect contradictories.[4]

Furthermore it is generally assumed that the following hold:

If $\Box p$ then p (if it is necessary that p, then p is true).
If p then $\Diamond p$ (if p is true, then p is possible).

These reflect the idea that we are interested here in alethic (and thus truth-implicating) necessity and its siblings.

We now turn to some of the analogies involved in what is a corresponding bit of deontic folk logic: "The Traditional Scheme" [McNamara, 1990; 1996a]. This is a minor elaboration of what can be found in [von Wright, 1953] and [Prior, 1962, [1955]].

1.2 The Traditional Scheme and the Modal Analogies

The five normative statuses of the Traditional Scheme are:[5]

---

[4]This key will be relied on throughout for similar diagrams. Recall that propositions are *contraries* if they can't both be true, *sub-contraries* if they can't both be false, and *contradictories* if they always have opposing truth-values. The square can be easily augmented as a hexagon by including nodes for contingency [McNamara, 1996a]. Cf. the deontic hexagon below.

[5]Only deontic operators will appear in boldface. These abbreviations are not standard. O is routinely used instead of **OB**, and O is often read as "It ought to be the case that". P is used instead of **PE**, and if used at all, F (for "forbidden") instead of **IM** and I (for "indifference") instead of **OP**. Deontic non-necessity, here denoted by **GR** is seldom ever named, and even in English it is hard to find a term for this condition. The double letter choices used here are easy mnemonics expressing *all five* basic conditions (which, from a logical standpoint, are on a par), and they will facilitate later discussion involving just what notions to take SDL and kin to be modeling, and how it might be enriched to handle other related normative notions. Both deontic logic and ethical theory is fraught with difficulties when it comes to interchanging allegedly equivalent expressions for one another. Here we choose to read the basic operator as "it is obligatory that" so that all continuity with permissibility, impermissibility, and optionality

> it is obligatory that (**OB**)
> it is permissible that (**PE**)
> it is impermissible that (**IM**)
> it is gratuitous that (**GR**)
> it is optional that (**OP**).

The first three are familiar, but the fourth is widely ignored, and the fifth has regularly been conflated with "it is a matter of *indifference* that p" (by being defined in terms of one of the first three), which is not really part of the traditional scheme (more below). Typically, one of the first two is taken as basic, and the others defined in terms of it, but any of the first four can play the same sort of purported defining role. The most prevalent approach is to take the first as basic, and define the rest as follows:

$$\textbf{PE}p \leftrightarrow \ \sim\!\textbf{OB}\sim p$$
$$\textbf{IM}p \leftrightarrow \textbf{OB}\sim p$$
$$\textbf{GR}p \leftrightarrow \ \sim\!\textbf{OB}p$$
$$\textbf{OP}p \leftrightarrow (\sim\!\textbf{OB}p\,\&\,\sim\!\textbf{OB}\sim p).^{6}$$

These assert that something is permissible iff (if and only if) its negation is not obligatory, impermissible iff its negation is obligatory, gratuitous iff it is not obligatory, and optional iff neither it nor its negation is obligatory. Call this "The Traditional *Definitional* Scheme (TDS)". If one began with **OB** alone and considered the formulas on the right of the equivalences above, one could easily be led to consider them as at least candidate defining conditions for those on the left. Although not uncontestable, they are natural, and this scheme is still widely employed. Now if the reader looks back at our use of the necessity operator in defining the remaining four alethic modal operators, it will be clear that that definitional scheme is perfectly analogous to the deontic one above. From the formal standpoint, the one is merely a syntactic variant of the other: just replace **OB** with $\Box$, **PE** with $\Diamond$, etc.

In addition to the TDS, it was traditionally assumed that the following, call it "The Traditional Threefold Classification (TTC)" holds:

---
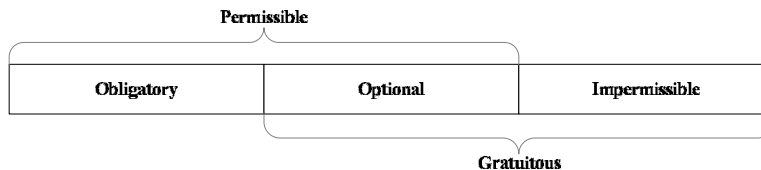
is not lost, as it would be with the "it ought to be the case that" reading [McNamara, 1996c]. A choice must be made. "It is obligatory that" may also be read personally, but non-agentially as "it is obligatory for Jones that" [Krogh and Herrestad, 1996; McNamara, 2004a]. We will return to these issues again below.

[6]In this essay we will generally call such equivalences "definitions", sloughing over the distinction between abbreviatory definitions of operators not officially in the formal language, and axiom systems with languages containing these operators, and axioms directly encoding the force of such definitions as equivalences.

Here too, all propositions are divided into three jointly exhaustive and mutually exclusive deontic classes: every proposition is obligatory, optional, or impermissible, but no proposition falls into more than one of these three categories. Furthermore, the permissible propositions are those that are either obligatory or optional, and the gratuitous propositions are those that are impermissible or optional. The reader can easily confirm that this natural scheme is also perfectly analogous to the threefold classification we gave above for the alethic modal notions.

Furthermore, "The Deontic Square (DS)" is part of the Traditional Scheme:
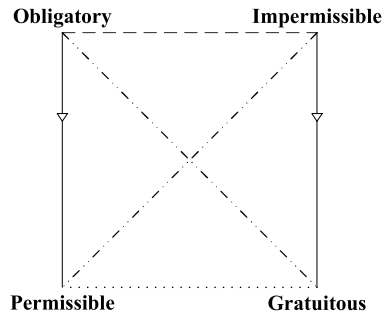
The logical operators at the corners are to be interpreted as in the modal square of opposition. The two squares are plainly perfectly analogous as well. If we weave in nodes for optionality, and shift to formuli, we get a deontic hexagon:
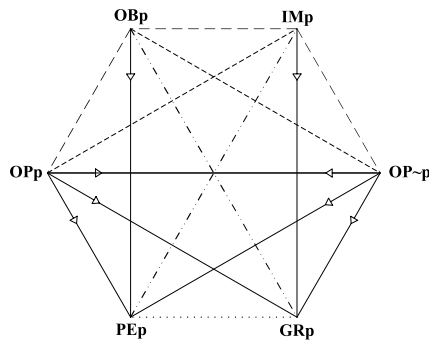
Given these correspondences, it is unsurprising that our basic operator, read here as "it is obligatory that", is often referred to as "deontic *necessity*". However, there are also obvious dis-analogies. Before, we saw that these two principles are part of the traditional conception of alethic modality:

> If $\Box p$ then p (if it is necessary that p, then p is true).
> If p, then $\Diamond p$ (if p is true, then it is possible).

But their deontic analogs are:

> If **OB**p then p (if it is obligatory that p, then p is true).
> If p, then **PE**p (if p is true, then it is permissible).

The latter two are transparently false, for obligations can be violated, and impermissible things do happen.[7] However, as researchers turned to generalizations of alethic modal logic, they began considering wider classes of modal logics, including ones where the necessity operator was not truth-implicating. This too encouraged seeing deontic necessity, and thus deontic logic, as falling within modal logic so-generalized, and in fact recognizing possibilities like this helped to fuel the generalizations of what began with a focus on alethic modal logic [Lemmon, 1957; Lemmon and Scott, 1997].

1.3 Toward Deontic Logic Proper

It will be convenient at this point to introduce a bit more regimentation. Let's assume that we have a simple propositional language with the usual suspects, an infinite set of propositional variables (say, $P_1, \ldots, P_n, \ldots$) and complete set of truth-functional operators (say, $\sim$ and $\rightarrow$), as well as the one-place deontic operator, **OB**.

Deontic Wffs: Here is a more formal definition. Suppose that we have:

A set of Propositional Variables (PV): $P_1, \ldots, P_i, \ldots$ — where "i" is a numerical subscript; three propositional operators: $\sim, \rightarrow,$ **OB**; and a pair of parentheses: (,).

The set of D-wffs (deontic well-formed formuli) is then the smallest set satisfying the following conditions (lower case "p" and "q" are metavariables):

  FR1. PV is a subset of D-wffs.
  FR2. For any p, p is in D-wffs only if $\sim p$ and **OB**p are also in D-wffs.
  FR3. For any p and q, p and q are in D-wffs only if $(p \rightarrow q)$ is in D-wffs.

We then assume the following abbreviatory definitions:

  DF1-3. $\&, \vee, \rightarrow$ as usual.
  DF4. **PE**$p =_{df} \sim$**OB** $\sim p$.
  DF5. **IM**$p =_{df}$ **OB** $\sim p$.
  DF6. **GR**$p =_{df} \sim$**OB**p.
  DF7. **OP** $=_{df}$ ($\sim$**OB**p & $\sim$**OB** $\sim p$).

---

[7]The logic of [Mally, 1926] was saddled with the T-analog above. Mally reluctantly embraced it since it seemed to follow from premises he could find no fault with. See [Lokhorst, 2004].

Unless otherwise stated, we will only be interested in deontic logics that contain classical propositional calculus (PC). So let's assume we add that as the first ingredient in specifying any deontic logic, so that, for example, $\mathbf{OB}p \rightarrow \sim\sim\mathbf{OB}p$, can be derived in any system to be considered here.

Above, in identifying the Traditional Definitional Scheme, we noted that we could have taken any of the first four of the five primary normative statuses listed as basic and defined the rest in terms of that one. So we want to be able to generate the corresponding equivalences derivatively from the scheme we did settle on, where $\mathbf{OB}$ is basic. But thus far we cannot. For example, it is obviously desirable to have $\mathbf{OB}p \rightarrow \sim\mathbf{PE} \sim p$ as a theorem from the traditional standpoint. After all, this wff merely expresses one half of the equivalence between what would have been *definiens* and *definiendum* had we chosen the alternate scheme of definition in which "$\mathbf{PE}$" was taken as basic instead of "$\mathbf{OB}$". However, $\mathbf{OB}p \rightarrow \sim\mathbf{PE} \sim p$ is not thus far derivable. For $\mathbf{OB}p \rightarrow \sim\mathbf{PE} \sim p$ is definitionally equivalent to $\mathbf{OB}p \rightarrow \sim\sim\mathbf{OB} \sim\sim p$, which reduces by PC to $\mathbf{OB}p \rightarrow \mathbf{OB} \sim\sim p$, but the latter formula is not tautological, so we cannot complete the proof. So far we have deontic wffs and propositional logic, but no deontic logic. For that we need some distinctive principles governing our deontic operator, and in particular, to generate the alternative equivalences that reflect the alternative definitional schemes alluded to above, we need what is perhaps the most fundamental and least controversial rule of inference in deontic logic, and the one characteristic of "classical modal logics" [Chellas, 1980]:

$\quad\quad\quad$ $\mathbf{OB}$-RE: If $p \leftrightarrow q$ is a theorem, then so is $\mathbf{OB}p \leftrightarrow \mathbf{OB}q$.

This rule tells us that if two formulas are provably equivalent, then so are the results of prefacing them with our basic operator, $\mathbf{OB}$. With its aid (and the Traditional Definitional Scheme's), it now easy to prove the equivalences corresponding to the alternative definitional schemes. For example, since $\vdash p \leftrightarrow \sim\sim p$, by $\mathbf{OB}$-RE, we get $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB} \sim\sim p$, i.e. $\vdash \mathbf{OB}p \leftrightarrow \sim\sim\mathbf{OB} \sim\sim p$, which generates $\vdash \mathbf{OB}p \leftrightarrow \sim\mathbf{PE} \sim p$, given our definitional scheme. To the extent that the alternative definitional equivalences are supposed to be derivable, we can see RE as presupposed in the Traditional Scheme.

All systems we consider here will contain RE (whether as basic or derived). They will also contain, unless stated otherwise, one other principle, a thesis asserting that a logical contradiction (conventionally denoted by "$\perp$") is always gratuitous:

$\quad\quad\quad$ OD : $\sim\mathbf{OB} \perp$.

So, for example, OD implies that it is a logical truth that it is not obligatory that my taxes are paid and not paid. Although OD is not completely uncontestable,[8]

---

[8]If Romeo solemnly promised Juliet to square the circle did it thereby become obligatory that he do so?

it is plausible, and like RE, has been pervasively presupposed in work on deontic logic. In this essay, we will focus on systems that endorse both RE and OD.

Before turning to our first full-fledged system of deontic logic, let us note one very important principle that is not contained in all deontic logics, and about which a great deal of controversy in deontic logic and in ethical theory has transpired.

1.4   The Fundamental Presupposition of the Traditional Scheme

Returning to the Traditional Scheme for a moment, its Threefold Classification, and Deontic Square of Opposition can be expressed formally as follows:

DS: $(\mathbf{OB}p \leftrightarrow \sim\mathbf{GR}p)\,\&\,(\mathbf{IM}p \leftrightarrow \sim\mathbf{PE}p)\,\&\, \sim(\mathbf{OB}p\,\&\,\mathbf{IM}p)\,\&$
$\sim(\sim\mathbf{PE}p\,\&\, \sim\mathbf{GR}p)\,\&\,(\mathbf{OB}p \rightarrow \mathbf{PE}p)\,\&\,(\mathbf{IM}p \rightarrow \mathbf{GR}p).$

TTC: $(\mathbf{OB}p \vee \mathbf{OP}p \vee \mathbf{IM}p)\,\&\,[\sim(\mathbf{OB}p\,\&\,\mathbf{IM}p)\,\&\, \sim(\mathbf{OB}p\,\&\,\mathbf{OP}p)\,\&$
$\sim(\mathbf{OP}p\,\&\,\mathbf{IM}p)].$

Given the Traditional Definitional Scheme, it turns out that DS and TTC are each *tautologically* equivalent to the principle that obligations cannot conflict (and thus to one another):

NC: $\sim(\mathbf{OB}p\,\&\,\mathbf{OB} \sim p).$[9]

So the Traditional Scheme rests squarely on the soundness of NC (and the traditional definitions of the operators). Indeed, the Traditional Scheme is nothing other than a disguised version of NC, given the definitional component of that scheme.

NC is not to be confused in content with the previously mentioned principle, OD($\sim\mathbf{OB} \perp$). OD asserts that no single logical contradiction can be obligatory, whereas NC asserts that there can never be two things that are each separately obligatory, where the one obligatory thing is the negation of the other. The presence or absence of NC arguably represents one of the most fundamental divisions among deontic schemes. As, until recently, in modern normative ethics (see [Gowans, 1987]), early deontic logics presupposed this thesis. Before turning to challenges to NC, we will consider a number of systems that endorse it, beginning with what has come to be routinely called "Standard Deontic Logic", the benchmark system of deontic logic.

---

[9]For DS becomes $(\mathbf{OB}p \leftrightarrow \sim\sim\mathbf{OB}p)\,\&\,(\mathbf{OB} \sim p \leftrightarrow \sim\sim\mathbf{OB} \sim p)\,\&\, \sim(\mathbf{OB}p\,\&\, \mathbf{OB} \sim p)\,\&$
$\sim(\sim\sim\mathbf{OB} \sim p\,\&\, \sim\sim\mathbf{OB}p)\,\&\,(\mathbf{OB}p \rightarrow\sim\mathbf{OB} \sim p)\,\&\,(\mathbf{OB} \sim p \rightarrow\sim\mathbf{OB}p)$, and although the first two conjuncts are tautologies, the remaining four are each tautologically equivalent to NC above. Similarly, TTC becomes $(\mathbf{OB}p \vee (\sim\mathbf{OB}p\,\&\, \sim\mathbf{OB} \sim p) \vee \mathbf{OB} \sim p)\,\&\,[\sim(\mathbf{OB}p\,\&\,\mathbf{OB} \sim p)\,\&\, \sim$ $(\mathbf{OB}p\,\&\,(\sim\mathbf{OB}p\,\&\, \sim\mathbf{OB} \sim p))\,\&\, \sim((\sim\mathbf{OB}p\,\&\, \sim\mathbf{OB} \sim p)\,\&\,\mathbf{OB} \sim p)]$, and the exhaustiveness clause is tautological, as are the last two conjuncts of the exclusiveness clause, but the first conjunct of that clause is just NC again. Likewise for the assumptions that *the gratuitous* is the disjunction of *the permissible* and *the obligatory* and that *the permissible* is the disjunction of *the obligatory* and *the optional*. (See [McNamara, 1996a, pp. 422–46].)

2 STANDARD DEONTIC LOGIC

2.1 Standard Syntax

Standard Deontic Logic (SDL) is the most cited and studied system of deontic logic, and one of the first deontic logics axiomatically specified. It builds upon propositional logic, and is in fact essentially just a distinguished member of the most studied class of modal logics, "normal modal logics". It is a monadic deontic logic, since its basic deontic operator is a one-place operator (like $\sim$, and unlike $\rightarrow$): syntactically, it applies to a *single* sentence to yield a compound sentence.[10]

Assume again that we have a language of classical propositional logic with an infinite set of propositional variables, the operators $\sim$ and $\rightarrow$, and the operator, **OB**. SDL is then often axiomatized as follows:

SDL:	A1.	All tautologous wffs of the language	(TAUT)
	A2.	$\mathbf{OB}(p \rightarrow q) \rightarrow (\mathbf{OB}p \rightarrow \mathbf{OB}q)$	(**OB**-K)
	A3.	$\mathbf{OB}p \rightarrow \sim\mathbf{OB} \sim p$	(**OB**-D)
	MP.	If $\vdash p$ and $\vdash p \rightarrow q$ then $\vdash q$	(MP)
	R2.	If $\vdash p$ then $\vdash \mathbf{OB}p$	(**OB**-NEC).[11]

SDL is just the normal modal logic "D" or "KD", with a suggestive notation expressing the intended interpretation.[12] TAUT is standard for normal modal systems. **OB**-K, which is the K axiom present in all normal modal logics, tells us that if a material conditional is obligatory, and its antecedent is obligatory, then so is its consequent.[13] **OB**-D tells us that p is obligatory only if its negation isn't. It is just "No Conflicts" again, but it is also called "D" (for "Deontic") in normal modal logics. MP is just Modus Ponens, telling us that if a material conditional and its antecedent are theorems, then so is the consequent. TAUT combined with MP gives us the full inferential power of the Propositional Calculus (often referred to, including here, as "PC"). As noted earlier, PC has no distinctive deontic import. **OB**-NEC tells us that if anything is a theorem, then the claim that that thing is obligatory is also a theorem. Note that this guarantees that something is always obligatory (even if only logical truths).[14] Each of the distinctively deontic principles, **OB**-K, **OB**-D, and **OB**-NEC are contestable, and we will consider criticisms of them shortly. However, to avoid immediate confusion for those new to deontic logic, it is perhaps worth noting that **OB**-NEC is generally deemed a

---

[10]In a monadic system one can easily *define* dyadic deontic operators *of sorts* [Hintikka, 1971]. For example, we might define "deontic implication" as follows: $p {}_d\!\rightarrow q =_{df} \mathbf{OB}(p \rightarrow q)$. We will consider non-monadic systems later on.

[11]"$\vdash$" before a formula indicates it is a theorem of the relevant system.

[12]Note that this axiomatization, and all others here, use "axiom schema": schematic specifications by syntactic pattern of classes of axioms (rather than particular axioms generalized via a substitution rule). We will nonetheless slough over the distinction here.

[13]It is also justifies a version of *Deontic Detachment*, from **OB**p and **OB**$(p \rightarrow q)$ derive **OB**q, an inference pattern to be discussed later.

[14]Compare the rule that contradictions are not permissible: if $\vdash\sim p$ then $\vdash\sim\mathbf{PE}p$. *R2* is often said to be equivalent to "not everything is permissible", and thus to rule out only "normative systems" that have no normative force at all.

convenience that, among other things, assures that SDL is in fact just one of the well-studied normal modal logics with a deontic interpretation. Few have spilled blood to defend its cogency substantively, and these practical compromises can be strategic, especially in early stages of research.

Regarding SDL's expressive powers, advocates typically endorse the Traditional Definitional Scheme noted earlier. Below we list some theorems and two important derived rules of SDL.[15]

$\mathbf{OB}\top$	($\mathbf{OB}$-N)
$\sim\mathbf{OB}\bot$[16]	($\mathbf{OB}$-OD)
$\mathbf{OB}(p \,\&\, q) \rightarrow (\mathbf{OB}p \,\&\, \mathbf{OB}q)$	($\mathbf{OB}$-OD)
$(\mathbf{OB}p \,\&\, \mathbf{OP}q) \rightarrow \mathbf{OB}(p \,\&\, q)$	($\mathbf{OB}$-C / Aggregation)
$\mathbf{OB}p \vee \mathbf{OP}p \vee \mathbf{IM}p$	($\mathbf{OB}$-Exhaustion)
$\mathbf{OB}p \rightarrow\, \sim\mathbf{OB} \sim p$	($\mathbf{OB}$-NC or $\mathbf{OB}$-D)
If $\vdash p \rightarrow q$ then $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$	($\mathbf{OB}$-RM)
If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB}q$	($\mathbf{OB}$-RE)

We will be discussing a number of these subsequently. For now, let's briefly show that RM is a derived rule of SDL. We note some simple corollaries as well.

Show
If $\vdash p \rightarrow q$, then $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$. ($\mathbf{OB}$-RM)

Proof
Suppose $\vdash p \rightarrow q$. Then by $\mathbf{OB}$-NEC, $\vdash \mathbf{OB}(p \rightarrow q)$, and then by K, $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$.

Corollary 1
$\vdash \mathbf{OB}p \rightarrow \mathbf{OB}(p \vee q)$. (Weakening)

Corollary 2
If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB}q$. ($\mathbf{OB}$-RE)[17]

Although the above axiomatization is standard, alternative axiomatizations do have certain advantages. One such axiomatization is given in Appendix A2 and shown to be equivalent to the one above.

---

[15]We ignore most of the simple definitional equivalences mentioned above, as well as DS and TTC.

[16]Compare $\mathbf{OB}$-N and $\mathbf{OB}$-D with $\mathbf{OB}(p \vee \sim p)$ and $\sim\mathbf{OB}(p \,\&\, \sim p)$, respectively.

[17]RE is the fundamental rule for "Classical Systems of modal logic", a class that includes normal modal logics as a proper subset. See [Chellas 1980].

von Wright's 1951 System and SDL: A quick comparison of SDL with the famous system in the seminal piece [von Wright, 1951] is in order. It is fair to say that von Wright [1951] launched deontic logic as an area of active research. There was a flurry of responses, and not a year has gone by since without published work in this area. von Wright's 1951 system is an important predecessor of SDL, but the variables there ranged over *act types* not propositions. As a result, the deontic operator symbols (e.g. **OB**) were interpreted as applying not to sentences, but to names of act types (cf. "to attend" or "attending") to yield a sentence (e.g. "it is obligatory to attend" or "attending is obligatory"). So iterated deontic sequences (e.g. **OBOB**A) were not well-formed formulas and shouldn't have been on his intended interpretation, since **OB**A (unlike A) is a sentence, not an act description, so not suitable for having **OB** as a preface to it (cf. "it is obligatory it is obligatory to run" or "running is obligatory is obligatory"). However, von Wright did think that there can be negations, disjunctions and conjunctions of act types, and so he used standard connectives to generate not only complex normative sentences (e.g. **OB**A & **PE**A), but complex act descriptions (e.g. A & ∼B), and thus complex normative sentences involving them (e.g. **OB**(A & ∼B) → **PE**(A & ∼B)). The standard connectives of PC are thus used in a systematically ambiguous way in von Wright's initial system with the hope of no confusion, but a more refined approach (as he recognized) would call for the usual truth-functional operators and a second set of act-type-compounding analogues to these.[18] Mixed formulas (e.g. A → **OB**A) were not well-formed in his 1951 system and shouldn't have been on his intended interpretation, since if **OB**A is well-formed, then A must be a name of an act type not a sentence, but then it can't suitably be a preface to →, when the latter is followed by an item of the sentence category (e.g. **OB**A). (Cf. "If to run then it is obligatory to run".) However, this also means that the standard violation condition for an obligation (e.g. **OB**p & ∼p) is not expressible in his system. von Wright also rejected NEC, but otherwise accepts analogues to the basic principles of SDL.

Researchers quickly opted for a syntactic approach where the variables and operators are interpreted propositionally as they are in PC (Prior 1962 [1955], Anderson [1956], Kanger 1971 [1957] and Hintikka [1957]), and von Wright soon adopted this course himself in his key early revisions of his "old system" (e.g. von Wright [1968; 1971] (originally published in 1964 and 1965). Note that this is essentially a return to the approach in Mally's deontic logic of a few decades before.

SDL can be strengthened in various ways, in particular, we might consider adding axioms where deontic operators are embedded within one another. For

---

[18]Compare the deontic logic in [Meyer, 1988], where a set of operators for action (drawn from dynamic logic) are used along with a separate set of propositional operators.

example, suppose we added the following formula as an axiom to SDL. Call the result "SDL+" for easy reference here:

A4.   **OB**(**OB**$p \to p$)

   This says (roughly) that it is required that obligations are fulfilled.[19] This is not a theorem of SDL (as we will see in the next section), so SDL+ is a genuine strengthening of SDL. Furthermore, it makes a logically *contingent* proposition (i.e. that **OB**$p \to p$) obligatory *as a matter of deontic logic.* SDL does not have this substantive feature. With this addition to SDL, it is easy to prove **OBOB**$p \to$ **OB**p, a formula involving an iterated occurrence of our main operator.[20] This formula asserts that if it is obligatory that p be obligatory, then p *is* obligatory. (Cf. "the only things that are required to be obligatory are those that actually are").[21]

2.2   *Standard Semantics*

The reader familiar with elementary textbook logic will have perhaps noticed that the deontic square and the modal square both have even better-known analogs for the quantifiers as interpreted in classical predicate logic ("all $x : p$" is read as all objects x satisfy condition p; similarly for "no $x : p$" and "some $x : p$"):

---
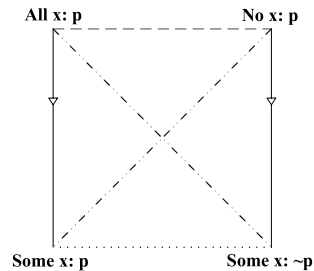
[19]Equivalently, **OB**($p \to$ **PE**p), it is required that only permissible things are true.

[20]For **OB**(**OB**$p \to p$) $\to$ (**OBOB**$p \to$ **OB**p) is just a special instance of **OB**-K. So using $A4$ above, and MP, we get **OBOB**$p \to$ **OB**p directly.

[21]See Chellas [1980, pp. 193–194] for a concise critical discussion of the comparative plausibility of these two formula. (Note that Chellas' chapters on deontic logic in this exceptional textbook are gems generally.) However, where Chellas states that if there are any unfulfilled obligations (i.e. **OB**p and $\sim p$ both hold), then "ours in one of the worst of all possible worlds", this is misleading, since the semantics does not rank worlds other than to sort them into acceptable and unacceptable ones (relative to a world). The illuminating underlying point is that for any world j whose alternatives are all p-worlds, but where p is false, it follows that not only can't j be an acceptable alternative to itself, but it can't be an acceptable alternative to any other world, i, either. Put simply, $A4$ implies that any (**OB**p & $\sim p$)-world is *universally* unacceptable. However, though indeed significant, this does not express a *degree or extent* of badness: given some ranking principle allowing for indefinitely better and worse worlds relative to some world i (such as in preference semantics for dyadic versions of SDL and kin — see below), j might be among the absolute best of the i-unacceptable worlds (i.e. ranked second only to those that are simply i-acceptable through and through), for all $A4$ implies.

Though less widely noted in textbooks, there is also a threefold classification for classical quantifiers:
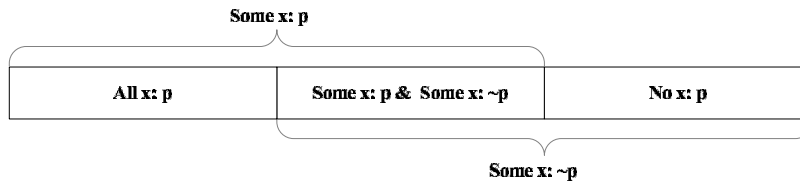
Here all conditions are divided into three jointly exhaustive and mutually exclusive classes: those that hold for all objects, those that hold for none, and those that hold for some and not for others, where no condition falls into more than one of these three categories. These deep quantificational analogies reflect much of the inspiration behind what is most often called "possible worlds semantics" for alethic modal and deontic logics, to which we now turn.[22] Once the analogies are noticed, this sort of semantics seems all but inevitable.

We now give a standard "Kripke-style" possible world semantics for SDL. Informally, we assume that we have a set of possible worlds, W, and a relation, A, relating worlds to worlds, with the intention that Aij iff j is a world where everything obligatory in i holds (i.e. no violations of the obligations holding in i occur in j). For brevity, we will call all such worlds so related to i, "i-Acceptable" worlds and denote them by A^i.[23] We then add that the acceptability relation is "serial": for every world, i, there is at least one i-acceptable world. Finally, propositions are either true or false at a world, never both, and when a proposition, p, is true at a world, we will often indicate this by referring to that world as a "p-world". The truth-functional operators have their usual behavior at each world. Our focus will be on the contribution deontic operators are taken to make.

The fundamental idea here is that the normative status of a proposition from the standpoint of a world i can be assessed by looking at how that proposition fairs at the i-acceptable worlds. Let's see how. For any given world, i, we can easily picture the i-accessible worlds as all corralled together in logical space as follows (where seriality is reflected by a small dot representing the presence of at least one world):

---

[22][von Wright, 1953] and Prior 1962 [1955] (already noted in the 1st ed., [1955]).
[23]The worlds related to i by A are also often called "ideal worlds". This language is not innocent [McNamara, 1996c].

A^i
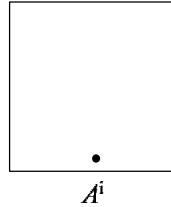
The intended truth-conditions, relative to i, for our five deontic operators can now be pictured as follows:

OBp: All p · A^i **PEp:** Some p · A^i **IMp:** No p · A^i **GRp:** Some ~p · A^i **OPp:** Some p & Some ~p · A^i
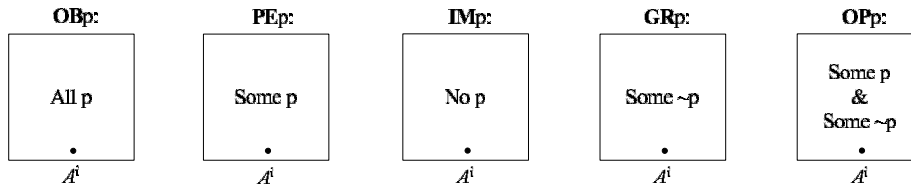
Thus, p is *obligatory* iff it holds in all the i-acceptable worlds, *permissible* iff it holds in some such world, *impermissible* iff it holds in no such world, *gratuitous* iff its negation holds in some such world, and *optional* iff p holds in some such world, and so does $\sim p$. When a formula must be true at any world in any such model of serially-related worlds, then the formula is *valid*.

Kripke-Style Semantics for SDL: A more formal characterization of this semantic framework follows

We define the frames (structures) for modeling SDL as follows.
F is a Kripke-SDL (or KD) Frame: $F = \langle W, A \rangle$ such that:

1. W is a non-empty set

2. A is a subset of $W \times W$

3. A is serial: $\forall i \exists j A i j$.

A model can be defined in the usual way, allowing us to then define truth at a world in a model for all sentences of SDL (and SDL+):

> *M is an Kripke-SDL Model*: $M = \langle F, V \rangle$, where F is an SDL Frame, $\langle W, A \rangle$, and V is an assignment on F : V is a function from the propositional variables to various subsets of W (the truth sets for the variables — the worlds where the variables are true for this assignment).

Let "$M \vDash_i p$" denote p's truth at a world, i, in a model, M.

Basic Truth-Conditions at a world, i, in a Model, M

[PC]: (Standard Clauses for the operators of Propositional Logic.)

[**OB**]: $M \vDash_i \mathbf{OB}p : \forall j$ [if Aij then $M \vDash_j p$].

Derivative Truth-Conditions

[**PE**]: $M \vDash_i \mathbf{PE}p : \exists j(Aij \,\&\, M \vDash_j p)$

[**IM**]: $M \vDash_i \mathbf{IM}p :\sim\exists j(Aij \,\&\, M \vDash_j p)$

[**GR**]: $M \vDash_i \mathbf{GR}p : \exists j(Aij \,\&\, M \vDash_j \sim p)$

[**OP**]: $M \vDash_i \mathbf{OP}p : \exists j(Aij \,\&\, M \vDash_j p) \,\&\, \exists j(Aij \,\&\, M \vDash_j \sim p)$.

p is true in the model, $M(M \vDash p) : p$ is true at every world in M.

p is valid $(\vDash p) : p$ is true in every model.

Metatheorem:
SDL is sound and complete for the class of all Kripke-SDL models.[24]

To illustrate the workings of this framework, consider NC (**OB**-D), $\mathbf{OB}p \rightarrow \sim$ **OB** $\sim p$. This is valid in this framework. For suppose that **OB**p holds at any world i in any model. Then each i-accessible world is one where p holds, and by the seriality of accessibility, there must be at least one such world. Call it j. Now we can see that $\sim\mathbf{OB} \sim p$ must hold at i as well, for otherwise, **OB**$\sim p$ would hold at i, in which case, $\sim p$ would have to hold at all the i-accessible worlds, including j. But then p as well as $\sim p$ would hold at j itself, which is impossible (by the semantics for "$\sim$"). The other axioms and rules of SDL can be similarly shown to be valid, as can all the principles listed above as derivable in SDL

However, $A4$, the axiom we added to SDL above to get SDL+, is not valid in the standard serial models. In order to validate $A4$, **OB**(**OB**$p \rightarrow p$), we need the further requirement of "secondary seriality": that any i-acceptable world, j, must be in turn acceptable to itself. We can illustrate such an i and j as follows:
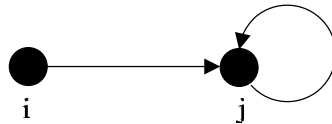
Here we imagine that the arrow connectors indicate relative acceptability, thus here, j (and only j) is acceptable to i, and j (and only j) is acceptable to j.

---

[24]That is, any theorem of SDL is valid per this semantics (soundness), and any formula valid per this semantics is a theorem of SDL (completeness).

If all worlds that are acceptable to any given world have this property of self-acceptability, then our axiom is valid. For suppose this property holds throughout our models, and that for some arbitrary world i, $\mathbf{OB}(\mathbf{OB}p \to p)$ is false at i. Then not all i-acceptable worlds are worlds where $\mathbf{OB}p \to p$ is true. So, there must be an i-acceptable world, say j, where $\mathbf{OB}p$ is true, but p is false. Since $\mathbf{OB}p$ is true at j, then p must be true at all j-acceptable worlds. But by stipulation, j is acceptable to itself, so p must be true at j, but this contradicts our assumption that p was false at j. Thus $\mathbf{OB}(\mathbf{OB}p \to p)$ must be true at all worlds, after all.

Two Counter-Models Regarding Additions to SDL

Here we show that $A4$, $\mathbf{OB}(\mathbf{OB}p \to p)$, is not derivable in SDL and that SDL $+ \mathbf{OBOB}p \to \mathbf{OB}p$ does not imply $A4$.
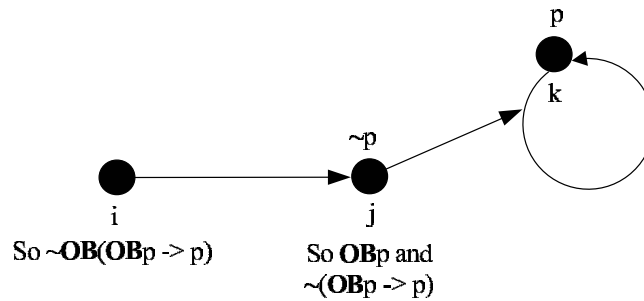
We first provide a counter-model to show that $A4$ is indeed a genuine (non-derivable) addition to SDL:

> Here, seriality holds, since each of the three worlds has at least one world acceptable to it (in fact, exactly one), but *secondary* seriality fails, since although j is acceptable to i, j is not acceptable to itself. Now look at the top annotations regarding the assignment of truth or falsity to p at j and k. The lower deontic formuli derive from this assignment and the accessibility relations. (The value of p at i won't matter.) Since p holds at k, which exhausts the worlds acceptable to j, $\mathbf{OB}p$ must hold at j, but then, since p itself is false at j, $(\mathbf{OB}p \to p)$ must be false at j. But j is acceptable to i, so not all i-acceptable worlds are ones where $(\mathbf{OB}p \to p)$ holds, so $\mathbf{OB}(\mathbf{OB}p \to p)$ must be false at i.[25] We have already proven that seriality, which holds in this model, automatically validates $\mathbf{OB}$-D. It is easy to show that the remaining ingredients of SDL hold here as well.[26]

---

[25] Note that this is in contrast to j itself, where the latter formula does hold, for the reader can easily verify that $(\mathbf{OB}p \to p)$ holds at k in this model, and k is the only world acceptable to j.

[26] The remaining items hold independently of seriality. Completing the proof amounts to both a proof of SDL's soundness with respect to our semantics, and of $A4$'s independence (non-

We proved above that ($\mathbf{OBOB}p \rightarrow \mathbf{OB}p$) is derivable from $A4$. Here is a model that shows that the converse fails. It is left to the reader to verify that given the accessibility relations and indicated assignments to p at j and k, $\mathbf{OBOB}p \rightarrow \mathbf{OB}p$ must be (vacuously) true at i, while $\mathbf{OB}(\mathbf{OB}p \rightarrow p)$ must be false at i.
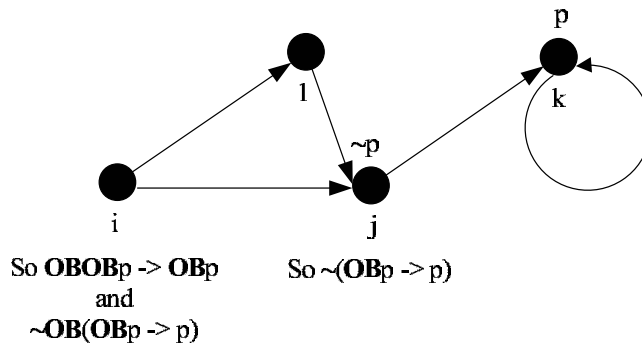
We should also note that one alternative semantic picture for SDL is where we have a set of world-relative ordering relations, one for each world i in W, where $j \geq_i k$ iff j is as good as k (and perhaps better) relative to i, where not all worlds in W need be in the purview (technically, the field) of the ordering relation associated with i. We then assume that from the standpoint of any world i, (a) each world in its purview is as good as itself, (b) if one is as good as a second, and the second is as good as a third, then the first is as good as the third, (c) and for any two worlds in its purview, either the first is as good as the second or vice versa (i.e. respectively, each such $\geq i$ is reflexive, transitive, and connected in the field of $\geq_i$). $\mathbf{OB}p$ is then true at a world i iff there is some world k that is first of all as good as itself relative to i, and all worlds ranked as good as k from the standpoint of i are p-worlds. Thus, roughly, $\mathbf{OB}p$ is true at i iff p is true from somewhere on up in the subset of worlds in W ordered relative to i. It is widely recognized that this approach will also determine SDL, but proofs of this are not widely available.[27]

However, if we add "The Limit Assumption", that for each world i, there is always at least one world as good (relative to i) as all worlds in i's purview (i.e. one i-best world), we can easily generate our earlier semantics for SDL derivatively. We need only derive the natural analogue to our prior truth-conditions for $\mathbf{OB}$: $\mathbf{OB}p$ is true at a world i iff p is true at all the i-best worlds.

---

derivability from) SDL.

[27]But see [Goble, Forthcoming-b].

OBp:

all p-worlds here ⟶

the i-ranked worlds
(the higher the level,
the better the worlds
within it, relative to i)
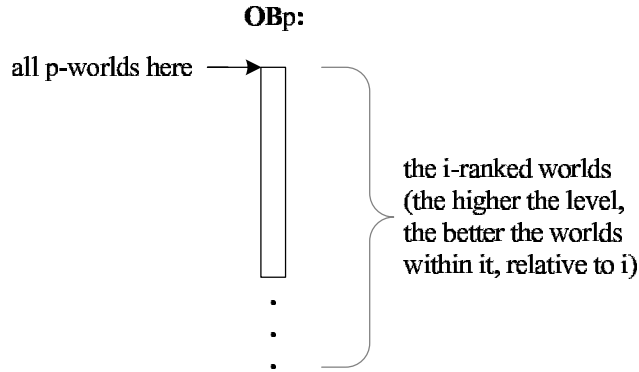
Essentially, the ordering relation coupled with the Limit Assumption just gives us a way to *generate* the set of i-acceptable worlds instead of taking them as primitive in the semantics: j is i-acceptable iff j is i-best. Once generated, we look only at what is going on in the i-acceptable worlds to interpret the truth-conditions for the various deontic operators, just as with our simpler Kripke-Style semantics. The analogue to the seriality of our earlier i-acceptability relation is also assured by the Limit Assumption, since it entails that for each world i, there is always some i-acceptable (now i-best) world. Although this ordering semantics approach appears to be a bit of overkill here, it became quite important later on in the endeavor to develop expressively richer deontic logics (ones going beyond the linguistic resources of SDL). We will return to this later.

For now, we turn to the second-most well known approach to monadic deontic logic, one in which SDL will emerge *derivatively.*

3 THE ANDERSONIAN–KANGERIAN REDUCTION

The *Andersonian–Kangerian* reduction is dually-named in acknowledgement of Kanger's and Anderson's independent formulation of it around the same time.[28] As [Hilpinen, 2001a] points out, the approach is adumbrated much earlier in Leibniz. We follow Kanger's development here, noting Anderson's toward the end.

3.1 Standard Syntax

Assume that we have a language of classical modal propositional logic, with a distinguished (deontic) propositional constant:

"d" for "all (relevant) normative demands are met".

Now consider the following axiom system, "Kd":

---

[28]Kanger [1971 [1957]] (circulating in 1950 as a typescript) and Anderson [1967 [1956]] and [Anderson, 1958].

Kd: A1   All Tautologies   (TAUT)
        A2: $\Box(p \to q) \to (\Box p \to \Box q)$ (K)
        A3: $\Diamond d$ $(\Diamond d)$
        R1:   If $\vdash p$ and $\vdash p \to q$ then $\vdash q$ (MP)
        R2:   If $\vdash p$ then $\vdash \Box q$ (NEC).

Kd is just the normal modal logic K with $A3$ added.[29] $A3$ is interpreted as telling us that it is possible that all normative demands are met. In import when added to system K, it is similar to (though stronger than) the "No Conflicts" axiom, $A3$, of SDL. *All* of the Traditional Scheme's deontic operators are defined operators in Kd:

$$\mathbf{OB}p =_{\mathrm{df}} \Box(d \to p)$$
$$\mathbf{PE}p =_{\mathrm{df}} \Diamond(d\&p)$$
$$\mathbf{IM}p =_{\mathrm{df}} \Box(p \to\sim d)$$
$$\mathbf{GR}p =_{\mathrm{df}} \Diamond(d\& \sim p)$$
$$\mathbf{OP}p =_{\mathrm{df}} \Diamond(d\&p)\&\Diamond(d\& \sim p).$$

So in Kd, p is *obligatory* iff p is necessitated by all normative demands being met, *permissible* iff p is compatible with all normative demands being met, *impermissible* iff p is incompatible with all normative demands, *gratuitous* iff p's negation is compatible with all normative demands, and *optional* iff p is compatible with all normative demands, and so is $\sim p$. Since none of the operators of the Traditional Scheme are taken as primitive, and the basic logic is a modal logic with necessity and possibility as the basic modal operators, this is referred to as "a reduction" (of deontic logic to modal logic).

Proofs of SDL-ish wffs are then just K-proofs of the corresponding modal formulae involving "d".

---

[29] K is the basic (weakest) normal modal logic. (See the entry in this volume on modal logics by Rob Goldblatt.) Traditionally, and in keeping with the intended interpretation, the underlying modal logic had T as a theorem, indicating that necessity was truth-implicating. We begin with K instead because T generates a system *stronger* than SDL. We will look at the addition of T shortly. Åqvist 2002 [1984] is an excellent source on the meta-theory of the relationship between SDL-ish deontic logics and corresponding Andersonian–Kangerian modal logics, as well as the main dyadic (primitive conditional operator) versions of these logics. Smiley [1963] is a landmark in the comparative study of such deontic systems. McNamara [1999] gives determination results for various deontic logics that employ three deontic constants allowing for a "reduction" of other common sense normative concepts.

Two Simple Proofs in Kd:
First consider the very simple proof of **OB**d:

> By PC, we have $d \rightarrow d$ as a theorem. Then by $R2$, it follows that $\square(d \rightarrow d)$, that is, **OB**d.

Next consider a proof of NC, **OB**$p \rightarrow \sim$**OB** $\sim\sim p$. As usual, in proofs of wffs with deontic operators, we make free use of the rules and theorems that carry over from the normal modal logic K. Here it is more perspicuous to lay the proof out in a numbered-lined stack:

1. Assume $\sim($**OB**$p \rightarrow \sim$**OB** $\sim p)$. (For reductio)
2. That is, assume
   $\sim(\square(d \rightarrow p) \rightarrow \sim\square(d \rightarrow \sim p))$. (Def of "**OB**")
3. So $\square(d \rightarrow p)\,\&\,\square(d \rightarrow \sim p)$.   (2, by PC)
4. So $\square(d \rightarrow (p\,\&\,\sim p))$.   (3, derived rule of modal logic, K)
5. But $\lozenge d$ (A3)
6. So $\lozenge(p\,\&\,\sim p)$.   (4 and 5, derived rule of modal logic, K)
7. But $\sim\lozenge(p\,\&\,\sim p)$.   (a theorem of modal logic, K)
8. So **OB**$p \rightarrow \sim$**OB** $\sim p$.   (1–7, PC)

Part of the point of the Andersonian–Kangerian reduction is to find a way to *generate* SDL from non-SDL resources, which can be easily done in Kd (as the next box shows).

SDL Containment Proof
We give a proof that SDL is indeed contained in Kd.
Recall SDL:

$A1$:	All tautologous wffs of the language (TAUT)	
$A2$:	**OB**$(p \rightarrow q) \rightarrow ($**OB**$p \rightarrow$ **OB**$q)$	(**OB**-K)
$A3$:	**OB**$p \rightarrow \sim$**OB** $\sim p$	(**OB**-NC)
$R1$:	If $\vdash p$ and $\vdash p \rightarrow q$ then $\vdash q$	(MP)
$R2$:	If $\vdash p$ then $\vdash$ **OB**p	(**OB**-NEC)

We have already shown **OB**-NC is derivable in Kd above, and TAUT and MP are given, since they hold for *all* formulas of Kd. So we need only derive **OB**-K and **OB**-NEC of SDL, which we will do in reverse order. Note that RM, if $\vdash r \rightarrow s$, then $\vdash \square r \rightarrow \square s$, is derivable in Kd, and so we rely on it in the second proof.[30]

---

[30] An examination of our earlier proof that RM for **OB** was one of the derived rules within SDL reveals that *for any* system with NEC and K governing a necessity operator, the rule RM is derivable. Here it is again adapted for 2:

  Show: If $\vdash p \rightarrow q$ then $\vdash \square p \rightarrow q$. (RM)
  Proof: Suppose $\vdash p \rightarrow q$. Then by NEC, $\vdash \square(p \rightarrow q)$, and then by $K \vdash \square p \rightarrow \square q$.

Show: If $\vdash p$ then $\vdash \mathbf{OB}p$. (**OB**-NEC)
Proof: Assume $\vdash p$. It follows by PC that $\vdash d \rightarrow p$. So by NEC for $\square$, we get $\vdash \square(d \rightarrow p)$, that is, $\mathbf{OB}p$.

Show: $\vdash \mathbf{OB}(p \rightarrow q) \rightarrow (\mathbf{OB}p \rightarrow \mathbf{OB}q)$. (K of SDL)
Proof: Assume $\mathbf{OB}(p \rightarrow q)$ and $\mathbf{OB}p$. From PC alone, $\vdash (d \rightarrow (p \rightarrow q)) \rightarrow [(d \rightarrow p) \rightarrow (d \rightarrow q)]$. So by RM for $\square$, we have $\vdash \square(d \rightarrow (p \rightarrow q)) \rightarrow \square[(d \rightarrow p) \rightarrow (d \rightarrow q)]$. But the antecedent of this is just $\mathbf{OB}(\rightarrow q)$ in disguise, which is our first assumption. So we have $\square[(d \rightarrow p) \rightarrow (d \rightarrow q)]$ by MP. Applying K for $\square$ to this, we get $\square(d \rightarrow p) \rightarrow \square(d \rightarrow q)$. But the antecedent to this is just our second assumption, $\mathbf{OB}p$. So by MP, we get $\square(d \rightarrow q)$, that is, $\mathbf{OB}q$.

Metatheorem: SDL is derivable in Kd.

Note that showing that the pure deontic fragment of Kd contains *no more than* SDL is a more complex matter. The proof relies on already having semantic metatheorems available. An excellent source for this is Åqvist 2002 [1984].[31]

In addition to containing all theorems of SDL, we note a few theorems specific to Kd because of the non-overlapping syntactic ingredients, $d, \square,$ and $\lozenge$:

$$\vdash \mathbf{OB}d$$
$$\vdash \square(p \rightarrow q) \rightarrow (\mathbf{OB}p \rightarrow \mathbf{OB}q) \quad \text{(RM}')$$
$$\vdash \square p \rightarrow \mathbf{OB}p \quad \text{(Nec}')$$
$$\vdash \mathbf{OB}p \rightarrow \lozenge p \quad \text{(Kant's Law)}$$
$$\vdash \sim\lozenge(\mathbf{OB}p \,\&\, \mathbf{OB} \sim p) \quad \text{(NC}').$$

These are easily proven.[32]

Although our underlying modal system is just K, adding further *non-deontic* axiom schemata (i.e. those neither abbreviate-able via SDL wffs, nor involving d specifically) can nonetheless have a deontic impact. To illustrate, suppose we added a fourth axiom, one to the effect that necessity is here truth-implicating, called axiom "T":

$$T : \square p \rightarrow p.$$

Call the system that results from adding this formula to our current system "KTd". Axiom T is certainly plausible enough here, since, as mentioned above, this approach to deontic logic is more sensible if necessity is interpreted as truth-implicating, since it takes obligations to be things necessitated by all normative demands being met, but in what sense, if not a truth-implicating sense of necessity?

---

[31] The "pure deontic fragment" is the set of theorems of Kd that can be abbreviated using only the truth-functions and the five standard deontic operators.

[32] We proved the first above, and given our definition of **OB**, RM′ and NEC′ follow from standard features of the modal logic K alone, but Kant's Law and NC′ also depend on the distinctive deontic axiom, $\lozenge d$.

Now with T added to Kd, we have gone beyond SDL, since we can now prove things expressible in SDL's language that we have already shown are not theorems of SDL. The addition of T makes derivable our previously mentioned axiom $A4$ of SDL+, which we have shown is not derivable in SDL itself:

$$\vdash \mathbf{OB}(\mathbf{OB}p \to p).^{33}$$

So, reflecting on the fact that SDL+ is derivable in KTd, we see that the Andersonian–Kangerian reduction must either rely on a non-truth-implicating conception of necessity in order for its pure deontic fragment to match SDL, or SDL itself is not susceptible to the Andersonian–Kangerian reduction. Put another way, the most plausible version of the Andersonian–Kangerian reduction can't help but view "Standard Deontic Logic" as too weak.

Determinism and Deontic Collapse in the Classic A-K-Framework:
Note that adding $T, \Box p \to p$, allows us to explore a classical issue connected with determinism and deontic notions. Given axiom T is now naturally taken to encode a truth-implicating notion of necessity in systems containing it. For this reason, we can now easily augment KTd with an axiom expressing determinism:

$$\vdash p \to \Box p. \quad \text{(Determinism)}$$

It is obvious on a moments reflection that, along with T, Determinism (as an axiom schemata), yields a collapse of modal distinctions, since $p \leftrightarrow \Box p$, and $p \leftrightarrow \Diamond p$ would then be provable. However, we can also explore, the classical question of what happens to moral distinctions if determinism holds. This question is also settled from the perspective of KTd, since the following is a derivable rule of that system:

$$\text{If } \vdash p \to \Box p, \text{ then } \vdash p \leftrightarrow \mathbf{OB}p.$$

To prove this, assume Determinism, $\vdash p \to \Box p$.

a) We first show $\vdash p \to \mathbf{OB}p$. Assume p. Then by Determinism, $\Box p$. So by NEC$'$, namely $\Box p \to \mathbf{OB}p$, we get $\mathbf{OB}p$, and thus $\vdash p \to \mathbf{OB}p$.

---

[33] *Proof*: By $T, \vdash \Box(d \to p) \to (d \to p)$. Then by PC we can get $\vdash d \to [\Box(d \to p) \to p]$. From this in turn, by NEC, we have $\vdash \Box(d \to [\Box(d \to p) \to p])$, that is $\mathbf{OB}(\mathbf{OB}p \to p)$.

b) Next, we show $\vdash \mathbf{OB}p \to p$. Assume $\mathbf{OB}p$ and $\sim p$ for *reductio*. By Determinism, we have $\sim p \to \Box \sim p$. So $\Box \sim p$. This yields $\mathbf{OB} \sim p$, by NEC$'$. But then we have $\mathbf{OB}p \,\&\, \mathbf{OB} \sim p$, which contradicts a prior demonstrated theorem, NC. So $\vdash \mathbf{OB}p \to p$.

So, from the standpoint of the classic Andersonian–Kangerian reduction, where the notion of necessity is truth-implicating (and thus axiom T is intended), the addition of the most natural expression of determinism entails that truth and deontic distinctions collapse. This in turn is easily seen to imply these corollaries:

$$\text{If } \vdash p \to \Box p, \text{ then } \vdash p \leftrightarrow \mathbf{PE}p$$
$$\text{If } \vdash p \to \Box p, \text{ then } \vdash \mathbf{OB}p \leftrightarrow \mathbf{PE}p$$
$$\text{If } \vdash p \to \Box p, \text{ then } \vdash \sim p \leftrightarrow \mathbf{IM}p$$
$$\text{If } \vdash p \to \Box p, \text{ then } \vdash \sim\mathbf{OP}p.$$

For example, consider the last corollary. By definition, p is optional iff neither p nor $\sim p$ is obligatory. But given determinism, this would entail that neither p nor $\sim p$ is true, which is not possible. So nothing can be morally optional if determinism is true.

Anderson's approach is practically equivalent to Kanger's. First, consider the fact that we can easily define another constant in Kd, as follows:

$$s =_{\mathrm{df}} \sim d,$$

where this new constant would now be derivatively read as follows:

"some (relevant) normative <u>demands</u> has been violated".

Clearly our current axiom, $\Diamond d$, could be replaced with $\sim\Box s$, asserting that it is not necessary that some normative demand is violated. We could then define $\mathbf{OB}$ as:

$$\mathbf{OB}p =_{\mathrm{df}} \Box(\sim p \to s),$$

and similarly for the other four operators.

Essentially, Anderson took this equivalent course with "s" being his primitive, and initially interpreted as standing for something like "the sanction has been invoked" or "there is a liability to sanction", and then $\sim\Box s$ was the axiom added to some modal system (at least as strong as modal system KT) to generate SDL and kin.

We should also note that Anderson was famous as a founding figure in relevant logic. (See the entry in this volume on relevant logics by Greg Restall.) Instead of using strict implication, $\Box(p \to q)$, he explored the use of a relevant (and thus neither a material nor strict) conditional, $\Rightarrow$, to express the reduction as:
$\mathbf{OB}p =_{\mathrm{df}} \quad \sim p \Rightarrow s$. (A bit more on this can be found in [Lokhorst, 2004]. See

references there, but also see [Mares, 1992].) This alternative reflects the fact that
there is an issue in both Kanger's and Anderson's strict necessitation approaches
of just what notion of "necessity" we can say is involved in claiming that meetings
all normative demands (or avoiding the sanction) *necessitates p*?

As a substantive matter, how should we think of these "reductions"? For exam-
ple, should we view them as giving us an *analysis* of what it is for something to be
obligatory? Well, taking Kanger's course first, it would seem that d must be read
as a distinctive deontic ingredient, if we are to get the derivative deontic reading
for the "reduced" deontic operators. Also, as our reading suggests, it is not clear
that d does not, at least by intention, express a complex quantificational notion
involving the very concept of obligation (demand) as a proper part, namely that all
obligations have been fulfilled, so that the "reduction", *presented as an analysis*,
would appear to be circular. If we read d instead as "ideal circumstances obtain",
the claim of a substantive reduction or analysis appears more promising, until we
ask, "Are the circumstances ideal only with respect to meeting normative *demands
or obligations*, or are they ideal in other (for example supererogatory) ways that go
beyond merely satisfying normative *demands*?" Anderson's "liability to sanction"
approach may appear more promising, since the idea that something is obligatory
if (and only if) and because non-compliance necessitates (in some sense) liability to
(or perhaps desert of) punishment does not appear to be circular (unless the notion
of "liability" itself ultimately involves the idea of permissibility of punishment),
but it is still controversial (e.g. imperfect obligations are often thought to include
obligations where no one has a right to sanction you for violations). Alternatively,
perhaps a norm that is merely an *ideal* cannot be *violated*, in which case perhaps
norms that have been *violated* can be distinguished (as a subset) from norms that
have not been *complied* with, and then the notion of an obligation as something
that must obtain unless some norm is violated will not be obviously circular. The
point here is that there is a substantive philosophical question lingering here that
the language of a "reduction" brings naturally to the surface. The formal utility
of the reduction does not hinge of this, but its philosophical significance does.

3.2 *Standard Semantics*

The semantic elements here are in large part analogous to those for SDL. We have
a binary relation again, but this time instead of a relation interpreted as relating
worlds *acceptable* to a given world, here we will have a relation, R, relating worlds
"accessible" to a given world (e.g. possible relative to the given world). The only
novelties are two: (1) we add a simple semantic element to match our syntactic
constant "d", and (2) we add a slightly more complex analog to seriality, one that
links the accessibility relation to the semantic element added in order to model d.
We introduce the elements in stages.

Once again, assume that we have a set of possible worlds, W, and assume that
we have a relation, R, relating worlds to worlds, with the intention that Rij iff j
is accessible to i (e.g. j is a world where everything true in j is possible relative

to i).[34] For brevity, we will call all worlds possible relative to i, "*i-accessible*
worlds" and denote them by R^i. For the moment, no restrictions are placed on
the relation R. We can illustrate these truth-conditions for our modal operators
with a set of diagrams analogous to those used for giving the truth-conditions for
SDL's deontic operators. We use obvious abbreviations for <u>nec</u>essity, <u>pos</u>sibility,
<u>imp</u>ossibility, <u>non</u>-necessity, and <u>con</u>tingency:

NECp:	POSp:	IMPp:	NONp:	CONp:
All p	Some p	No p	Some ~p	Some p & Some ~p
R^i	R^i	R^i	R^i	R^i
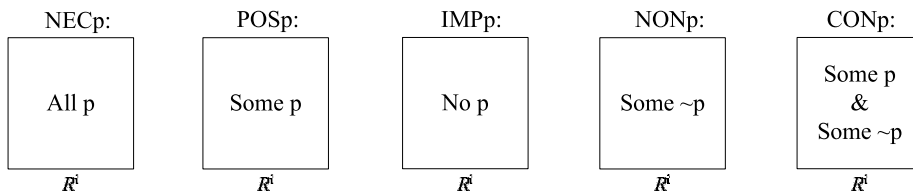
Here we imagine that for any given world, i, we have corralled all the i-accessible
worlds together. We then simply look at the quantificational status of p (and/or
$\sim p$) in these i-accessible worlds to determine p's modal status back at i: at a given
world i, p is *necessary* if p holds throughout R^i, *possible* if p holds somewhere in
R^i, *impossible* if p holds nowhere in R^i, *non-necessary* if $\sim p$ holds somewhere in
R^i, and *contingent* if p holds somewhere in R^i, and so does $\sim p$.

The only *deontic* element in the syntax of Kd is our distinguished constant,
d, intended to express the fact that all normative demands are met. To model
that feature, we simply assume that the worlds are divided into those where all
normative demands are met and those that are not. We denote the former subset
of worlds by "DEM" in a model. Then d is true at a world j iff j belongs to DEM.
Here is a picture where d is true at an arbitrary world j:

d:

DEM

•j
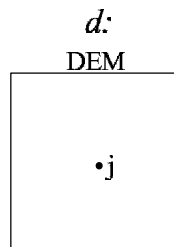
[34]Note that this means that, for generality, we assume that what is possible may vary from
one world to another. This is standard in this sort of semantics for modal logics. For example, if
we wanted to model physical possibility and necessity, what is physically possible for our world,
may not be so for some other logically possible world with different fundamental physical laws
than ours. By adding certain constraints, we can generate a picture where what is possible does
not vary at all from one world to another. (See the entry in this volume on modal logics by Rob
Goldblatt.)

Since j is contained in DEM, that means all normative demands are met at j.[35]

Corresponding to simple seriality for SDL (that there is always an i-acceptable world), we assume what I will call "strong seriality" for Kd: for every world i, there is an i-accessible world that is among those where all normative demands are met. In other words, for every world i, the intersection of the i-accessible worlds with those where all normative demands are met is non-empty. Given the truth conditions for d, strong seriality validates $\Diamond d$, ensuring that for any world i, there is always some *i-accessible* world where d is true:

<div align="center">

POSd:

</div>
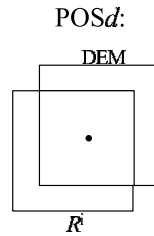
Given these semantic elements, if you apply them to the definitions of the five deontic operators of Kd, you will see that in each case, the normative status of p at i depends on p's relationship to *this intersection* of the i-accessible worlds and the worlds where all normative demands are met:
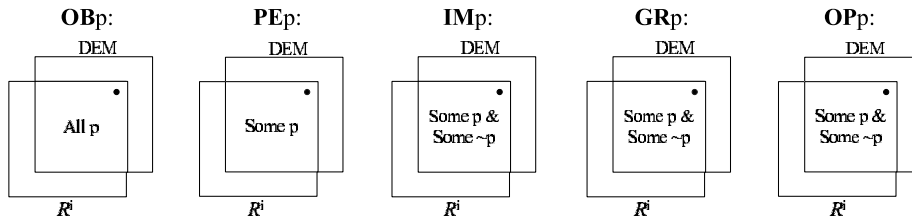
If that inter-section is permeated by p-worlds, p is *obligatory*, if it contains some p-world, p is *permissible*, if it contains no p-world, p is *impermissible*, if it contains some $\sim p$-world, p is *gratuitous*, and if it contains some p-world as well as some $\sim p$-world, then p is *optional*.[36]

---

[35]Note that we could add an ordering-relation semantics like that described at the end of our section on SDL in order to generate the DEM component of these models. The main difference would be that instead of a set of world-relative ordering relations, one for each world (e.g. $\geq_i$) there would be just one ordering relation, $\geq$, whereby all worlds in W (in a given model) would be ranked just once. This relation would be reflexive, transitive, and connected, while satisfying the Limit Assumption in W. DEM would then be the set of all the best worlds in W, and then the truth conditions for d and the five deontic operators would be cast via DEM so generated.

[36]More explicitly, since $\mathbf{OB}p =_{\mathrm{df}} \Box(d \to p)$, we need only look at $\Box(d \to p)$. The latter will be true at a world i, iff $(d \to p)$ is true at all the i-accessible worlds. But given the truth-conditions for the material conditional "$\to$", that just amounts to saying p is true at all those i-accessible worlds (if any) where d is true, which in turn holds iff p is true at all the i-accessible worlds falling within DEM, i.e. at there intersection (which is non-empty by strong seriality). Similarly for the other four deontic operators.

Kripke-Style Semantics for Kd:

A more formal characterization of this semantic picture is given here. We define the frames for modeling Kd as follows:

F is a Kd frame: $F = \langle W, R, DEM \rangle$ such that

1. W is a non-empty set

2. R is a subset of $W \times W$

3. DEM is a subset of W

4. $\forall i \exists j (Rij \,\&\, j \in DEM)$.

A model can be defined in the usual way, allowing us to then define truth at a world in a model for all sentences of Kd (as well as for KTd).

> *M is a Kd Model*
>
> $M = \langle F, V \rangle$, where F is a Kd Frame, $\langle W, R, DEM \rangle$, and V is an assignment on F: V is a function from the propositional variables to various subsets of W.
>
> *Basic Truth-Conditions at a world, i, in a Model, M*:
>
> [PC]:   (Standard Clauses for the operators of Propositional Logic.)
>
> [$\Box$]: $M \vDash_i \Box p$ iff $\forall j$ (if Rij then $M \vDash_j p$).
>
> [d]: $M \vDash_i d$ iff i is in DEM.
>
> *Derivative Truth-Conditions*:
>
> [$\Diamond$]: $M \vDash_i \Diamond p$: $\exists j (Rij \,\&\, M \vDash_j p)$
>
> [**OB**]: $M \vDash_i \mathbf{OB}p : \forall j[\text{if } Rij \,\&\, j \in DEM \text{ then } M \vDash_j p]$.
>
> [**PE**]: $M \vDash_i \mathbf{PE}p$: $\exists j (Rij \,\&\, j \in DEM \,\&\, M \vDash_j p)$
>
> [**IM**]: $M \vDash_i \mathbf{IM}p : \forall j[if\, Rij \,\&\, j \in DEM \text{ then } M \vDash_j {\sim}p]$
>
> [**GR**]: $M \vDash_i \mathbf{GR}p$: $\exists j (Rij \,\&\, j \in DEM \,\&\, M \vDash_j {\sim}p)$
>
> [**OP**]: $M \vDash_i \mathbf{OP}p : \exists j (Rij \,\&\, j \in DEM \,\&\, M \vDash_j p) \,\&\, \exists j (Rij \,\&\, j \in DEM \,\&\, M \vDash_j {\sim}p)$.
>
> (Truth in a model and validity are defined just as for SDL.)

Metatheorem: Kd is sound and complete for the class of all Kd models.

If we wish to validate $T, \Box p \to p$ (and derivatively, $A4$, $\mathbf{OB}(\mathbf{OB}p \to p)$), we need only stipulate that the accessibility relation, R, is reflexive: that each world i is i-accessible (possible relative to itself):

i
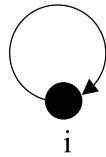
For then $\Box p \to p$ must be true at any world i, for if $\Box p$ is true at i, then p is true at each i-accessible world, which includes i, which is self-accessible. This will indirectly yield the result that $\mathbf{OB}(\mathbf{OB}p \to p)$ is true in all such models as well.

We turn now to a large variety of problems attributed to the preceding closely related systems.

4   CHALLENGES TO STANDARD DEONTIC LOGICS

Here we consider some of the "paradoxes" attributed to "Standard Deontic Logics" like those above (SDLs). Although the use of "paradox" is widespread within deontic logic and it does conform to a technical use in philosophical logic, namely the distinction between "paradox" and "antinomy" stemming from Quine's seminal "The Ways of Paradox" [Quine 1976 [1962]], I will also use "puzzle", "problem" and "dilemma" below.

To paraphrase von Wright, the number of outstanding problems in deontic logic is large, and most of these can be framed as problems or limitations attributed to SDLs. In this section we will list and briefly describe most of them, trying to group them where feasible under crucial principles of SDL or more general themes.

4.1   A Puzzle Centering Around the Very Idea of a Deontic Logic

Jorgensen's Dilemma [Jorgensen, 1937]

A view still held by many researchers within deontic logic and metaethics, and particularly popular in the first few decades following the emergence of positivism, was that evaluative sentences are not the sort of sentences that can be either true or false. But then how can there be a logic of normative sentences, since logic is the study of what follows from what, and one thing can follow from another only if the things in question are the sort that can be either true or false? So there can be no deontic logic. On the other hand, some normative sentences do seem to follow from others, so deontic logic must be possible. What to do? That's Jorgesson's dilemma.

A widespread distinction is that between a *norm* and a *normative proposition*.[37] The idea is that a normative sentence such as "You may park here for one hour" may be used by an authority to provide permission on the spot or it may be used by a passerby to report on an already existing norm (e.g. a standing municipal regulation). The activity of using a normative sentence as in the first example is sometimes referred to as "norming" — it creates a norm by granting permission by the very use. The second use is often said to be descriptive, since the sentence is then not used to grant permission, but to report that permission to do so is a standing state. It is often maintained that the two uses are mutually exclusive, and only the latter use allows for truth or falsity. Some have challenged the exclusiveness of the division, by blending semantics and speech-act theory (especially regarding performatives), thereby suggesting that it may be that one who is in authority to grant a permission not only grants it in performing a speech act by uttering the relevant sentence (as in the first example), but also thereby makes what it said true (that the person is permitted to park).[38],[39]

4.2   A Problem Centering Around NEC

The Logical Necessity of Obligations Problem

Consider

1.   Nothing is obligatory.

A natural representation of this in the language of SDL would be:

1′. $\sim$**OB**q, for all q.

We noted above that **OB**-*NEC* entails **OB**-*N* (i.e. $\vdash$ **OB**$\top$); but given 1′), we get $\sim$**OB**$\top$, and thus a contradiction. SDL seems to imply that it is a truth of logic that something is always obligatory. But it seems that what 1) expresses, an absence of obligations, is possible. For example, consider a time when no rational agents existed in the universe. Why should we think that any obligations existed then?

---

[37]von Wright [1963]; Hedenius [1963 [1941]]; Alchourron and Bulygin [1971; 1981]; Makinson [1999]. von Wright [1963] attributes the distinction to Ingemar Hedenius [1963 [1941]]. See [Makinson and van der Torre, 2003] for a recent attempt to provide a logic of norms.

[38]See [Lemmon 1962a; Kamp 1974; 1979]. It is often thought that performative utterances generally work this way [Kempson, 1977]. For example, if a marriage ceremony conducted by a legitimate authority requires that authority to end the ceremony with the proverbial (but dated) "You are now man and wife" in order to complete an act of marriage, the speech act utilizing this sentence not only marries the couple (in the context), but it appears to also be a true description of their state as of that moment.

[39]Perhaps this is as good a place as any to direct the interested reader to a problem much discussed in metaethics since Hume: the so-called "Is-Ought Problem": [Schurz, 1997] is an excellent full length study employing the techniques of deontic and modal logic in investigating this problem.

von Wright [1951] notes that since the denial of $\sim$**OB**$\top$ is provably equivalent to **PE**$\bot$ (given the traditional definitional scheme and **OB**-*RE*), and since both **OB**$\top$ and **PE**$\bot$ are odd, we should opt for a "principle of contingency", which says that **OB**$\top$ and $\sim$**PE**$\bot$ are both logically contingent. von Wright [1963, pp. 152–154] argues that **OB**$\top$ (and **PE**$\bot$) do not express real prescriptions. Føllesdal and Hilpinen [1971, p. 13] suggests that excluding **OB**-*N* only excludes "*empty normative systems*" (i.e. normative systems with no obligations), and perhaps not even that, since no one can fail to fulfill **OB**$\top$ anyway, so why worry?[40] However, since it is dubious that anyone can bring it about that $\top$, it would seem to be equally dubious that anyone can "fulfill" **OB**$\top$, and thus matters are not so simple. al-Hibri [1978] discusses various early takes on this problem, rejects **OB**-*N*, and later develops a deontic logic without it. Jones and Pörn [1985] explicitly rejects **OB**-*N* for "ought" in the system developed there, where the concern is with what people ought to do. If we are reading **OB** as simply "it ought to be the case that", it is not clear that there is anything counterintuitive about **OB**$\top$ (now read as, essentially, "it ought to be that contradictions are false"), but there is also no longer any obvious connection to what is obligatory or permissible for that reading, or to what people ought to do.

4.3   Puzzles Centering Around RM

Free Choice Permission Paradox [Ross, 1941]

Consider:

1.   You may either sleep on the <u>s</u>ofa-bed or sleep on the <u>g</u>uest room bed.[41]

2.   You may sleep on the sofa-bed and you may sleep on the guest room bed.

The most straightforward symbolization of these in SDL appears to be:

1′.   **PE**$(s \vee g)$

2′.   **PE**$s \,\&\, $**PE**$g$

Now it is also natural to see 2) as following from 1): if you permit me to sleep in either bed, it would seem that I am permitted to sleep in the first, *and* I am permitted to sleep in the second (though perhaps not to sleep in both, straddling the two, as it were). But 2′) does not follow from 1′) and the following is not a theorem of SDL:

∗.   **PE**$(p \vee q) \rightarrow ($**PE**$p \,\&\, $**PE**$q)$.

---

[40]Cf. [Prior, 1958].

[41]I will underline key letters to serve as cues for symbolization schemes left implicit, but hopefully clear enough.

Furthermore, suppose $*$ were added to a system that contained SDL. Disaster would result. For it follows from $\mathbf{OB} - RM$ that $\mathbf{PE}p \rightarrow \mathbf{PE}(p \vee q)$.[42] So with $*$ it would follow that $\mathbf{PE}p \rightarrow (\mathbf{PE}p \,\&\, \mathbf{PE}q)$, for any q, so we would get

$**$. $\mathbf{PE}p \rightarrow \mathbf{PE}q$,

that if anything is permissible, then everything is, and thus it would also be a theorem that nothing is obligatory, $\vdash \sim\!\mathbf{OB}p$.[43]

Some have argued for two senses of "permissibility" here.[44]

The Violability Puzzle:[45] Here is another puzzle centering around RM. It would seem that it is of the very nature of obligations that they are violable in principle, unlike simple assertions, so that the following seems to be a conceptual truth:

1. If it is logically impossible that p is false, then it is logically impossible that p is obligatory.

But in SDL, this would naturally be expressed as a rule of inference:

$$\text{If } \vdash p \text{ then } \vdash\sim\!\mathbf{OB}p \qquad \text{(Violability)}$$

But since $\top$ is a logical truth, Violability would yield $\sim\!\mathbf{OB}\top$, which directly contradicts theorem $\mathbf{OB}$-N. Thus, SDL seems to make it a logical truth that there are inviolable obligations. But the idea that it is obligatory that it is either raining or not raining, something that couldn't be otherwise on logical grounds, seems counterintuitive. Furthermore, even in a system that lacked the force of $\mathbf{OB}$-NEC and $\mathbf{OB}$-N, if it has the force of just the rule RM (if $\vdash p \rightarrow q$ then $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$), then were we to also countenance the Violability rule in such a system, we would be immediately forced to conclude that nothing is obligatory,

---

[42]This follows from RM and the definition of $\mathbf{PE}$: Suppose $\vdash p \rightarrow q$. Then $\vdash \sim\!q \rightarrow \sim\!p$. So by RM, $\vdash \mathbf{OB} \sim\!q \rightarrow \mathbf{OB} \sim\!p$, and thus $\vdash \sim\!\mathbf{OB} \sim\!p \rightarrow \sim\!\mathbf{OB} \sim\!q$, i.e. $\vdash \mathbf{PE}p \rightarrow \mathbf{PE}q$. Now just let q be $(p \vee q)$.

[43]For suppose something was obligatory, say $\mathbf{OB}p$. Then by NC, it follows that $\mathbf{PE}p$. One instance of $**$ above is $\mathbf{PE}p \rightarrow \mathbf{PE} \sim\!p$. So we would then have $\mathbf{PE} \sim\!p$, which by RE, PC and the definition of $\mathbf{PE}$ amounts to $\sim\!\mathbf{OB}p$, contradicting our assumption. Thus nothing could be obligatory.

[44]For example, one sense would be as in SDL (the simple absence of a prohibition), the other being a stronger sense of permission [von Wright, 1968] with a distinct logic that would, for example, ratify $*$, but not $**$, above. Another approach was to say that this is a pseudo-problem, since the conjunctive use of "or" in the context of a permission word can be expressed as a conjunction of permitting conjuncts, $\mathbf{PE}p \,\&\, \mathbf{PE}q$ [Føllesdal and Hilpinen, 1971]. Kamp [1974; 1979] contain detailed analyses of these issues, one sensitive to both the semantics and pragmatics of permission.

[45]von Wright [1963, p. 154] comes very close to stating this objection.

$\vdash\sim\mathbf{OB}p$, thus rendering the system inapplicable.[46]   von Wright [1963, p. 154] comes close to endorsing Violability, but the context there is more complex and less straightforward than that above. Jones and Porn [1985] provides a system designed explicitly to accommodate violability (among other things) for their analysis of "ought".

Ross's Paradox [Ross, 1941]

Consider:

1. It is obligatory that the letter is mailed.

2. It is obligatory that the letter is <u>m</u>ailed or the letter is <u>b</u>urned.

In SDLs, these seem naturally expressible as:

1'. $\mathbf{OB}m$

2'. $\mathbf{OB}(m \lor b)$

But $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}(p \lor q)$ follows by RM from $\vdash p \rightarrow (p \lor q)$. So 2' follows from 1', but it seems rather odd to say that an obligation to mail the letter entails an obligation that can be fulfilled by burning the letter (something presumably forbidden), and one that would appear to be violated by not burning it if I don't mail the letter.

The Good Samaritan Paradox [Prior, 1958] [47]

Consider

1. It ought to be the case that Jones helps Smith who has been robbed.

2. It ought to be the case that Smith has been robbed.

Now it seems that the following must be true:

> Jones helps Smith who has been robbed if and only if Jones <u>h</u>elps Smith and Smith has been <u>r</u>obbed.

But then it would appear that a correct way to symbolize 1) and 2) in SDLs is:

1'. $\mathbf{OB}(h \,\&\, r)$

---

[46]For suppose $\mathbf{OB}p$. Then since by PC, $\vdash p \rightarrow \top$, it follows by $\mathbf{OB}$-RM that $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}\top$. But since by PC, $\vdash \top$, by Violability, it follows that $\vdash\sim\mathbf{OB}\top$. So by PC, $\vdash\sim\mathbf{OB}p$, for any p.

[47]Prior cast it using this variant of RM: If $\vdash p \rightarrow q$ then $\vdash \mathbf{IM}q \rightarrow \mathbf{IM}p$ (the impermissibility of Smith being robbed then appears to wrongly imply the impermissibility of helping him who has been robbed). See also [Åqvist, 1967], which has been very influential.

2′. **OB**r.

But it is a thesis of PC that $(h \,\&\, r) \to r$, so by RM, it follows that **OB**$(h \,\&\, r) \to$ **OB**r, and then we can derive 2′ from 1′ by MP. But it hardly seems that if helping the robbed man is obligatory it follows that his being robbed is likewise obligatory.[48]

The Paradox of Epistemic Obligation [Åqvist, 1967]

This is a much-discussed variant of the preceeding paradox. Consider:

1. The bank is being <u>r</u>obbed.

2. It ought to be the case that Jones (the guard) knows that the bank is being robbed.

3. It ought to be the case that the bank is being robbed.

Let us symbolize "Jones knows that the bank is being robbed" by "$K_j r$". Then it would appear that a correct way to symbolize (1)–(3) in SDLs (augmented with a "K" operator) is:

1′. r

2′. **OB**$K_j r$

3′. **OB**r.

But it is a logical truth that if one knows that p then p is the case (surely Jones knows that the bank is being robbed only if the bank is in fact being robbed). So $\vdash K_j r \to r$ would hold in any system augmented with a faithful logic of knowledge. So in such a system, it would follow by RM that $\vdash$ **OB**$K_j r \to$ **OB**r, but then we can derive 3′ from this conditional and 2′ by MP.[49] But it hardly seems to follow from the fact that it is obligatory that the guard knows that the bank is being robbed, that it is likewise obligatory that the bank is being robbed. It seems that SDL countenances inferences from patently impermissible states of affairs that

---

[48]This paradox can also be cast equivalently with just one agent, and via **IM** as easily as **OB**: "The Victims Paradox" notes that the victim of the crime helps herself only if there was a crime. If it is impermissible that there be a crime, it will follow under similar symbolization that it is impermissible for the victim of the crime to help herself, which hardly sounds right. Similarly for "The Robber's (Repenter's) Paradox", where now we focus on the robber making amends (or repenting) for his crime, and again we seem to get the result that it is impermissible for the robber to make amends for his crime, suggesting a rather convenient argument against all obligations to ever make amends for one's crimes. These early variations were used to show that certain initially proposed solutions to the Good Samaritan Paradox didn't really solve the problem. Both versions are found in [Nowell Smith and Lemmon, 1960].

[49](1′) is not really essential here, it just helps to clarify that (2) does not express some strange standing obligation but a transient one that emerges as a result of the *de facto* robbery.

someone is obliged to know hold when they hold to the conclusion that the same impermissible states of affairs are obligatory.[50]

Some RM-Related Literature: One standard response to Ross's Paradox, the Good Samaritan Paradox (and the Paradox of Epistemic Obligation) is to try to explain them away. For example, Ross's Paradox is often quickly rejected as elementary confusion [Føllesdal and Hilpinen, 1971] or it is rejected on the grounds that the inference is only pragmatically odd in ways that are independently predictable by any adequate theory of the pragmatics of deontic language [Castañeda, 1981]. Similarly, it has been argued that the Good Samaritan Paradox is really a conditional obligation paradox, and so RM is not the real source of the paradox [Castañeda, 1981; Tomberlin 1981]. However, since these paradoxes all at least *appear* to depend on **OB**-RM, a natural solution to the problems is to undercut the paradoxes by rejecting **OB**-RM itself. Two accessible and closely related examples of approaches to deontic logic that reject **OB**-RM from a principled philosophical perspective are [Jackson, 1985] and [Goble, 1990a]. Jackson [1985] argues for an approach to "ought to be" that links it to counterfactuals, and he informally explores its semantics and logic; Goble [1990a] makes a similar case for "good" and "bad" (as well as "ought"), formally tying these to logical features of counterfactuals explicitly. ([Goble, 1990b] contains the main technical details.) Interestingly, their approaches also intersect with the issue of "actualism" and "possibilism" as these terms are used in ethical theory. Roughly, *possibilism* is the view that I ought to bring about p if p is part of the best overall outcome I *could* bring about, even if the goodness of this overall outcome, depends on all sorts of other things that I *would* not in fact bring about were I to bring about p. In contrast, *actualism* is the view that I ought to bring about p if doing so *would* in fact be better than not doing so, and this, of course, can crucially depend on what else I would do (ideal or not) were I to bring about p. (See [Jackson and Pargetter, 1986; Jackson, 1988], and for early discussions of this issue, see [Goldman, 1976] and [Thomason, 1981a].) In [Hansson, 1990], and more elaborately in [Hansson, 2001], S. O. Hansson develops systems of deontic logic where he analyzes prohibitive and prescriptive deontic notions in terms of abstract properties of various preference orderings (e.g. a normative status is prohibitive whenever anything worse than something that has that status also has it). He also sees **OB**-RM as the main culprit in the paradoxes of standard deontic logic, and thus he methodically explores non-standard frameworks where **OB**-RM is not sound. Hansson [2001] is also important for its extensive and original work on preference logic and preference structures, which, as we have already noted, are used regularly in deontic logic (and elsewhere). A very useful general source that covers some of the issues surrounding **OB**-RM, along with many others, is [van der Torre, 1997].

---

[50]Theoretically one could claim that we have a conflict of obligations here, but this seems quite implausible. The banks' being robbed appears to be definitely non-obligatory.

4.4   Puzzles Centering Around NC, OD and Analogues

Sartre's Dilemma and Conflicting Obligations [Lemmon, 1962b][51]

A conflict of obligations is a situation where there are two obligations and it is not possible for both to be fulfilled.

Consider the following conflict:

1. It is obligatory that I now meet Jones (say, as promised to Jones, my friend).

2. It is obligatory that I now do not meet Jones (say, as promised to Smith, another friend).

Here it would seem that I have a conflict of obligations, in fact a quite direct one, since what I promised one person would happen, I promised another would not happen. People do (e.g. under pressure or distraction) make such conflicting promises, and it appears that they incur conflicting obligations as a result.[52] But consider the natural representation of these in SDLs:

1′. $\mathbf{OB}j$

2′. $\mathbf{OB} \sim j$

But since NC, $\mathbf{OB}p \rightarrow \sim\mathbf{OB} \sim p$, is a theorem of all SDLs, we can quickly derive a contradiction from (1) and (2), which means that (1′) conjoined with (2′) represents a logically inconsistent situation. Yet, the original hardly seems logically incoherent.[53]

Kant's Law and Unpayable Debts[54] Here is a simple puzzle about $\mathbf{OB}p \rightarrow \Diamond p$. Consider:

1. I'm obligated to pay you back $10 tonight.

2. I can't pay you back $10 tonight (e.g. I just gambled away my last dime).

Since this puzzle typically involves some notion of possibility, let us represent the above sentences in KTd, which includes SDL, but also has a possibility operator:

---

[51]I change the example to have the conflict be direct and explicit. Sartre's much cited example is of a man obligated to join the resistance (to avenge his brother's death and fight the Nazi occupation) and obligated to stay home and aid his ailing mother (devastated by the loss of the man's brother, her son, and deeply attached to the one son still alive).

[52]Whether or not these obligations are both all-things-considered-obligations is a further issue. For our purposes here, the point is that they appear to be obligations. See the upcoming puzzle, Plato's Dilemma, for further issues.

[53]von Wright [1968] refers to a conflict of obligations as a "predicament" and illustrates with the much-cited example of Jephthah (from the *Book of Judges*), who promises God to sacrifice the first living being he meets upon returning home from war, if God gives him victory, which wish is granted, but his daughter is the first living being he meets upon his return.

1′. **OB**p

2′. $\sim\lozenge p$.

(1) and (2) appear to be consistent. It seems to be a sad fact that often, people are unable to fulfill their financial obligations, just as it seems to be a truism that financial obligations are obligations. But in KTd, it is a theorem that **OB**$p \rightarrow \lozenge p$. So we derive a contradiction from this symbolization and the assumption that 1′ and 2′ are true.

A variant example is:

1. I owe you ten dollars, but I can't pay you back.

2. I'm obligated to pay you ten dollars, but I can't.

(2) seems to follow from (1), and (1) hardly seems contradictory, since owing money clearly does not entail being able to pay the money owed. Thomason [1981b] suggests a distinction between deliberative contexts of evaluation and judgmental contexts, where in the latter context evaluations such as 1) above need not satisfy Kant's law since, roughly, we go back in time and evaluate the present in terms of where things would now be relative to optimal past options that were accessible but no longer are.

---

[54]*Kant's* law is more accurately rendered as involving agency (if Doe is obligated to bring something about then Doe is able to do so), but the label is often used in deontic logic for almost any implication from something's being obligatory to something's being possible, roughly whatever formula comes closest to Kant's in the system.

Conflation of Conflicts with Impossible Obligations: Here is another puzzle associated with NC and OD, one showing that SDL conflates logically distinct situations.

We saw above that Kant's Law, when represented as $\mathbf{OB}p \rightarrow \Diamond p$, is a theorem of KTd. If we interpret possibility here as practical possibility, then as the indebtedness example above suggests, it is far from evident that it is in fact true. However, a stronger claim than that of Kant's Law is that something cannot be obligatory unless it is at least *logically* possible. In SDL, this might be expressed by the rule:

$$\text{If } \vdash \sim p \text{ then } \vdash \sim \mathbf{OB}p.$$

This is derivable in SDL, since if $\vdash \sim p$, then $\vdash \mathbf{OB} \sim p$ by $\mathbf{OB}$-NEC, and then by $\mathbf{OB}$-NC, we get $\vdash \sim \mathbf{OB}p$. Claiming that Romeo is obligated to square the circle because he solemnly promised Juliet to do so is less convincing as an objection than the earlier financial indebtedness case. So SDL is somewhat better insulated from this sort of objection, and, as we noted earlier, we are confining ourselves here to theories that endorse $\mathbf{OB}$-OD (i.e. $\vdash \sim \mathbf{OB}\bot$).[55]

However, this points to another puzzle for SDL. The rule above is equivalent to $\vdash \mathbf{OB}\text{-}OD$ in any system with $\mathbf{OB}\text{-}RE$, and in fact, in the context of SDL, these are both equivalent to $\mathbf{OB}$-NC. That is, we could replace the latter axiom with either the former rule above or $\mathbf{OB}\text{-}OD$ to get a system equivalent to SDL. In particular, in any system with K and RE, $(\mathbf{OB}p \,\&\, \mathbf{OB} \sim p) \leftrightarrow \mathbf{OB}\bot$ is a theorem.[56] But it seems odd that there is no distinction between a contradiction being obligatory, and having two distinct conflicting obligations. It seems that one can have a conflict of obligations without it being obligatory that some logically impossible state of affairs obtains. A distinction seems to be lost here. Separating $\mathbf{OB}$-NC from $\mathbf{OB}$-D is now quite routine in conflict-allowing deontic logics.

Some early discussions and attempted solutions to the last two problems can be found in [Chellas, 1980] and [Schotch and Jennings, 1981], both of whom use non-normal modal logics for deontic logic.[57] Brown [1996b] uses a similar approach to Chellas' for modeling conflicting obligations, but with the addition of an ordering relation on obligations to model the relative stringency of obligations, thus moving in the direction of a model addressing Plato's Dilemma as well.

---

[55]However see [Da Costa and Carnielli, 1986] which develops a deontic logic in the context of paraconsistent logic.

[56]For first suppose $\mathbf{OB}p \,\&\, \mathbf{OB} \sim p$ holds. Then one instance of K is $\mathbf{OB}(\sim p \rightarrow \bot) \rightarrow (\mathbf{OB} \sim p \rightarrow \mathbf{OB}\bot)$. But $\mathbf{OB}\text{-}RE, \mathbf{OB}(\sim p \rightarrow \bot)$ is equivalent to just $\mathbf{OB}p$, so we get $\mathbf{OB}p \rightarrow (\mathbf{OB} \sim p \rightarrow \mathbf{OB}\bot)$ by PC. So given $\mathbf{OB}p \,\&\, \mathbf{OB} \sim p$, we get $\mathbf{OB}\bot$ by PC. Second assume $\mathbf{OB}\bot$. By $PC, \vdash \bot \rightarrow p$. So by RM, we get $\mathbf{OB}\bot \rightarrow \mathbf{OB}p$, and then $\mathbf{OB}p$. We can then generate $\mathbf{OB} \sim p$

Let me note that a long-ignored and challenging further puzzle for conflicting obligations, called "van Fraassen's Puzzle" [van Fraassen, 1973], has deservedly received increasing attention of late: [Horty, 1994; 2003; van der Torre and Tan, 2000; McNamara, 2004a; Hansen, 2004] and [Goble, Forthcoming-a].

The Limit Assumption Problem: Recall our earlier mention of an ordering semantics approach to SDL, and our mention there of the Limit Assumption:

> that for each world i, there is always at least one world as good
> (relative to i) as all worlds in i's purview (i.e. one i-best world).

Although some in deontic logic have operated as if the Limit Assumption is true, it is a questionable assumption to make, especially as a matter of logic. It seems that there are possible scenarios in which the ordering of worlds in the purview of some world i have no upper bound on their goodness. Blake Barley gave a nice example in an unpublished paper, "The Deontic Dial", circulated at the University of Massachusetts-Amherst in the early 1980s: you have a dial that you can turn anywhere from 0 to 1, where both 0 and 1 yield disaster, but all the numbers in between yield better and better value, increasing with the natural order of the numbers (cf. [McMichael, 1978]). In such a case, there seems to be no real sense to the old maxim: "Do the best you can!". This rules out the most natural simple clause for **OB** per optimizing theories:

> **OB**p is true at i iff p holds in all the i-best worlds,

for plainly in scenarios where there are no i-best worlds, everything is obligatory and nothing is permissible by this clause, but this seems wrong: even in the dial case, it seems clearly not obligatory to turn the dial to 1.

Lewis [1973] famously argued that the Limit Assumption (as used here or as used for modeling counterfactuals) is an unjustified assumption, and that our clauses for deontic (and counterfactual) operators must reflect this fact. Most logicians agreed. This led to more complex clauses such as the one used earlier:

> **OB**p is true at i iff p is true from somewhere on up
> in the subset of worlds in W ordered relative to i,

---

the same way.

[57] Normal modal logics won't do since K and RE hold in all such logics. Chellas uses minimal models and Schotch and Jennings generalize Kripke models.

and this in turn leads to greater complexity in the metatheory for such logics. The new clause has some odd features, for example, in the case of the dial, for each number between 0 and 1, you are obligated to turn the dial past that number, but in the scenario, the set of such obligatory things together entail that you turn the dial to 1, which is also forbidden. Although no conflict will show up in the system (no formula of the form **OB**p and **OB** $\sim p$ will be validated), you nonetheless have an infinite set of obligations which cannot be jointly fulfilled, and thus a conflict of sorts, for which you can hardly be faulted. Lewis [1978] argues contra [McMichael, 1978], that the related problem McMichael there refers to (called "The Confinement Problem" in [McNamara, 1995]) is a problem for utilitarianism, not for deontic logic; but see [Fehige, 1994; p. 42], who suggests that there are still choices a logician must make and that "...When the best options are lacking, then so are flawless accounts of the lack". Fehige provides a systematic critical discussion of deontic logicians approaches to the Limit Assumption.

Plato's Dilemma and Defeasible Obligations [Lemmon, 1962b] [58]

1. I'm obligated to meet you for a light lunch meeting at the restaurant.

2. I'm obligated to rush my choking child to the hospital.

Here we seem to have an indirect conflict of obligations, if we assume that satisfying both obligations is practically impossible. Yet here, unlike in our prior example, where the two promises might naturally have been on a par, we would all agree that the obligation to help my child *overrides* my obligation to meet you for lunch, and that the first obligation is *defeated* by the second obligation, which *takes precedence.* Ordinarily, we would also assume that no other obligation overrides my obligation to rush my son to the hospital, so that this obligation is an *all things considered obligation*, but not so for the obligation to meet you for lunch. Furthermore, we are also prone to say that the situation is one where the general obligation we have to keep our appointments (or to keep our promises, still more generally) has an *exception* — the circumstances are extenuating. Once we acknowledge conflicts of obligation, there is the further issue of representing the logic of reasoning about conflicting obligations where some override others, some are defeated, some are all things considered obligations, some are not, some hold generally, but not unexceptionally, etc. So the issue here is one of conflicting obligations of different weight and the defeasibility of one of two obligations. Clearly, there is no mechanism in SDL for this, since SDL does not allow for conflicts to begin with, yet this is an issue that goes well beyond that of merely having a logic

---

[58]Here too I change the example. Plato's case involves returning weapons as promised to someone who now in a rage intends to unjustly kill someone with the weapon. Lemmon interprets the issues raised by Sartre's dilemma a bit differently than I do here.

that allows for conflicts. There have been a variety of approaches to this dilemma, and to defeasibility among conflicting obligations.

Some Literature on Defeasible Obligations: von Wright [1968] suggested that minimizing evil is a natural approach to conflict resolution, thereby suggesting that a sort of minimizing (and thus reliance on an ordering) is apt. Alchourron and Makinson [1981] provide an early formal analysis of conflict resolution via partial orderings of regulations and regulation sets. Chisholm [1964] has been very influential conceptually, as witnessed, for example, by [Loewer and Belzer, 1983]. In ethical theory, the informal conceptual landmark is [Ross, 1939]. Horty [1994] is a very influential discussion forging a link between Reiter's default logic developed in AI (see [Brewka, 1989]), and an early influential approach to conflicts of obligation, [van Fraassen, 1973], which combines a preference ordering with an imperatival approach to deontic logic. Prakken [1996] discusses Horty's approach and an alternative that strictly separates the defeasible component from the deontic component, arguing that handling conflicts should be left to the former component only. See also [Makinson, 1993] for a sweeping discussion of defeasiblity and the place of deontic conditionals in this context. Other approaches to defeasibility in deontic logic that have affinities to semantic techniques developed in artificial intelligence for modeling defeasible reasoning about defeasible conditionals generally are [Asher and Bonevac, 1996] and [Morreau, 1996], both of which attempt to represent W. D. Ross-like notions of *prima facia* obligation, etc. Also notable are the discussions of defeasibility and conditionality in [Alchourron, 1993; 1996], where a revision operator (operating on antecedents of conditionals) is relied on in conjunction with a strict implication operator and a strictly monadic deontic operator. Note that [Alchourron, 1996; Prakken, 1996; Asher and Bonevac, 1996; Morreau, 1996] and [Prakken, 1996] are all found in *Studia Logica* **57**, 1996 (guest edited by A. I. J. Jones and M. Sergot). Nute [1997] is dedicated to defeasibility in deontic logic and is the best single source on the topic, with articles by many of the key players, including Nute himself. See [Bartha, 1999] for an approach to contrary-to-duty conditionals and to defeasible conditionals layered over a branching time framework with an agency operator. Smith [1994] contains an interesting informal discussion of conflicting obligations, defeasibility, violability and contrary-to-duty conditionals. Since it is very much a subject of controversy and doubt as to whether deontic notions contribute anything special to defeasible inference relations (as opposed to defeasible conditionals), we leave this issue aside here, and turn to conditionals, and the problem in deontic logic that has received the most concerted attention.

4.5   Puzzles Centering Around Deontic Conditionals

The Paradox of Derived Obligation/Commitment [Prior, 1954]

Consider the following statements:

1. (a) Bob's promising to meet you commits him to meeting you.

   (b) It is obligatory that Bob meets you if he promises to do so.

It was suggested that these might be represented in either of two ways in SDL:

1'. $p \to \mathbf{OB}m$[59]

1''. $\mathbf{OB}(p \to m)$.[60]

Consider (1') first. The following are both simply tautologies: $\sim r \to (r \to \mathbf{OB}s)$ and $\mathbf{OB}s \to (r \to \mathbf{OB}s)$. So if 1' reflected a proper analysis of 1a/b), anything false would commit us to anything whatsoever (e.g. since I am not now standing on my head, it would follow that my standing on my head commits me to giving you all my money) and everything commits us to anything obligatory (e.g. if I'm obligated to call you, then my standing on my head commits me to doing so). What of 1'' then? The following are theorems of SDL: $\mathbf{OB} \sim r \to \mathbf{OB}(r \to s)$ and $\mathbf{OB}s \to \mathbf{OB}(r \to s)$. So if 1'' reflected an apt analysis of commitment, it would follow from SDL that anything impermissible commits us to everything, and once again, everything commits us to anything obligatory. So, these seem to be troublesome candidates for symbolizing 1a) or 1b) in SDL. The problems are reminiscent of paradoxes about material implication (reading 1'), and strict implication (reading 1''), respectively.[61] So the question arose, are there any *special* problems associated with the interaction of deontic notions and conditionality? The next paradox (Chisholm's), increased the perception that there might very well be. Many consider it to be the most challenging and distinctive puzzle of deontic logic.

Contrary-to-Duty (or Chisholm's) Paradox [Chisholm, 1963a]

Consider the following:

1. It ought to be that Jones goes to the assistance of his neighbors.

2. It ought to be that if Jones does go then he tells them he is coming.

3. If Jones doesn't go, then he ought not tell them he is coming.

---

[59]In the 1st edition of Prior [1962 [1955]].

[60]In [von Wright, 1951].

[61]In the case of symbolization 1'', since $(r \to s)$ is logically equivalent to $(\sim r \vee s)$, and the two troublesome formulas associated with this symbolization reduce to $\mathbf{OB} \sim r \to \mathbf{OB}(\sim r \vee s)$ and $\mathbf{OB}s \to \mathbf{OB}(\sim r \vee s)$, these are also instances of Ross' Paradox given this SDL interpretation of the sentences.

4. Jones doesn't go.

This certainly appears to describe a possible situation. It is widely thought that (1)–(4) constitute a *mutually consistent* and *logically independent* set of sentences. We treat these two conditions as desiderata. Note that (1) is a *primary obligation*, saying what Jones ought to do unconditionally.[62] (2) is a *compatible-with-duty obligation*, appearing to say (in the context of (1)) what else Jones ought to do on the condition that Jones fulfills his primary obligation. In contrast, (3) is a *contrary-to-duty obligation or "imperative" (a "CTD")* appearing to say (in the context of (1)) what Jones ought to do conditional on his violating his primary obligation. (4) is a factual claim, which conjoined with (1), implies that Jones *violates* his primary obligation. Thus this puzzle also places not only deontic *conditional* constructions, but the *violability* of obligations, at center stage. It raises the challenging question: what constitutes proper reasoning about what to do in the face of violations of obligations?

How might we represent this quartet in SDL? The most straightforward symbolization is;

1′. **OB**g.

2′. **OB**$(g \to t)$.

3′. $\sim g \to$ **OB** $\sim t$.

4′. $\sim g$.

But Chisholm points out that from (2′) by principle **OB**-K we get **OB**$g \to$ **OB**t, and then from (1′) by MP, we get **OB**t; but by MP alone we get **OB** $\sim t$ from (3′) and (4′). From these two conclusions, by PC, we get $\sim($**OB**$t \to \sim$ **OB**$\sim t)$, contradicting NC of SDL. Thus (1′)–(4′) leads to inconsistency per SDL. But (1)–(4) do not seem inconsistent at all, so the representation cannot be a faithful one. Various less plausible representations in SDL are similarly unfaithful. For example, we might try reading the second and third premises uniformly, either on the model of (2′) or on the model of (3′). Suppose that instead of (3′) above, we use (3″) **OB**$(\sim g \to \sim t)$. The trouble with this is (3″) is derivable from (1′) in SDL, but there is no reason to think (3) in fact follows from (1), so we have an unfaithful representation again. Alternatively, suppose that instead of (2′) above, we use (2″) $g \to$ **OB** $\sim t$. This is derivable from (4′) in PC (and thus in SDL). But there is no reason to think (2) follows from (4). So again, we have an unfaithful representation.

The following displays in tabular form the difficulties trying to interpret the quartet in SDL:[63]

---

[62]Here we follow tradition (albeit self-consciously) in sloughing over the differences between what ought to be, what one ought to do, and what is obligatory.

[63]The remaining truly strained combination would replace 2′ with 2″ *and* 3′ with 3″, but that just doubles the trouble with the second and third readings, so it is routinely ignored.

First Try :   Second Try: Third Try:
1′. **OB**g 1′. **OB**g 1′. **OB**g
2′. **OB**$(g \rightarrow t)$ 2′. **OB**$(g \rightarrow t)$ 2″. $g \rightarrow$ **OB**t
3′. $\sim g \rightarrow$ **OB** $\sim t$ 3″. **OB**$(\sim g \rightarrow \sim t)$ 3′. $\sim g \rightarrow$ **OB** $\sim t$
4′. $\sim g$ 4′. $\sim g$ 4′. $\sim g$

From 1′, 2′, **OB**t.   3″ follows from 1′.   2″ follows from 4′.
From 3′, 4′, **OB** $\sim t$. So independence is lost.   So independence is lost.
By NC, **OB**$t \rightarrow \sim$**OB** $\sim t$.
So $\perp$; consistency is lost.

Each reading of the original violates one of our desiderata: mutual consistency or joint independence.

If von Wright launched deontic logic as an academic specialization, Chisholm's Paradox was the booster rocket that provided the escape velocity deontic logic needed from subsumption under normal modal logics, thus solidifying deontic logic's status as a distinct specialization. It is now virtually universally acknowledged that Chisholm was right: the sort of conditional deontic claim expressed in 3) can't be faithfully represented in SDL, nor more generally by a composite of some sort of unary deontic operator and a material conditional. This is one of the few areas where there is nearly universal agreement in deontic logic. Whether or not this is because some special *primitive dyadic deontic conditional* is operating or because it is just that some *non-material* conditional is essential to understanding important deontic reasoning is still a hotly contested issue.

Some Literature on Contrary-to-Duty Obligations: von Wright [1956; 1971] take the now-classic non-componential dyadic operator approach to the syntax of CTDs. Danielsson [1968], Hansson [1969], Lewis [1973; 1974], and Feldman [1986] provide samples of a "next best thing" approach: the interpretation of conditional obligations via a primitive non-componential dyadic operator, in turn interpreted via a preference ordering of the possible worlds where the (perhaps obligation-violating) antecedent holds; see also [Åqvist, 2002 [1984]] for an extensive systematic presentation of this sort of approach (among other things), and [al-Hibri, 1978] for an early widely-read systematic discussion of a number of approaches to CTDs (among other things). van Fraassen [1972], Loewer and Belzer [1983], and Jones and Porn [1985] also offer influential discussions of CTDs and propose distinct formal solutions, each also employing orderings of outcomes, but offering some twists on the former more standard pictures. An important forthcoming source on the metatheory of classical and near-classical logics via classic and near-classic ordering structures for the dyadic operator is [Goble, Forthcoming-b]. Mott [1973] and [Chellas, 1974] (and [Chellas, 1980]) offer influential analyses of the puzzle by combining a *non-material*

conditional and a *unary* deontic operator to form a genuine componential compound, $p \Rightarrow \mathbf{OB}q$, for representing conditionals like (3) above; [DeCew, 1981] is an important early critical response to this sort of approach. Tomberlin [1983] contains a very influential informal discussion of various approaches. Bonevac [1998] is a recent argument against taking conditional obligation to be a primitive non-componential operator, suggesting roughly that techniques like those developed in AI (see [Brewka, 1989]) for defeasible reasoning suffice for handling woes with CTDs. Smith [1993; 1994] contain important discussions stressing the difference between violability and defeasibility, and the relevance of the former rather than the latter to CTDs. Åqvist and Hoepelman [1981], and van Eck [1982] (and again, [Loewer and Belzer, 1983]) are classic representatives of attempts to solve the puzzle by incorporating temporal notions into deontic logic. Jones [1990] contains an influential argument against any temporal-based general solution to the puzzle. Castañeda [1981] argued that by carefully distinguishing between (roughly) propositions and actions in the scope of deontic operators, Chisholm's puzzle, as well as most puzzles for deontic logic, can be resolved; Meyer [1988] offers a version of this general approach using dynamic logic. Prakken and Sergot [1996] contains an influential argument against any such action-based general solution to the puzzle. For recent work on CTDs in the context of a branching time framework with agency represented *a la* Horty–Belnap, see [Horty, 1996; 2001; Bartha, 1999], and Bartha's contribution to [Belnap, 2001; Chapter 11]. A recent source that reviews a good deal of the literature on CTDs and proposes its own solution is [Carmo and Jones, 2002]; but see also material on this problem in [Nute, 1997] (especially [van der Torre and Tan, 1997], and [Prakken and Sergot, 1997]).

Appendix A2 contains additional discussion of this very important paradox. One newer puzzle often discussed in either the context of **OB**-RM or in the context of discussing conditional obligations is the following.

The Paradox of the Gentle Murderer [Forrester, 1984] [64]

Consider:

1. It is obligatory that John Doe does not kill his mother.

2. If Doe does <u>kill</u> his mother, then it is obligatory that Doe kills her <u>gently</u>.

3. Doe does kill his mother (say for an inheritance).

Then it would appear that a correct way to symbolize (1) and (2) in SDLs is:

1′. **OB** $\sim k$

2′. $k \rightarrow \mathbf{OB}g$

---

[64]Also called "Forrester's Paradox".

$3'$. k.

First, from $2'$ and $3'$, it follows that **OB**g by MP. But now add the following unexceptionable claim:

Doe kills his mother gently only if Doe kills his mother.

Assuming this, symbolized as $g \to k$, is a logical truth in an expanded system, by **OB**-RM it follows that **OB**$g \to$ **OB**k, and so by MP again we get **OB**k. This seems bad enough, for it hardly seems that from the fact that if I kill my mother then I must kill her gently and that I will kill her (scoundrel that I am), we can conclude that I am actually obligated to kill my mother. Add to this that from **OB**k in turn, we get $\sim$**OB** $\sim k$ by NC of SDL, and thus we have a contradiction as well. So we must either construe (2) so that it does not satisfy *modus ponens* or we must reject **OB**-RM.[65]

4.6 Problems Surrounding (Normative) Expressive Inadequacies of SDL

Here we look at some monadic normative notions that appear to be inexpressible in SDL.

The Normative Gaps Puzzle [von Wright, 1968] [66]

In some normative systems, permissions, prohibitions and obligations are explicitly given. So it would seem to be possible for there to be normative systems with gaps: where something is neither obligatory, impermissible, nor permissible. Yet **OB**$p \lor$ (**PE**p & **PE** $\sim p$) $\lor$ **IM**p is a thesis ("exhaustion") of SDL (given the Traditional Definitional Scheme), which makes any such gaps impossible.

Urmson's Puzzle — Indifference versus Optionality [1958]

Consider:

1. It is optional that you attend the meeting, but not a matter of indifference that you do so.

---

[65] Some have suggested this is a problem stemming from scope difficulties, others have argued that the problem is that **OB**-RM is in fact invalid, and rejecting it solves the problem. ([Sinnott-Armstrong, 1985] argues for a scope solution; [Goble, 1991] criticizes the scope solution approach, and argues instead for rejecting **OB**-RM.) We have listed this puzzle here rather than under the Good Samaritan Puzzle (in turn under puzzles associated with **OB**-RM) since, unlike the Standard Good Samaritan, this puzzle seems to crucially involve a contrary-to-duty conditional, and so it is often assumed that a solution to the Chisholm Paradox should be a solution to this puzzle as well (and vice versa). Alternatively, one might see the puzzle as one where we end up obligated to kill our mother gently because of our decision to kill her (via factual detachment), and then by **OB**-RM, we would appear obligated to kill her, which has no plausibility by anyone's lights, and thus calls for rejecting **OB**-RM. However, this would still include a stance on contrary-to-duty conditionals and detachment.

[66] See also [Alchourron and Bulygin, 1971].

This seems to describe something quite familiar: optional matters that are nonetheless not matters of indifference. But when deontic logicians and ethicists gave an operator label for the condition ($\sim$**OB**$p\,\&\,\sim$**OB** $\sim p$), it was almost invariably "It is indifferent that p", "**IN**p". But then it would seem to follow from the theorem **OB**$p \vee (\sim$**OB**$p\,\&\,\sim$**OB** $\sim p) \vee$**IM** $\sim p$, that $(\sim$**OB**$p\,\&\,\sim$**IM**$p) \rightarrow$**IN**p, that is, everything that is neither obligatory nor prohibited is a matter of indifference. But many actions, including some heroic actions, are neither obligatory nor prohibited, yet they are hardly matters of indifference. We might put this by saying that SDL can represent optionality, but not indifference, despite the fact that the latter concept has been a purported target for representation since nearly its beginning (see also [Chisholm, 1963b] and [McNamara, 1996a]).

The Supererogation Problem [Urmson, 1958]

Some things are beyond the call of duty or supererogatory (e.g. volunteering for a costly or risky good endeavor where others are equally qualified and no one person is obligated). SDL has no capacity to represent this complex concept.[67]

The Must versus Ought Dilemma [McNamara, 1990; 1996c]

Consider:

1. Although you can skip the meeting, you ought to attend.[68]

This seems perfectly possible, even in a situation where no conflicting obligations are present, as we will suppose here. 1) appears to imply that it is optional that you attend — that you *can* attend and that you *can* fail to attend. It seems clear that the latter two uses of "can" express permissibility. Yet "ought" is routinely the reading given for deontic necessity in deontic logic (and in ethical theory), and then "permissibility" is routinely presented as its dual. But then if we symbolize 1) above accordingly, we get,

1′. **PE** $\sim p\,\&\,$**OB**p

which is just $\sim$**OB**$p\,\&\,$**OB**p in disguise (given **OB**-RE and the Traditional Definitional Scheme). So (1′), given **OB**-NC, yields a contradiction. Another way to put this is that the "can" of permissibility is much more plausibly construed as the dual of "must" than as the dual of "ought". This yields a dilemma for standard deontic logic (really for most work in deontic logic):

> Either deontic necessity represents "ought", in which case, its dual does not represent permissibility (and neither does any other construction in SDL), or permissibility is represented in SDL, but "ought" is inexpressible in it despite the ubiquitous assumption otherwise.

---

[67]For some attempts to accommodate supererogation in deontic logic, see [Chisholm, 1963b; Chisholm and Sosa, 1966; Humberstone, 1974; Forrester, 1975; Sajama, 1985; Hrushka and Joerden, 1987; McNamara, 1990; 1996a; 1996b; 1999].

[68]Cf. [Chisholm, 1963b].

That "ought" is the dual of permissibility is really a largely overlooked pervasive bipartisan presupposition in both ethical theory and deontic logic.[69]

The Least You Can Do Problem [McNamara, 1990]

1. You should have come home on time; the least you could have done was called, and you didn't do even that.

The expression in the second clause has been completely ignored in the literature on deontic logic and ethical theory both. 1) appears to express the idea that there is some minimal but acceptable alternative (and the criticism suggested in the emphatic third clause is that not even that minimal acceptable option was taken, much less the preferable option identified in the first clause using "ought"). This notion of what is minimally acceptable among the permissible options is not expressible in SDL.

Regarding the last four problems, McNamara [1990; 1996a; 1996c; 1996b] and Mares and McNamara [1997] attempt to devise logics distinguishing "must" from "ought", indifference from optionality, as well as distinctly representing "the least you can do" idiom, and using this unstudied idiom to analyze one central sense of "supererogation". Appendix A3 contains a sketch of this framework for common sense morality.

4.7 Agency in Deontic Contexts

We routinely talk about both what *ought to be* and what people *ought to do*. These hardly look like the same things (for example, the latter notion calls for an agent, the former does not). This issue, and the general issue of representing agency in deontic logic has been much discussed, and continues to be an area of active concern.

The Jurisdictional Problem and the Need for Agency[70]

---

[69]Jones and Porn [1986] gives an early attempt to distinguish the two, although "must" ends up looking more like *practical* necessity in their framework (that which holds in all scenarios — permissible or not) than *deontic* necessity. McNamara [1990; 1996c] provide cumulative case arguments that "must" not "ought" is the dual of permissibility, and thus that it is this almost universally ignored term "must", not "ought", that tracks the traditional concern in ethical theory and deontic logic with permissibility. Forrester [1975] is an early attempt to sketch an operator scheme distinguishing "ought" from "obligatory".

[70]The following formulation of the problem has the status of reconstructed deontic folklore in the form of an argument or problem explicitly showing the inability of SDL to be taken to represent agential obligations. The need to eventually represent agency in order to represent agential obligations was so widely recognized early on that arguments for it are hard to find. The earliest reference I have found that comes close to formulating the problem in just the following way is [Lindahl, 1977, p. 94], which explicitly uses the "none of your business" terminology. However, it was surely known to Kanger, and fair to say it was presupposed by him in his attempted analysis of rights-related notions as far back as his seminal paper, Kanger [1971 [1957]]. Cf. also [von Wright, 1968].

Consider the following:

1. Jane Doe is obligated to not bring it about that *your child is disciplined.*

2. Jane Doe is obligated to not bring it about *that your child is not disciplined.*

Suppose you have a child. For almost any Jane Doe, (1) is then true: she is obligated to not bring it about that *your* child is disciplined, since that is none of her business. Similarly, for 2): she also is obligated to not bring it about that *your* child is not disciplined, since bringing that about is also none of her business. How might we represent these in SDL? Suppose we try to read the **OB** of SDL as "Jane Doe is obligated to bring it about that __";[71] then how do we express (1) and (2)? The closest we appear to be able to come is:

1′. $\sim$**OB**p.

2′. $\sim$**OB** $\sim p$.

But these won't do. Collectively, (1′) and (2′) amount to saying that two obligations are absent, that it is neither obligatory that Jane Doe brings it about that your child is disciplined nor obligatory that she brings it about that your child is not disciplined. But this is compatible with its being the case that both (1) and (2) above are false. After all, suppose now that you are Jane Doe, the single parent of your child. Then in a given situation, it may be that you, the child's sole parent and guardian, are both permitted to bring it about that the child is disciplined and permitted to bring it about that the child is not disciplined, in which case both (1) and (2) are false. (These permissions in fact appear to be equivalent to the negations of (1) and (2).) But the falsity of (1) (and the first permission) implies the truth of (2′) on the current reading, and the falsity of (2) (and the second permission) implies the truth of (1′) on the current reading. So clearly (1) and (1ı) are not equivalent, nor are (2) and (2′).

Alternatively, on the proposed reading of **OB**, shifting the outer negation signs to the right of the operators in (1′) and (2′) will just get us this conflicting pair:

1″. **OB** $\sim p$

2″. **OB** $\sim\sim p$

which are hopeless candidates for symbolizing (1) and (2), which do not conflict with one another.

Also, consider this traditional equivalence:

$$\textbf{IM}p \leftrightarrow \textbf{OB} \sim p.$$

If we are going to read "**OB**" as having agency built into it, presumably we want to do the same for the other operators, and so **IM**p above will be read as "it is

impermissible for Jane Doe to bring it about that p". However, this renders the left to right implication in the equivalence above unsound, for it may be true that it is impermissible for me to discipline your child, but false that it is obligatory for me to see to it that your child is positively not disciplined. The matter must be left up to you.

On the face of it, the "not"s in (1) and (2) are not external to the deontic operators, as it were, nor are they directly operating on p; rather they pertain to Jane Doe's agency with respect to p. They come "between" the deontic element and the agential element, so reading **OB** as an amalgamation of a deontic and agential operator does not allow for the "insertion" of any such negation. So, unsurprisingly, it looks like we simply must have some explicit representation of agency if we are to represent agential obligations like those in (1) and (2) above.

A Simple Kangerian Agency Framework

So let us introduce a standard operator for this missing element,

BA: Jane Doe brings it about that __.[72]

Then clearly the following relations expressing an agent's simple position with respect to a proposition, p, are to be distinguished:

BAp: Jane Doe brings it about that p
$BA \sim p$:   Jane Doe brings it about that $\sim p$
$\sim BAp$:   Jane Doe does not bring it about that p
$\sim BA \sim p$: Jane Doe does not bring it about that $\sim p$.

Plainly, if neither of the first two hold, then the conjunction of the last two holds. In such a case we might say that Jane Doe is *passive* with respect to p, or more adequately, passive with respect to herself bringing about p or its negation.[73] Let's introduce such an operator:

$$PVp =_{df} \sim BAp \,\& \, \sim BA \sim p.$$

Clearly we have here another potential set of modal operators, and we can introduce rough analogues to our traditional definitional schemes for alethic modal operators and deontic operators as follows:

$$ROp =_{df} BA \sim p$$
$$NRp =_{df} \sim BA \sim p$$
$$NBp =_{df} \sim BAp$$
$$PVp =_{df} \sim BAp \,\& \, \sim BA \sim p.$$

---

[72] "E" is often used for this operator. With two or more agents, we would need to represent agents explicitly: $BA_s p, BA_{s'} p$, etc.

[73] This "passivity" terminology, although used elsewhere, is perhaps not ideal and can't seriously be viewed as an analysis of "passivity" *per se*, since one might bring about neither a proposition nor its negation, and yet be quite influential regarding it (e.g. intentionally and actively increasing its probability without making it happen), thus the longer and more cumbersome expression.

The first says that *it is ruled out by what our agent does that* p if and only if our agent brings it about that $\sim p$. Note that this notion does not apply to all things that are ruled out *per se*, but only to those that are specifically ruled out *by* our agent's exercise of her agency. So contradictions, the negations of laws of nature and of past events, are not ruled out *by* what our agent now does. The second says *it is not ruled out by anything our agent does that* p if and only if our agent does not bring it about that $\sim p$. Laws of logic (which are necessarily ruled in) as well as contradictions (which are necessarily ruled out) are not things that are ruled out *by* our agent. The third says *our agent does not bring it about that* p (p is not ruled *in* by anything our agent does) if and only if it is false that our agent brings it about that p. This is, of course, compatible with p's holding for some other reason, such as that it is a law of logic or nature, or because it holds due to another person's exercise of her own agency. The fourth says *our agent is passive regarding* p (does nothing herself that determines the status of p) if and only if our agent neither brings about p nor rules p out by what she does do (if anything). Again, it does not follow from the fact that our agent leaves something open that it is open *per se*. PVp is consistent with its being fixed that p and consistent with its being fixed that $\sim p$, as long as neither is fixed *by* anything our agent has done. These notions are all intended to have a strong agential reading.

It is quite plausible to think that the first five agential operators satisfy the conditions of the traditional square and the traditional threefold classification scheme:
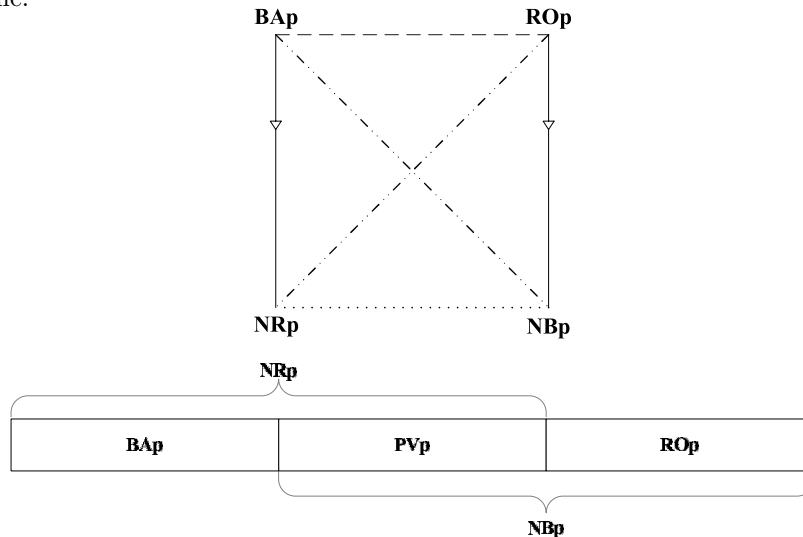
For example, in the latter case, for every agent Jane Doe, and any proposition, p, either Doe brings about p, or Doe brings about $\sim p$, or Doe brings about neither, and furthermore, no more than one of these three can hold (i.e. the three are mutually exclusive and jointly exhaustive). We will come back to this in a moment.

Virtually all accounts of this operator take it to satisfy the rule,

If $p \leftrightarrow q$ is a theorem, so is $BAp \leftrightarrow BAq$ (BA-RE),

as well as the scheme,

$BAp \rightarrow p \ (BA-T)$

(if an agent brings about p, then p holds — "success" clause), and the schema,

$(BAp \& BAq) \rightarrow BA(p \& q) \ (BA\text{-}C)$[74]

It is also the majority opinion that this operator satisfies this scheme:

$\sim BA\top (BA\text{-}NO).$

Consider again:

1. BAp $1')$ $\sim BAp$

2. $BA \sim p,$ $2')$ $\sim BA \sim p,$

and consider pairing these with one another. Pruning because of the commutativity of conjunction, we get six combinations:

a.	$BAp \& BA \sim p.$	(Contradiction given BA-T axiom)
b.	$BAp \& \sim BAp.$	(PC contradiction)
c.	$BAp \& \sim BA \sim p.$	(The second clause is implied by the first)
d.	$BA \sim p \& \sim BAp.$	(The second clause is implied by the first)
e.	$BA \sim p \& \sim BA \sim p.$	(PC contradiction)
f.	$\sim BAp \& \sim BA \sim p.$	(i.e. PVp).

Recall that because of the BA-T axiom, (1) above implies $(2')$, and (2) implies $(1')$. So the following three pruned down statuses for a proposition, p, and an agent, s, are the only pairs that remain of the six above (redundancies are also eliminated):

$$BAp,$$
$$BA \sim p,$$
$$PVp.$$

For the reasons alluded to already, it is easy to prove (using the above principles) that these three statuses (regarding an agent) and a proposition, p, are indeed mutually exclusive and jointly exhaustive, as anticipated.

---

[74]Where here we read the antecedent as implying that **BA**p and **BA**q both *now* hold.

Inaction versus Refraining/Forebearing: Another operator of considerable pre-theoretic interest is briefly discussed here. It can be defined via a condition involving embedding of "BA":

$$RFp =_{df} BA \sim BAp.$$

This expresses a widely endorsed analysis of refraining (or "forebearing").[75] In quasi-English, *it is a case of Refraining by our agent that p* if and only if our agent brings it about that she does not bring it about that p. The importance of this in agency theory is based on the assumption that refraining from doing something is distinct from simply not doing something. In the current agential framework, the importance of the above is reflected in the denial of this claim:

$$*: \quad \sim BAp \rightarrow RFp.$$

No agent brings about logical truths, but neither does an agent bring it about by what she does do that she doesn't bring about such truths. It has nothing to do with what she does. That $*$ can't hold is easily proven given any consistent system with BA-RE and BA-NO.[76] So refraining from p is not equivalent to merely not bringing about p. Whether or not it is of great importance in deontic logic itself is a more controversial matter. It would hinge on matters like whether or not there is a difference between being obligated to not bring it about that p and being obligated to bring it about that you don't bring it about that p. For example, if it is true that the only things it can be obligatory for me to not bring about are things I can only not bring about by what I do bring about instead, then it would seem that I am obligated to not bring about p iff I am obligated to bring it about by what I do do that I do not bring it about that p. In this case, I would be obligated to not bring p about iff I am also obligated to bring it about that I don't bring it about that p. An alternative account sometimes given of refraining is that of inaction coupled with ability: to refrain from bringing it about that p is to be *able* to bring it about that p and to not bring it about that p ([von Wright, 1963], on "forbearance"). This might be expressed as follows:
$$RFp =_{df} \quad \sim BAp \,\&\, ABp,$$
where "AB" is interpreted as an agential *ability* operator, perhaps a compound operator of the broad form "$\Diamond BAp$", with "$\Diamond$" suitably constrained (e.g. as what is now still possible or still possible relative to our agent). In some frameworks, the two proposed analysans of RF are provably equivalent (e.g. [Horty, 2001]).[77] Informally one might argue that if I am able to bring it about that p and don't, then I don't bring it about that p by whatever it is that I do bring about, and so I refrain per the first analysis; and if I truly bring it about by what I do that I don't bring it about that p, then I must have been able to bring it about that p even though I didn't, so I refrain per the second analysis.

---

[75]It has been most utilized by Belnap and coworkers. See [Belnap, 2001], and its references to prior papers.

It is beyond the scope of this essay to delve non-superficially into the logic of agency,[78] and here we can only barely touch on the more complex interaction of such logics with deontic logics by keeping the agency component exceedingly simple. Appendix A4 contains a brief sketch of a less abstract and more detailed influential framework for agency, STIT theory.

The Meinong–Chisholm Reduction for Agential Obligations [Chisholm, 1964] [79]

Let us set aside the jurisdictional problem as having established the need to go beyond SDL in order to represent agential obligations. Returning to deontic matters, the question arises: how do we represent not just agency, but agential obligation? With an agency operator in hand, we might now invoke the famous "Meinong–Chisholm Reduction": the idea that Jane Doe's obligation *to do* some thing is equivalent to what it is obligatory that Jane do (cf. what Jane ought to do is what it ought to be that Jane does). If we regiment this a bit using our operator for agency, we get the following version of the "reduction":

> *Meinong–Chisholm Reduction*: Jane Doe is obligated to bring it about
> that p iff it is obligatory that Jane Doe brings it about that p.

This is sometimes taken to be a reduction of personal obligation to impersonal obligation and agency (or it is sometimes rephrased as a reduction of the personal "ought to do" to the impersonal "ought to be" and agency).[80] Although not uncontested (e.g. see [Horty, 2001]), by relying on this analysis we appear to have a way to represent the troublesome sentences, (1) and (2) of the jurisdictional problem:

1″. **OB** $\sim BAp$,

2″. **OB** $\sim BA \sim p$.

These might be taken to assert that Jane Doe is positively obligated to *not* bring it about that p and that she is also positively obligated to *not* bring it about that $\sim p$. Here we can properly express the fact that she is positively obligated to be non-agential with respect to the status of both p and $\sim p$. These are easily distinguished from the claims that Jane Doe is *not* obligated to bring about p (i.e.

---

[76]Suppose S is any consistent system with BA-NO, BA-RE and PC: For *reductio* assume $\vdash \sim BAp \rightarrow RFp$. By BA-NO,$\vdash \sim BA\top$. So by our assumption, $\vdash RF\top$. Now by PC, $\vdash \sim BA\top \leftrightarrow \top$. So by BA-RE, $\vdash \sim BA \sim BA\top$. So by definition of RF, we have $\vdash \sim RF\top$, and hence an inconsistent set of theorems.

[77]But not so for the "achievement" agency operator in [Belnap, 2001].

[78]See the following sources, and the references therein: [Segerberg, 1982; Elgesem, 1993; 1997; Hilpinen, 1997a; 1997b; Segerberg 1997; Belnap, 2001].

[79]Meinong 1972 [1917]; Chisholm [1964] attributes the idea's endorsement to Nicolai Hartmann as well.

[80]More generally, it can be seen as a reduction of an agential deontic operator to a non-agential deontic operator (but not necessarily an impersonal one) and a non-deontic agency operator ([Krogh and Herrestad, 1996] and [McNamara, 2004a]).

$\sim$**OB**BAp) and that she is *not* obligated to bring about $\sim p$ (i.e. $\sim$**OB**$BA \sim p$). Similar remarks hold for our earlier equivalence **IM**$p \leftrightarrow$ **OB** $\sim p$.

Generally, if we substitute "BAp" for p in the traditional definitional scheme's equivalences, we get:

$$\textbf{IM}BAp \leftrightarrow \textbf{OB} \sim BAp$$
$$\textbf{PE}BAp \leftrightarrow \sim\textbf{OB} \sim BAp$$
$$\textbf{GR}BAp \leftrightarrow \sim\textbf{OB}BAp$$
$$\textbf{OP}BAp \leftrightarrow \sim\textbf{OB}BAp \,\&\, \sim\textbf{OB} \sim BAp.$$

If we now read each deontic operator as "it is ____ for Jane Doe that", so that it is impersonal but not agential,[81] the earlier problem with **IM**$p \leftrightarrow$ **OB** $\sim p$, coupled with trying to read the agency into the deontic operators, disappears. For the deontic-agential compound above gets things right: it is impermissible that Jane Doe brings it about that your child is disciplined iff it is obligatory that she does not bring it about that your child is disciplined. We can now clearly and distinctly express the idea that something is simply out of Jane Doe's jurisdiction.

This general approach to obligations to do things has been very widely employed in deontic logic.[82] Recently, [Krogh and Herrestad, 1996] and [McNamara, 2004a] reinterpret the analysis so that the deontic operator is *personal*, yet not agential. This is arguably a more plausible way to preserve a componential analysis of agential obligation. [McNamara, 2004a] also makes the case that a person's being obligated to be such that a certain condition holds (e.g. being obligated to be at home at noon, as promised) is the more basic idiom, and being obligated to bring about something is just being obligated to be such that you do bring it about.

Appendix A4 contains a brief discussion of the Meinong–Chisholm Analysis in the context of STIT theory.

A Glimpse at the Theory of Normative Positions [Kanger, 1971; 1957]

One way in which the Meinong–Chisholm analysis has been fruitfully employed is in the study of what are called "normative positions". A set of normative positions is intended to describe the set of all possible mutually exclusive and

---

[81]McNamara [2004a].

[82]For example, see [Kanger, 1971 [1957]; Lindahl, 1977; Porn, 1970; 1977; 1989; Horty, 1996; Jones and Sergot, 1996; Santos and Carmo, 1996; Belnap, 2001]. As indicated earlier, [Horty, 1996] and [Horty, 2001] is of interest for (among other things) its argument against the Meinong–Chisholm reduction (see Appendix), and for providing an alternative non-componential analysis of agential obligation in the context of a branching-time analysis of agency. McNamara [2004b] provides a critical exposition of the basic framework. This is in contrast to the branching-time approach to deontic contexts in [Belnap and Bartha, 2001], where agential obligation is a componential compound of an agency operator and an obligation operator (one in turn analyzed via an Andersonian–Kangerian reduction). Another alternative to the major trend above, one that would unfortunately also take us too far afield, is the adaptation of modal logics for representing computer programs (e.g. dynamic logic) to represent actions in deontic logic. A classic source here is [Meyer, 1988] which combines a dynamic logic approach to action with an adaptation of the Andersonian–Kangerian reduction to generate deontic notions. See also [Segerberg, 1982].

jointly exhaustive positions that a person or set of persons may be in regarding a proposition and with respect to a set of selected primitive normative statuses and a set of agency operators. For a given proposition, p, recall the partition regarding how Jane Doe may be positioned agentially with respect to p:

$$(BAp \vee ROp \vee PVp)\,\&\, \sim(BAp\,\&\,ROp)\,\&\, \sim(BAp\,\&\,PVp)\,\&\, \sim(ROp\,\&\,PVp).$$

Now also recall our partition with respect to obligations:

$$(\mathbf{OB}p \vee \mathbf{IM}p \vee \mathbf{OP}p)\,\&\, \sim(\mathbf{OB}p\,\&\,\mathbf{IM}p)\,\&\, \sim(\mathbf{OB}p\,\&\,\mathbf{OP}p)\,\&\, \sim(\mathbf{IM}p\,\&\,\mathbf{OP}p).$$

We might consider "merging" these two partitions, as it were, and try to get a representation of the possible ways Jane Doe may be positioned normatively with respect to her agency regarding p. Given certain choices of logic for BA and for **OB**, we might get a set of mutually exclusive and exhaustive "normative positions" for Jane Doe regarding p, the basic normative status, **OB**, and the basic agency operator, BA, such as that pictured below:
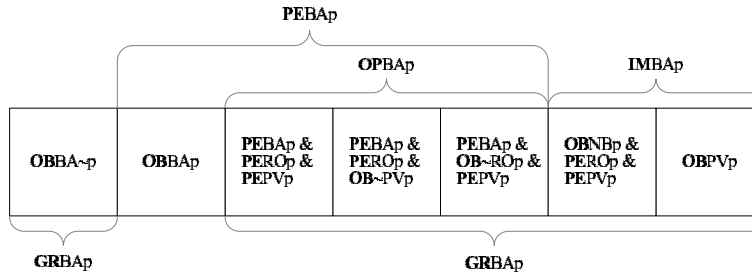
As usual, the partition above is intended to assert that the following seven classes are mutually exclusive and jointly exhaustive:

$\mathbf{OB}BA \sim p$

$\mathbf{OB}BAp$

$\mathbf{PE}BAp\,\&\,\mathbf{PE}ROp\,\&\,\mathbf{PE}PVp$ $(\sim\mathbf{OB} \sim BAp\,\&\, \sim\mathbf{OB} \sim BA \sim p\,\&\, \sim\mathbf{OB} \sim PVp)$

$\mathbf{PE}BAp\,\&\,\mathbf{PE}ROp\,\&\,\mathbf{OB} \sim PVp$ $(\sim\mathbf{OB} \sim BAp\,\&\, \sim\mathbf{OB} \sim BA \sim p\,\&\,\mathbf{OB} \sim PVp)$

$\mathbf{PE}BAp\,\&\,\mathbf{OB} \sim ROp\,\&\,\mathbf{PE}PVp$ $(\sim\mathbf{OB} \sim BAp\,\&\,\mathbf{OB} \sim BA \sim p\,\&\, \sim\mathbf{OB} \sim PVp)$

$\mathbf{OB}NBp\,\&\,\mathbf{PE}ROp\,\&\,\mathbf{PE}PVp$ $(\mathbf{OB} \sim BAp\,\&\, \sim\mathbf{OB} \sim BA \sim p\,\&\, \sim\mathbf{OB} \sim PVp)$

$\mathbf{OB}PVp$ (i.e. $\mathbf{OB}NBp\,\&\,\mathbf{OB}NRp$, given **OB**-$C$ & **OB**-M)

The respective cases where it is permissible, impermissible, optional or gratuitous to bring about p are indicated as well.

Some Literature on the Theory of Normative Positions: The theory of normative positions has been a dynamic area of research that we have barely touched on here. It has been an important and active area since its inception in Stig Kanger's seminal work [Kanger, 1971; [1957]; 1972], developed in a book-length study in [Lindahl, 1977], and thus sometimes referred to as "the Kanger–Lindahl theory". It has been used in attempts to analyze legal relations, like those made famous by [Hohfeld, 1919], among other things. The Kanger–Lindahl theory has been further developed by [Jones and Sergot, 1993; Sergot, 1999; Herrestad and Krogh, 1995] and [Lindahl, 2001]. See also [Allen, 1996] for a somewhat different approach to Hohfeldian legal relations, and Porn [1970; 1977] for a framework employed to analyze various normatively laden social positions and relations. Lindahl [2001] provides an excellent overview and orientation on Kanger's work in this area, and various problems informing subsequent research. (Other stunning contributions of Kanger to deontic logic are discussed in [Hilpinen, 2001b] in the same volume.) Sergot [1999] takes the formal work of normative positions to a new level of abstraction and precision, and the later work mentioned above by Lindahl, and Herrestad and Krogh continue the exploration of refinements of the earlier Kanger–Lindahl conceptual framework to adapt it better to the analysis of legal notions.

Deontic Compliments: One current issue in dispute is whether or not deontic operators call for agential complements or not. We outline the issue loosely here. Consider:

> *Libertarian Deontic Compliment Thesis (LDCT)*: Any of the fundamental five deontic operators followed by *any* sentential compliment is well-formed.

Let an *LDCT system* be any classical sentential modal logic containing any of the above deontic operators (but at least **OB**) that satisfies LDCT. In contrast, consider the

> *Strict Deontic Compliment Thesis (SDCT)*: Each fundamental deontic status must be followed immediately by an operator ascribing agency to an agent (here, by "BA") to be well-formed.

A *strict omission* is now a wff of the form RFp (i.e. $BA \sim BAp$). "$\sim BAp$" is just a *non-action*. *Strict deontic omissions* are deontic operators immediately followed by strict omissions.

Recall that if we substitute "BAp" for "p" in the equivalences associated with the Traditional Definitional Scheme, we get:

$$\mathbf{IM}BAp \leftrightarrow \mathbf{OB} \sim BAp$$
$$\mathbf{PE}BAp \leftrightarrow \sim\mathbf{OB} \sim BAp$$
$$\mathbf{GR}BAp \leftrightarrow \sim\mathbf{OB}BAp$$
$$\mathbf{OP}BAp \leftrightarrow \sim\mathbf{OB}BAp \,\&\, \sim\mathbf{OB} \sim BAp.$$

The instances above are all consistent with LDCT, but not SDCT. Essentially, non-action statements would have to be replaced by strict omissions. The needed replacements are given below with underlining stressing the trouble spots from the perspective of SDCT:

$\mathbf{IM}BAp \leftrightarrow \mathbf{OB}\underline{\sim BAp}$	$\mathbf{IM}BAp \leftrightarrow \mathbf{OB}\underline{BA \sim BAp}$
$\mathbf{PE}BAp \leftrightarrow \sim\mathbf{OB}\underline{\sim BAp}$	$\mathbf{PE}BAp \leftrightarrow \sim\mathbf{OB}\underline{BA \sim BAp}$
$\mathbf{GR}BAp \leftrightarrow \sim\mathbf{OB}\overline{BAp}$	(original is fine per SDCT)
$\mathbf{OP}BAp \leftrightarrow \sim\mathbf{OB}BAp \,\&\, \sim\mathbf{OB}\underline{\sim BAp}$	$\mathbf{OP}BAp \leftrightarrow \sim\mathbf{OB}BAp \,\&\, \sim\mathbf{OB}\underline{BA \sim BAp}.$

Belnap [2001] provisionally endorses SDCT. McNamara [2004a] raises doubts about SDCT. He notes that we are sometimes obligated to be a certain way (e.g. to be in our office), and furthermore, it is plausible to think that agential obligations reduce to this form — to obligations to be the agents of states of affairs, so that obligations to be a certain way are analytically prior to agential obligations.

An Obligation Fulfillment Dilemma [McNamara, 2004a][83]

Obligations can be fulfilled and violated. These are among the most characteristic features of obligations. It is often thought that fulfillment and violation conditions for what is obligatory are easily represented in SDL as follows:

$\mathbf{OB}p \,\&\, p$ (fulfillment)
$\mathbf{OB}p \,\&\, \sim p$ (violation).

Call this the "Standard Analysis". Now consider cases where p is itself some agential sentence, say BAq, where we continue to read this as saying that Jane Doe brings it about that q. The Standard Analysis then implies:

$\mathbf{OB}BAq \,\&\, BAq$ (fulfillment?)
$\mathbf{OB}BAq \,\&\, \sim BAq$ (violation?).

---

[83]This puzzle/dilemma is made explicit as such in [McNamara, 2004a], and one solution is there explored. However, the issue derives from [Krogh and Herrestad, 1996], who note that obligations can be yours yet fulfilled by someone else, and they use this distinction to offer a solution to the Leakage Problem below.

These suggest that Doe's obligation to bring it about that q is fulfilled iff she brings it about that q and is violated iff she doesn't. But if this is the proper analysis of obligation fulfillment, then it is hard to see how someone else could ever fulfill our obligations when we don't fulfill ours, for then our obligation would be unfulfilled and violated according to the Standard Analysis. Yet surely people can fulfill other people's obligations, and when they do so, it certainly seems to follow that our obligation is fulfilled. So the question then becomes, just what is obligatory? It would seem that it can't be that what is obligatory is that *Jane Doe* brings it about that p, for it is incoherent to say that someone else does that unless we mean that someone else gets Jane Doe to bring it about that p; but that is hardly the usual way in which we fulfill other's obligations. I might bring your book back to the library for you, thereby fulfilling one of your obligations without getting you to return the book yourself, at gunpoint say. So we face a dilemma:

> Since others can sometimes discharge our obligations, either *our* obligations are not always obligations for *us* to do things, and thus personal obligations need not be agential or obligation fulfillment is more complex than has been previously realized, and perhaps both [McNamara, 2004a].

The Leakage Problem [Krogh and Herrestad, 1996] [84] This is a problem closely related to the preceding one. As noted previously, when discussing two or more agents, subscripts are usually introduced to identify and distinguish the agents, for example $BA_i p \,\&\, BA_j q$ would indicate that i brings it about that p and j brings it about that q. Now let's assume that one agent can sometimes bring it about by what she does that another agent brings some thing about. For example, let's suppose that a parent can sometimes bring it about that a child brings it about that the child's room is cleaned (however rare this may in fact be). Carmo notes the following problem for the Meinong–Chisholm analysis. Consider:

1. $BA_i BA_j p \rightarrow BA_j p$

2. $\mathbf{OB} BA_i BA_j p \rightarrow \mathbf{OB} BA_j p$

(1) follows from $BA - T$, the virtually universally endorsed "success" condition for the intended agency operator. (1) is a logical truth. But then, in any system containing $\mathbf{OB}$-RM, (2) will be derivable from (1), and so if (1) is a theorem in that system, (2) will be as well. But given the Meinong–Chisholm analysis, this will imply that if I am obligated to bring it about that someone else does some thing, then she is obligated to do that thing as well. However, this is surely

---

[84]Krogh and Herrestad [1996] attributes the identification of this problem to Jose Carmo. They offer a solution there by distinguishing between personal and agential obligations.

false. If I am obligated to get my very young child to feed herself, it does not follow that she is herself, at her young age, obligated to feed herself, even if she is just becoming capable of doing so.

So it appears that the natural augmentation of SDL with an agency operator allows my obligation to implausibly "leak" beyond its proper domain and generate an obligation for her.

4.8   Challenges regarding Obligation, Change and Time

Although we have seen that obligations can be obligations to be (i.e. to satisfy a condition) as well as obligations to do, and that the former may be a special case of the latter, nonetheless, it is plausible to think that one is obligated *to do* something only if that thing is in the future. Thus even if attempts to solve Chisholm's contrary to duty paradox by invoking time do not look very plausible, this does not mean that there is no interesting work needed to forge relationships between time and obligations. For example, consider the system Kd. If we read d atemporally as all obligations past, present, and future are met, then the only relevant worlds are those so ideal that in them there has never been a single violation of a mandatory norm. But as a parent, I may be obligated to lock the front door at night even though this would not be a norm unless there had been past violations of other norms (e.g. against theft and murder). People also acquire obligations over time, create them for themselves and for others by their actions, discharge them, etc.[85]

<div align="center">CONCLUSION</div>

Plainly, there are a number of outstanding problems for deontic logic. Some see this as a serious defect; others see it merely as a serious challenge, even an attractive one. There is some antecedent reason to expect that the challenges will be great in this area. Normativity is challenging generally, not just in deontic logic. Normative notions appear to have strong semantic and pragmatic features. Normative notions must combine with notions for agency and with temporal notions to be of maximal interest — which introduces considerable logical complexity. There is also reason

---

[85]Two classics on time and deontic logic are [Thomason, 1981b] and [Thomason, 1981a], where temporal and deontic interactions are discussed, including an often invoked distinction between *deliberative 'ought's* (future-oriented/decision-oriented 'ought's) versus judgmental 'ought's (past, present or future oriented 'ought's from a purely evaluative, rather than action-oriented perspective). (Cf. the notion of "cues" for action in [van Eck, 1982].) Some other important earlier entries are [Loewer and Belzer, 1983; van Eck, 1982; Åqvist and Hoepelman, 1981], and [Chellas, 1969]. For a sample of some recent work, see [Bailhache, 1998] and her references to her earlier work and that of others, as well as [Brown, 1996a] for an attempt to develop a diachronic logic of obligations, representing obligations coming to be, and being discharged over time, where, for example, someone can now have an obligation to bring about p only if p is (now) false.

to think that there are hidden complexities in the interaction of normative notions and conditionals. Finally, there appears to be a wide array of normative notions with interesting interactions, some easily conflated with others (by ethicists as much as deontic logicians). Clearly, there is a lot of work to be done.

BIBLIOGRAPHY

[Alchourron, 1993] C. E. Alchourron. Philosophical Foundations of Deontic Logic and the Logic of Defeasible Conditionals. In *Deontic Logic in Computer Science*. J.-J. Ch Meyer, and R. J. Wieringa, eds., pp. 43–84. Chichester, John Wiley and Sons, 1993.
[Alchourron, 1996] C. E. Alchourron. Detachment and Defeasibility in Deontic Logic. *Studia Logica*, **57**: 5–18, 1996.
[Alchourron and Bulygin, 1971] C. E. Alchourron and E. Bulygin. *Normative Systems*. Wien; New York, Springer-Verlag, 1971.
[Alchourron and Bulygin, 1981] C. E. Alchourron and E. Bulygin. The Expressive Conception of Norms. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 95–124. Dordrecht, Reidel, 1981.
[Alchourron and Makinson, 1981] C. E. Alchourron and D. Makinson. Hierarchies of Regulations and Their Logic. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 125–148. Dordrecht, Reidel, 1981.
[al-Hibri, 1978] A. al-Hibri. *Deontic Logic: A Comprehensive Appraisal and a New Proposal*. Washington, DC, University Press of America, 1978.
[Allen, 1996] L. E. Allen. From the Fundamental Legal Conceptions of Hohfeld to Legal Relations: Refining the Enrichment of Solely Deontic Legal Relations. In *Deontic Logic, Agency and Normative Systems*, M. A. Brown and J. Carmo, eds., pp. 1–26. New York, Springer-Verlag, 1996.
[Anderson, 1956] A. R. Anderson. The Formal Analysis of Normative Systems. In *The Logic of Decision and Action*, N. Rescher, ed., pp. 147–213. Pittsburgh, University of Pittsburgh Press, 1956.
[Anderson, 1958] A. R. Anderson. A Reduction of Deontic Logic to Alethic Modal Logic. *Mind* **67**: 100–103, 1958.
[Anderson, 1967] A. R. Anderson. Some Nasty Problems in the Formal Logic of Ethics. *Nous*, **1**: 345–360, 1967 [1956].
[Åqvist, 1967] L. Åqvist. Good Samaritans, Contrary-to-Duty Imperatives, and Epistemic Obligations. *Nous*, **1**: 361–379, 1967.
[Åqvist, 2002] L. Åqvist. Deontic Logic. In *Handbook of Philosophical Logic, 2nd Edition*, D. Gabbay and F. Guenthner, eds. Dordrecht, Kluwer Academic Publishers, 2002 [1984].
[Åqvist and Hoepelman, 1981] L. Åqvist and J Hoepelman. Some Theorems About a 'Tree' System of Deontic Tense Logic. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 187–221. Dordrecht, Reidel, 1981.
[Asher and Bonevac, 1996] N. Asher and D. Bonevac. 'Prima Facie' Obligation. *Studia Logica*, **57**: 19–45, 1996.
[Bailhache, 1998] P. Bailhache. How to Mix Alethic, Deontic, Temporal, Individual Modalities. *Logica Trianguli*, **2**: 3–16, 1998.
[Bartha, 1999] P. Bartha. Moral Preference, Contrary-to-Duty Obligation and Defeasible Oughts. In *Norms, Logics and Information Systems*, P. McNamara and H. Prakken, eds., pp. 93–108. Amsterdam, IOS Press, 1999.
[Belnap et al., 2001] N. Belnap, M. Perloff and M. Xu. *Facing the Future: Agents and Choices in Our Indeterminist World*. New York, Oxford University Press, 2001.
[Bonevac, 1998] D. Bonevac. Against Conditional Obligation. *Nous*, **32**: 37–53, 1998.
[Brewka, 1989] G. Brewka. Non-Monotonic Logics — A Brief Overview. *AI Communications: The European Journal of Artificial Intelligence*, **2**: 88–97, 1989.
[Brown, 1996a] M. A. Brown. Doing as We Ought: Towards a Logic of Simply Dischargeable Obligations. In *Deontic Logic, Agency and Normative Systems*, M. A. Brown and J. Jose Carmo, eds., pp. 47–65. New York, Springer-Verlag, 1996.

[Brown, 1996b] M. A. Brown. A Logic of Comparative Obligation. *Studia Logica*, **57**: 117–137, 1996.

[Carmo and Jones, 2002] J. Carmo and A. I. J. Jones. Deontic Logic and Contrary-to-Duties. In *Handbook of Philosophical Logic*. D. Gabbay and F. Guenthner, eds., pp. 265–343. Kluwer Academic Publishers, 2002.

[Castañeda, 1981] H. N. Castañeda. The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in One Fell Swoop. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 37–85. Dordrecht, Reidel, 1981.

[Chellas, 1969] B. F. Chellas. The Logical Form of Imperatives. Philosophy Department. Stanford University. Stanford: Perry Lane Press, 1969.

[Chellas, 1974] B. F. Chellas. Conditional Obligation. In *Logical Theory and Semantic Analysis*, S. Stenlund, ed., pp. 23–33. Dordrecht, D. Reidel, 1974.

[Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge, Cambridge University Press, 1980.

[Chisholm, 1963a] R. M. Chisholm. Contrary-to-Duty Imperatives and Deontic Logic. *Analysis*, **24**: 33–36, 1963.

[Chisholm, 1963b] R. M. Chisholm. Supererogation and Offence: A Conceptual Scheme for Ethics. *Ratio*, **5**: 1–14, 1963.

[Chisholm, 1964] R. M. Chisholm. The Ethics of Requirement. *American Philosophical Quarterly*, **1**: 147–153, 1964.

[Chisholm and Sosa, 1966] R. M. Chisholm and E. Sosa. Intrinsic Preferability and the Problem of Supererogation. *Synthese*, **16**: 321–331, 1966.

[Da Costa and Carnielli, 1986] N. C. A. Da Costa and W. A. Carnielli. On Paraconsistent Deontic Logic. *Philosophia*, **16**: 293–305, 1986.

[Danielsson, 1968] S. Danielsson. *Preference and Obligation: Studies in the Logic of Ethics*. Uppsala, Filosofiska föreningen. 1968.

[DeCew, 1981] J. W. DeCew. Conditional Obligation and Counterfactuals. *Journal of Philosophical Logic*, **10**: 55–72, 1981.

[Elgesem, 1993] D. Elgesem. Action Theory and Modal Logic. Department of Philosophy. University of Oslo. Oslo, 1993.

[Elgesem, 1997] D. Elgesem. The Modal Logic of Agency. *Nordic Journal of Philosophical Logic*, **2**: 1–46, 1997.

[Fehige, 1994] C. Fehige. The Limit Assumption. In *Deontic (and Prohairetic) Logic in Analyomen 1*, G. Meggle, ed. Hawthorne, de Gruyter, 1994.

[Feldman, 1986] F. Feldman. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht; Boston, D. Reidel, 1986.

[Føllesdal and Hilpinen, 1971] D. Føllesdal and R. Hilpinen. Deontic Logic: An Introduction. In *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen, ed. Dordrecht, D. Reidel, 1971.

[Forrester, 1984] J. W. Forrester. Gentle Murder, or the Adverbial Samaritan. *Journal of Philosophy*, **81**: 193–196, 1984.

[Forrester, 1975] M. Forrester. Some Remarks on Obligation, Permission, and Supererogation. *Ethics*, **85**: 219–226, 1975.

[Goble, 1990a] L. Goble. A Logic of Good, Should, and Would: Part I. *Journal of Philosophical Logic*, **19**: 169–199, 1990.

[Goble, 1990b] L. Goble. A Logic of Good, Should, and Would: Part II." *Journal of Philosophical Logic* **19**: 253–276, 1990.

[Goble, 1991] L. Goble. Murder Most Gentle: The Paradox Deepens. *Philosophical Studies* **64**: 217–227, 1991.

[Goble, forthcoming-a] L. Goble. A Logic for Deontic Dilemmas. *Journal of Applied Logic*, forthcoming.

[Goble, forthcoming-b] L. Goble. Preference Semantics for Deontic Logic Part I — Simple Models. *Logique et Analyse*, **46**: 383–418, forthcoming.

[Goldman, 1976] H. S. Goldman. Dated Rightness and Moral Perfection. *The Philosophical Review*, **85**: 449–487, 1976.

[Gowans, 1987] C. Gowans. *Moral Dilemmas*. Oxford, Oxford University Press, 1987.

[Greenspan, 1975] P. S. Greenspan. Conditional Oughts and Hypothetical Imperatives. *Journal of Philosophy*, **72**: 259–276, 1975.

[Hansen, 2004] J. Hansen. Problems and Results for Logics About Imperatives. *Journal of Applied Logic*, **2**: 39–61, 2004.

[Hansson, 1969] B. Hansson. An Analysis of Some Deontic Logics. *Nous*, **3**: 373–398, 1969.

[Hansson, 1990] S. O. Hansson. Preference-Based Deontic Logic (PDL). *Journal of Philosophical Logic*, **19**: 75–93, 1990.

[Hansson, 2001] S. O. Hansson. *Structures of Values and Norms*. Cambridge, Cambridge University Press, 2001.

[Hedenius, 1963] I. Hedenius. *Om Ratt och Moral ('On Law and Morals')*. Stockholm, 1963 [1941].

[Herrestad and Krogh, 1995] H. Herrestad and C. Krogh. Obligations Directed from Bearers to Counterparts. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (International Conference on Artificial Intelligence and Law)*, pp. 210–218. College Park, Maryland, USA, ACM Press, 1995.

[Hilpinen, 1997a] R. Hilpinen. On Action and Agency. In *Logic, Action and Cognition — Essays in Philosophical Logic*, E. Ejerhed and S. Lindstrom, eds., pp. 3–27. Dordrecht, Kluwer Academic Publishers, 1997.

[Hilpinen, 1997b] R. Hilpinen. On States, Actions, Omissions and Norms. *Contemporary Action Theory (Individual Action)*, G. Holmstrom-Hintikka and R. Tuomela, eds., pp. 83–107. Dordrecht, Kluwer Academic Publishers, 1997.

[Hilpinen, 2001a] R. Hilpinen. Deontic Logic. In *The Blackwell Guide to Philosophical Logic*, L. Goble, ed., pp. 159–182. Malden, Blackwell Publishers, 2001.

[Hilpinen, 2001b] R. Hilpinen. Stig Kanger on Deontic Logic. In *Collected Papers of Stig Kanger with Essays on His Life and Work, Vol. II*, G. Holmstrom-Hintikka, S. Lindstrom and R. Sliwinski, eds., pp. 131–149. Dordrecht, Kluwer Academic Publishers, 2001.

[Hintikka, 1957] J. Hintikka. Quantifiers in Deontic Logic. *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum*, **23**: 1–23, 1957.

[Hintikka, 1971] J. Hintikka. Some Main Problems of Deontic Logic. In *Deontic Logic: Introductory and Systematic Readings*. R. Hilpinen, ed., pp. 59–104. Dordrecht, D. Reidel, 1971.

[Hohfeld, 1919] W. N. Hohfeld. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. New Haven, Yale University Press. 1919. (Edited by Walter Wheeler Cook.)

[Horty, 1994] J. F. Horty. Moral Dilemmas and Nonmonotonic Logic. *Journal of Philosophical Logic*, **23**: 35–65, 1994.

[Horty, 1996] J. F. Horty. Agency and Obligation. *Synthese*, **108**: 269–307, 1996.

[Horty, 2001] J. F. Horty. *Agency and Deontic Logic*. Oxford England; New York, Oxford University Press, 2001.

[Horty, 2003] J. F. Horty. Reasoning with Moral Conflicts. *Nous*, **37**: 557–605, 2003.

[Hrushka and Joerden, 1987] J. Hrushka and J. C. Joerden. Supererogation: Vom Deontologischen Sechseck Zum Deontologischen Zehneck. *Archiv fu Rechts und Sozialphilosophie*, **73**: 93–123, 1987.

[Humberstone, 1974] I. L. Humberstone. Logic for Saints and Heroes. *Ratio*, **16**: 103–114, 1974.

[Jackson, 1985] F. Jackson. On the Semantics and Logic of Obligation. *Mind*, **94**: 177–195, 1985.

[Jackson, 1988] F. Jackson. Understanding the Logic of Obligation. *Aristotelian Society*, **SUPP 62**: 255–270, 1988.

[Jackson and Pargetter, 1986] F. Jackson and R. Pargetter. Oughts, Options, and Actualism. *Philosophical Review*, **95**: 233–255, 1986.

[Jones, 1990] A. I. J. Jones. Deontic Logic and Legal Knowledge Representation. *Ratio Juris*, **3**: 237–244, 1990.

[Jones and Porn, 1985] A. I. J. Jones and I. Porn. Ideality, Sub-Ideality and Deontic Logic. *Synthese*, **65**: 275–290, 1985.

[Jones and Porn, 1986] A. I. J. Jones and I. Porn. 'Ought' and 'Must'. *Synthese*, **66**: 89–93, 1986.

[Jones and Sergot, 1993] A. I. J. Jones and M. Sergot. On the Characterization of Law and Computer Systems. In *Deontic Logic in Computer Science*, J.-J. Ch. Meyer and R. J. Wieringa, eds., pp. 275–307. Chichester, John Wiley and Sons, 1993.

[Jones and Sergot, 1996] A. I. J. Jones and M. Sergot. A Formal Characterization of Instituionalized Power. *Logic Journal of the IGPL* **4**: 429–445, 1996.

[Jorgensen, 1937] J. Jorgensen. Imperatives and Logic. *Erkenntnis*, **7**: 288–296, 1937.

[Kamp, 1974] H. Kamp. Free Choice Permission. *Proceedings of the Aristotelian Society*, **74**: 57–74, 1974.

[Kamp, 1979] H. Kamp. Semantics Versus Pragmatics. In *Formal Semantics and Pragmatics for Natural Languages*, F. Guenthner and S. J. Schmidt, eds., pp. 255–287. Dordrecht, D. Reidel, 1979.

[Kanger, 1971] S. Kanger. New Foundations for Ethical Theory. In *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen, ed., pp. 36–58. Dordrecht, D. Reidel, 1971 [1957].

[Kanger, 1972] S. Kanger. Law and Logic. *Theoria*, **38**: 105–132, 1972.

[Kempson, 1977] R. M. Kempson. *Semantic Theory*. Cambridge, Cambridge University Press, 1977.

[Knuuttila, 1981] S. Knuuttila. The Emergence of Deontic Logic in the Fourteenth Century. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 225–248. Dordrecht, Reidel, 1981.

[Krogh and Herrestad, 1996] C. Krogh and H. Herrestad. Getting Personal: Some Notes on the Relationship between Personal and Impersonal Obligation. In *Deontic Logic, Agency and Normative Systems*, M. A. Brown and J. Carmo, eds., pp.134–153. New York, Springer-Verlag, 1996.

[Lemmon, 1957] E. J. Lemmon. New Foundations for Lewis' Modal Systems. *The Journal of Symbolic Logic*, **22**: 176–186, 1957.

[Lemmon, 1962a] E. J. Lemmon. On Sentences Verifiable by Their Use. *Analysis*, **22**: 86–89, 1962.

[Lemmon, 1962b] E. J. Lemmon. Moral Dilemmas. *Philosophical Review*, **71**: 139–158, 1962.

[Lemmon and Scott, 1977] E. J. Lemmon and D. Scott. *The 'Lemmon Notes': An Introduction to Modal Logic*. Oxford, Basil Blackwell, 1977.

[Lewis, 1978] D. K. Lewis. Reply to McMichael's 'Too Much of a Good Thing: A Problem in Deontic Logic'. *Analysis*, **38**: 85–86, 1978.

[Lewis, 1973] D. K. Lewis. *Counterfactuals*. Oxford, Blackwell, 1973.

[Lewis, 1974] D. K. Lewis. Semantic Analyses for Dyadic Deontic Logic. In *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday*. S. Stenlund, ed., pp. 1–14. Dordrecht, D. Reidel, 1974.

[Lindahl, 1977] L. Lindahl. *Position and Change*. Dordrecht, D. Reidel, 1977.

[Lindahl, 2001] L. Lindahl. Stig Kanger's Theory of Rights. In *Collected Papers of Stig Kanger with Essays on His Life and Work, Volume II*, G. Holmstrom-Hintikka, S. Lindstrom and R. Sliwinski, eds., pp. 151–171. Dordrecht, Kluwer Academic Publishers, 2001.

[Loewer and Belzer, 1983] B. Loewer and M. Belzer. Dyadic Deontic Detachment. *Synthese*, **54**: 295–318, 1983.

[Lokhorst, 2004] G.-J. Lokhorst. Mally's Deontic Logic. In *The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)*, E. N. Zalta, ed. Stanford University, 2004. `http://plato.stanford.edu/archives/win2004/entries/mally-deontic`

[Makinson, 1993] D. Makinson. Five Faces of Minimality. *Studia Logica*, **52**: 339–379, 1993.

[Makinson, 1999] D. Makinson. On the Fundamental Problem of Deontic Logic. In *Norms, Logic and Information Systems*, P. McNamara and H. Prakken, eds., pp. 29–53. Amsterdam, IOS Press, 1999.

[Makinson and van der Torre, 2003] D. Makinson and L. W. N. van der Torre. What Is Input/Output Logic? In *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics*, B. Löwe, W. Malzkorn and T. Räsch, eds., pp. 163–174. Berlin, Springer-Verlag, 2003.

[Mally, 1926] E. Mally. *Gurndgesetz Des Sollens. Elemente Der Logik Des Willens*. Graz, Leuschner & Leubensky, 1926.

[Mares, 1992] E. D. Mares. Andersonian Deontic Logic. *Theoria*, **58**: 3–20, 1992.

[Mares and McNamara, 1997] E. D. Mares and P. McNamara. Supererogation in Deontic Logic: Metatheory for DWE and Some Close Neighbours. *Studia Logica*, **59**: 397–415, 1997.

[McMichael, 1978] A. McMichael. Too Much of a Good Thing: A Problem in Deontic Logic. *Analysis*, **38**: 83–84, 1978.

[McNamara, 1990] P. McNamara. The Deontic Quaddecagon. Philosophy Department. University of Massachusetts, Amherst, MA, 1990.

[McNamara, 1995] P. McNamara. The Confinement Problem: How to Terminate Your Mom with Her Trust. *Analysis*, **55**: 310–313, 1995.

[McNamara, 1996a] P. McNamara. Doing Well Enough: Toward a Logic for Commonsense Morality. *Studia Logica*, **57**: 167–192, 1996.

[McNamara, 1996b] P. McNamara. Making Room for Going Beyond the Call. *Mind*, **105**: 415–450, 1996.

[McNamara, 1996c] P. McNamara. Must I Do What I Ought? (or Will the Least I Can Do Do?). In *Deontic Logic, Agency and Normative Systems*, M. A. Brown and J. Carmo, eds., pp. 154–173. New York, Springer-Verlag, 1996.

[McNamara, 1999] P. McNamara. Doing Well Enough in an Andersonian–Kangerian Framework. In *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*, P. McNamara and H. Prakken, eds., pp. 181–198. Washington, DC, IOS Press, 1999.

[McNamara, 2004a] P. McNamara. Agential Obligation as Non-Agential Personal Obligation Plus Agency. *Journal of Applied Logic*, **2**: 117–152, 2004.

[McNamara, 2004b] P. McNamara. [Review of] Agency and Deontic Logic. *Mind*, **113**: 179–185, 2004.

[McNamara, 2005] P. McNamara. Deontic Logic. In *The Stanford Encyclopedia of Philosophy (Fall 2005 Edition)*, E. N. Zalta, ed., Stanford University, 2005. `http://plato.stanford.edu/archives/fall2005/entries/logic-deontic/`

[Meinong, 1972] A. Meinong. *On Emotional Presentation*. Evanston, Northwestern University Press, 1972 [1917].

[Meyer, 1988] J.-J. Ch. Meyer. A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic. *Notre Dame Journal of Formal Logic*, **29**: 109–136, 1988.

[Morreau, 1996] M. Morreau. 'Prima Facie' and Seeming Duties. *Studia Logica*, **57**: 47–71, 1996.

[Mott, 1973] P. L. Mott. On Chisholm's Paradox. *Journal of Philosophical Logic*, **2**: 197–210, 1973.

[Nowell Smith and Lemmon, 1960] P. H. Nowell Smith and E. J. Lemmon. Escapism: The Logical Basis of Ethics. *Mind*, **69**: 289–300, 1960.

[Nute, 1997] D. Nute, ed. *Defeasible Deontic Logic*. Vp;i,e 263 pf *Synthese Library*. Dordrecht; Boston, Kluwer Academic Publishers, 1997.

[Porn, 1970] I. Porn. *The Logic of Power*. Oxford, Blackwell, 1970.

[Porn, 1977] I. Porn. *Action Theory and Social Science*. Volume 120 of *Synthese Library*, Dordrecht; Boston, Reidel, 1977.

[Porn, 1989] I. Porn. On the Nature of a Social Order. In *Logic, Methodology and Philosophy of Science*, J. E. Fenstad, T. Frolov and R. Hilpinen, eds., pp. 553–567. New York, Elsevier Science, 1989.

[Prakken, 1996] H. Prakken. Two Approaches to the Formalisation of Defeasible Deontic Reasoning. *Studia Logica*, **57**: 73–90, 1996.

[Prakken and Sergot, 1996] H. Prakken and M. Sergot. Contrary-to-Duty Obligations. *Studia Logica*, **57**: 91–115, 1996.

[Prakken and Sergot, 1997] H. Prakken and M. Sergot. Dyadic Deontic Logic and Contrary-to-Duty Obligations. In Nute [1997, pp. 223–262]

[Prior, 1954] A. N. Prior. The Paradoxes of Derived Obligation. *Mind*, **63**: 64–65, 1954.

[Prior, 1958] A. N. Prior. Escapism: The Logical Basis of Ethics. In *Essays in Moral Philosophy*, A. I. Melden, ed., pp. 135–146. Seattle, University of Washington Press, 1958.

[Prior, 1962] A. N. Prior. *Formal Logic*. Oxford, Oxford University Press, 1962 [1955].

[Quine, 1976] W. V. O. Quine. The Ways of Paradox. In *The Ways of Paradox and Other Essays*, pp. 1–18. Cambridge, Harvard University Press, 1976 [1962].

[Ross, 1941] A. Ross. Imperatives and Logic. *Theoria*, **7**: 53–71, 1941.

[Ross, 1939] W. D. Ross. *Foundations of Ethics*. Oxford, Oxford University Press, 1939.

[Sajama, 1985] S. Sajama. Supererogation and High Values. *Theoria*, **51**: 77–88, 1985.

[Santos and Carmo, 1996] F. Santos and J. Carmo. Indirect Action, Influence and Responsibility. In *Deontic Logic, Agency and Normative Systems*, M. A. Brown and J. Carmo, eds., pp. 194–215. New York, Springer-Verlag, 1996

[Schotch and Jennings, 1981] P. K. Schotch and R. E. Jennings. Non-Kripkean Deontic Logic. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 149–162. Dordrecht, Reidel, 1981.

[Schurz, 1997] G. Schurz. *The Is-Ought Problem: An Investigation in Philosophical Logic*. Dordrecht; Boston, Kluwer Academic Publishers, 1997.

[Segerberg, 1982] K. Segerberg. A Deontic Logic of Action. *Studia Logica*, **41**: 269–282, 1982.

[Segerberg, 1997] K. Segerberg. Delta Logic and Brown's Logic of Ability. In *Logic, Action and Cognition — Essays in Philosophical Logic*, E. Ejerhed and S. Lindstrom, eds., pp. 29–45. Dordrecht, Kluwer Academic Publishers, 1997.

[Sergot, 1999] M. Sergot. Normative Positions. In *Norms, Logic and Information Systems*, P. McNamara and H. Prakken, eds., pp. 289–308. Amsterdam, IOS Press, 1999.

[Sinnott-Armstrong, 1985] W. Sinnott-Armstrong. A Solution to Forrester's Paradox of Gentle Murder. *Journal of Philosophy*, **82**: 162–168, 1985.

[Smiley, 1963] T. Smiley. The Logical Basis of Ethics. *Acta Philosophica Fennica*, **16**: 237–246, 1963.

[Smith, 1993] T. Smith. Violation of Norms. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pp. 60–65. Amsterdam, The Netherlands; New York, ACM Press, 1993.

[Smith, 1994] T. Smith. Legal Expert Systems: Discussion of Theoretical Assumptions. Faculty of Law. University of Utrecht. Utrecht, 1994.

[Thomason, 1981a] R. H. Thomason. Deontic Logic and the Role of Freedom in Moral Deliberation. In *New Studies in Deontic Logic*, R. Hilpinen, ed., pp. 177–186. Dordrecht, Reidel, 1981.

[Thomason, 1981b] R. H. Thomason. Deontic Logic as Founded on Tense Logic. In *New Studies in Deontic Logic*. R. Hilpinen, ed., p. 165–176. Dordrecht, Reidel, 1981.

[Tomberlin, 1981] J. E. Tomberlin. Contrary-to-Duty Imperatives and Deontic Logic. *Nous*, **15**: 357–375, 1981.

[Urmson, 1958] J. O. Urmson. Saints and Heroes. In *Essays in Moral Philosophy*, A. I. Melden, ed., pp. 198–216. Seattle, University of Washington Press, 1958.

[van der Torre, 1997] L. van der Torre. Reasoning About Obligations: Defeasibility in Preference-Based Deontic Logic. Tinbergen Institute. Erasmus University Rotterdam. Amsterdam, 1997.

[van der Torre and Tan, 1997] L. van der Torre and Y. H. Tan. The Many Faces of Defeasibility in Defeasible Deontic Logic. In Nute [1997, pp. 79–122].

[van der Torre and Tan, 2000] L. van der Torre and Y. H. Tan. Two-Phase Deontic Logic. *Logique et Analyse*, **171–172**: 411–456, 2000.

[van Eck, 1982] J. van Eck. A System of Temporally Relative Modal and Deontic Predicate Logic and Its Philosophical Applications (2). *Logique et Analyse*, **25**: 339–381, 1982.

[van Fraasssen, 1972] B. C. van Fraassen. The Logic of Conditional Obligation. *Journal of Philosophical Logic*, **1**: 417–438, 1972.

[van Fraasssen, 1973] B. C. van Fraassen. Values and the Heart's Command. *Journal of Philosophy*, **70**: 5–19, 1973.

[von Wright, 1951] G. H. von Wright. Deontic Logic. *Mind*, **60**: 1–15, 1951.

[von Wright, 1953] G. H. von Wright. *An Essay in Modal Logic*. New York, Humanities Press, 1953.

[von Wright, 1956] G. H. von Wright. A Note on Deontic Logic and Derived Obligation. *Mind*, **65**: 507–509, 1956.

[von Wright, 1963] G. H. von Wright. *Norm and Action: A Logical Enquiry*. New York, Humanities Press, 1963.

[von Wright, 1968] G. H. von Wright. *An Essay in Deontic Logic and the General Theory of Action*. Amsterdam, North Holland, 1968.

[von Wright, 1971] G. H. von Wright. A New System of Deontic Logic. In *Danish Yearbook of Philosophy*, **1**: 173–182, 1971.

APPENDICES

A.1 Alternative Axiomatization of SDL

The following alternative axiom system, which is provably equivalent to SDL, "breaks up" SDL into a larger number of "weaker parts" (SDL *a la carte*, as it were). This has the advantage of facilitating comparisons with other systems that reject one or more of SDL's theses in response to one or more of the problems discussed above.[86]

SDL′: $A1$.   All tautologous wffs of the language   (TAUT)
         $A2'$. $\mathbf{OB}(p\,\&\,q) \rightarrow (\mathbf{OB}p\,\&\,\mathbf{OB}q)$ ($\mathbf{OB}$-M)
         $A3'$. $(\mathbf{OB}p\,\&\,\mathbf{OB}q) \rightarrow \mathbf{OB}(p\,\&\,q)$ ($\mathbf{OB}$-C)
         $A4'$. $\sim\!\mathbf{OB}\bot$ ($\mathbf{OB}$-OD)
         $A5'$. $\mathbf{OB}\top$ ($\mathbf{OB}$-N)
         $R1$.   If $\vdash p$ and $\vdash p \rightarrow q$, then $\vdash q$ (MP)
         $R2'$.   If $\vdash p \leftrightarrow q$, then $\mathbf{OB}p \leftrightarrow \mathbf{OB}q$ ($\mathbf{OB}$-RE).

We recall SDL for easy comparison:

SDL: $A1$.   All tautologous wffs of the language (TAUT)
        $A2$. $\mathbf{OB}(p \rightarrow q) \rightarrow (\mathbf{OB}p \rightarrow \mathbf{OB}q)$ ($\mathbf{OB}$-K)
        $A3$. $\mathbf{OB}p \rightarrow \sim\!\mathbf{OB}\!\sim\!p$ ($\mathbf{OB}$-D)
        MP.   If $\vdash p$ and $\vdash p \rightarrow q$ then $\vdash q$ (MP)
        $R2$.   If $\vdash p$ then $\vdash \mathbf{OB}p$ ($\mathbf{OB}$-NEC).

Below is a proof that these two system are "equipollent": any formula derivable in the one is derivable in the other.

I. First, we need to prove that *each axiom (scheme) and rule of SDL′ can be derived in SDL*. $A1$ and $R1$ are common to both systems, so we need only show that $A2'$–$A5'$ and $R2'$ are derivable.

Recall that $\mathbf{OB}$-RM, and $\mathbf{OB}$-RE (i.e. $R2'$ are derivable in SDL:

Show: If $\vdash p \rightarrow q$, then $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$. ($\mathbf{OB}$-RM)
Proof: Assume $\vdash p \rightarrow q$. By $\mathbf{OB}$-NEC, $\vdash \mathbf{OB}(p \rightarrow q)$, and then by $\mathbf{OB}$-K, $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$.
Corollary: If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB}q$ ($R2'$ or $\mathbf{OB}$-RE)

So it remains to be shown that $A2'$–$A5'$ are derivable in SDL, and to do so we make free use of our already derived rules, $\mathbf{OB}$-RM and $\mathbf{OB}$-RE.

Show: $\vdash \mathbf{OB}(p\,\&\,q) \rightarrow (\mathbf{OB}p\,\&\,\mathbf{OB}q)$ ($A2'$ or $\mathbf{OB}$-M)
Proof: By PC, $\vdash (p\,\&\,q) \rightarrow p$. So by $\mathbf{OB}$-RM $\vdash \mathbf{OB}(p\,\&\,q) \rightarrow \mathbf{OB}p$. In the same manner, we can derive $\vdash \mathbf{OB}(p\,\&\,q) \rightarrow \mathbf{OB}q$. From these two, by PC, we then get $\mathbf{OB}(p\,\&\,q) \rightarrow (\mathbf{OB}p\,\&\,\mathbf{OB}q)$.

Show: $\vdash (\mathbf{OB}p\,\&\,\mathbf{OB}q) \rightarrow \mathbf{OB}(p\,\&\,q)$ ($A3'$ or $\mathbf{OB}$-C)

---

[86]The interrelationships between the rules and axioms which constitute the equivalence between these systems is taken for granted in work on deontic logic, and is thus useful to know.

Proof: By $PC, \vdash p \rightarrow (q \rightarrow (p \,\&\, q))$. So by **OB**-*RM* $\vdash$ **OB**$p \rightarrow$ **OB**$(q \rightarrow (p \,\&\, q))$. But by **OB**-*K*, we have $\vdash$ **OB**$(q \rightarrow (p \,\&\, q)) \rightarrow ($**OB**$q \rightarrow$ **OB**$(p \,\&\, q))$. So from these two, by PC, $\vdash$ **OB**$p \rightarrow ($**OB**$q \rightarrow$ **OB**$(p \,\&\, q))$, which is equivalent by PC to $\vdash ($**OB**$p \,\&\,$ **OB**$q) \rightarrow$ **OB**$(p \,\&\, q)$.

Show: $\vdash {\sim}$**OB**$\bot$ ($A4'$ or **OB**$-$OD)
Proof: (By reductio) Assume **OB**$\bot$. Since by $PC, \vdash \bot \leftrightarrow (p \,\&\, {\sim}p)$, by **OB**-*RE*, we get $\vdash$ **OB**$\bot \leftrightarrow$ **OB**$(p \,\&\, {\sim}p)$. So from this and our assumption, we get **OB**$(p \,\&\, {\sim}p)$. Given **OB**-*M*, this yields **OB**$p \,\&\,$ **OB** ${\sim}p$, and then from $A3$ of SDL, we get **OB**$p \,\&\, {\sim}$**OB** ${\sim}p$, a contradiction. So $\vdash {\sim}$**OB**$\bot$.

Show: $\vdash$ **OB**$\top$ ($A5'$ or **OB**-*N*)
Proof: By $PC, \vdash \top$. So By **OB** $- NEC$, we have $\vdash$ **OB**$\top$.

II. It remains for us to show that *each axiom (scheme) and rule of SDL can be derived in SDL$'$*. Again, $A1$ and $R1$ are common to both systems, so we need only show that $A2$, $A3$ and $R2$ are derivable in SDL$'$. It will be useful (but not necessary) to first show that **OB**-RM is derivable in SDL$'$, and then show the remaining items.

Show: If $\vdash p \rightarrow q$, then $\vdash$ **OB**$p \rightarrow$ **OB**q. (**OB**-RM)
Proof: Assume $\vdash p \rightarrow q$. By PC, it follows that $\vdash p \leftrightarrow (p \,\&\, q)$. So by $R2'$, $\vdash$ **OB**$p \leftrightarrow$ **OB**$(p \,\&\, q)$, and so by $PC, \vdash$ **OB**$p \rightarrow$ **OB**$(p \,\&\, q)$. But by $A2'$, $\vdash$ **OB**$(p \,\&\, q) \rightarrow ($**OB**$p \,\&\,$ **OB**$q)$. So from the last two results, by $PC, \vdash$ **OB**$p \rightarrow ($**OB**$p \,\&\,$ **OB**$q)$, and thus $\vdash$ **OB**$p \rightarrow$ **OB**q.

Corollary: **OB** $- RM$ is inter-derivable with **OB**-*RE* $+$ **OB**-*M*.

This follows from the preceding proof and the earlier proof of **OB**-*M* showing that SDL contains SDL$'$.

Show: $\vdash$ **OB**$(p \rightarrow q) \rightarrow ($**OB**$p \rightarrow$ **OB**$q)$ ($A2$ or **OB**-K)
Proof: By $PC, \vdash ((p \rightarrow q) \,\&\, p) \rightarrow q$. So by **OB**-RM, $\vdash$ **OB**$((p \rightarrow q) \,\&\, p) \rightarrow$ **OB**q. But by $A2'$ conjoined with $A3'$, we get $\vdash$ **OB**$((p \rightarrow q) \,\&\, p) \leftrightarrow ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p)$. So from the last two results, by PC, we get $\vdash ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p) \rightarrow$ **OB**q, and thus $\vdash ($**OB**$(p \rightarrow q) \rightarrow ($**OB**$p \rightarrow$ **OB**$q)$.[87]

Show: $\vdash$ **OB**$p \rightarrow {\sim}$**OB** ${\sim}p$ ($A3$ or **OB**-D)
Proof: (Reductio) Assume ${\sim}($**OB**$p \rightarrow {\sim}$**OB** ${\sim}p)$. By PC, (**OB**$p \,\&\,$ **OB** ${\sim}p)$. So by $A3'$, **OB**$(p \,\&\, {\sim}p)$, which is equivalent, by $R2'$ to **OB**$\bot$, which contradicts $A4'$.

Show: If $\vdash p$ then $\vdash$ **OB**p ($R2$ or **OB**-NEC)

---

[87]A direct proof of $A2$ without first proving RM is:

Show: $\vdash$ **OB**$(p \rightarrow q) \rightarrow ($**OB**$p \rightarrow$ **OB**$q)$ ($A2$ or **OB**-K)
Proof: By $PC, \vdash ((p \rightarrow q) \,\&\, p) \leftrightarrow (p \,\&\, q)$. So by $R2', \vdash$ **OB**$((p \rightarrow q) \,\&\, p) \leftrightarrow$ **OB**$(p \,\&\, q)$. But by $A2'$ conjoined with $A3'$, we get $\vdash$ **OB**$((p \rightarrow q) \,\&\, p) \leftrightarrow ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p)$. So from the last two results, by PC, we get $\vdash ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p) \leftrightarrow$ **OB**$(p \,\&\, q)$, and thus $\vdash ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p) \rightarrow$ **OB**$(p \,\&\, q)$. But by $A2'$, we have $\vdash$ **OB**$(p \,\&\, q) \rightarrow ($**OB**$p \,\&\,$ **OB**$q)$. So from the last two results, by PC, we get $\vdash ($**OB**$(p \rightarrow q) \,\&\,$ **OB**$p) \rightarrow ($**OB**$p \,\&\,$ **OB**$q)$, and thus $\vdash ($**OB**$(p \rightarrow q) \rightarrow ($**OB**$p \rightarrow$ **OB**$q)$.

Proof: Assume $\vdash p$. From this by *PC*, it follows that $\vdash p \leftrightarrow \top$. So by $R2'$, $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB}\top$. But then from $A5'$, we get $\vdash \mathbf{OB}p$. So if $\vdash p$ then $\vdash \mathbf{OB}p$.

A.2   A Bit More on Chisholm's Paradox

Recall the quartet and its most natural symbolization in SDL:

1. It ought to be that Jones goes to assist his neighbors. $1'$) $\mathbf{OB}g$.

2. It ought to be that if Jones goes, then he tells them he is coming. $2'$) $\mathbf{OB}(g \rightarrow t)$.

3. If Jones doesn't go, then he ought not tell them he is coming. $3'$) $\sim g \rightarrow \mathbf{OB} \sim t$.

4. Jones doesn't go. $4'$) $\sim g$.

There is a general point to be made regarding the key inferences that generate the paradox per the above symbolization. There is a sense in which the inference from $(1')$ and $(2')$ to $\mathbf{OB}t$ and the inference from $(3')$ & $(4')$ to $\mathbf{OB} \sim t$ involve "detachment" of an obligation from a pair of premises, one of which involves a deontic conditional in some way. Let us introduce a bit of regimentation. Let

$$\text{``}\mathbf{OB}(q/p)\text{''}$$

represent a shorthand for a conditional obligation or ought statement like that in the natural language sentence, (3), above.[88] So we will read $\mathbf{OB}(q/p)$ as "if p, then it ought to be (or it is obligatory) that q", in the manner of (3) above. Suppose we also assume, as almost all have,[89] that monadic obligations are disguised dyadic obligations, per the following analysis:

$$\mathbf{OB}p =_{df} \mathbf{OB}(p/\top).$$

With this in mind we distinguish between two relevant types of "detachment principles"[90] that we might ascribe to these iffy-ought's:

Factual Detachment (FD): $p \,\&\, \mathbf{OB}(q/p). \rightarrow \mathbf{OB}q$
Deontic Detachment (DD): $\mathbf{OB}p \,\&\, \mathbf{OB}(q/p). \rightarrow \mathbf{OB}q$

Factual detachment tells us that from the fact that p, and the deontic conditional to the effect that if p then it ought to be that q, we can conclude that it ought to be that q. *Deontic Detachment* in contrast tells us that from the fact that it ought to be that p and that if p, then it ought to be that q, we can conclude that it ought to be that q. If we interpret a deontic conditional as a material

---

[88]We continue to ignore the differences between "obligation" and "ought" for simplicity.
[89]Alchourron [1993] is a salient exception.
[90]Greenspan [1975].

conditional with an obligatory consequent (as in (3′) above), FD, but not DD is supported in SDL. Conversely, if we interpret deontic conditionals as obligatory material conditionals (as in (2′) above), DD, but not FD is supported in SDL.[91] Although we have shown earlier that neither of these interpretations is acceptable, the contrast reveals a general problem. *Carte blanche* endorsement of both types of detachment (without some restriction) is not tenable, since it leads implausibly to the conclusion that we are both obligated to tell (the neighbor we are coming) and obligated to not tell. Thus researchers tended to divide up over which principle of the two they endorse [Loewer and Belzer, 1983]. The Factual Detachment camp typically endorse the view that the conditional in 3) in the Chisholm Quartet needs to be interpreted as a non-material conditional, but otherwise things are as they seem in 3): we have a conditional obligation that is a simple composite of a non-deontic conditional and a pure unary deontic operator in the consequent:

> $\mathbf{OB}(q/p) =_{df} p \Rightarrow \mathbf{OB}q$, for some independent non-material conditional.[92]

Typically, the conditional was a non-classical conditional of the sort made famous by Stalnaker and Lewis.[93] It is then generally maintained that deontic detachment is flawed, since the conditional obligations like those in (2) tell us only what to do in ideal circumstances, but they do not necessarily provide "cues"[94]

---

[91]As already noted, some reject both analyses and think deontic conditionals are *sui generis*. Note also that 2) above has the conditional explicitly in the scope of the English "ought to be" operator, and this is not explicitly a deontic conditional as just characterized unless we add that it should be read as at least necessarily equivalent to "if Jones does go, then he ought to tell them he is coming". There is no uniform agreement about this, although often the Chisholm Paradox is characterized so that both (2) and (3) above would have the same superficial form ("if . . . , then it ought to be that. . . "), with the deontic term appearing in the second clause. We have instead followed Chisholm's original formulation. In either event, the inference from (1) and (2) to "it ought to be that Jones tells" is also called "deontic detachment" as is that from their formal analogues in SDL, where $\mathbf{OB}$-K validates the inference from (1′) and (2′) to $\mathbf{OB}t$.

[92]Ignoring the Chisholm quartet, Smith [1994] notes that adding factual detachment to SDL with $\mathbf{OB}(q/p)$ interpreted as $\mathbf{OB}(p \rightarrow q)$, yields Mally's problem: $\vdash \mathbf{OB}p \leftrightarrow p$. That SDL yields the first half, $p \rightarrow \mathbf{OB}p$, given factual detachment, is easily seen. Just substitute p for q in FD to yield $\vdash p \,\&\, \mathbf{OB}(p \rightarrow p). \rightarrow \mathbf{OB}p$. Then, since $\vdash \mathbf{OB}(p \rightarrow p)$ by $\mathbf{OB}$-N, it can fall out and we get $\vdash p \rightarrow \mathbf{OB}p$. Note that the proof depends crucially on the highly controversial rule of necessitation. However, Smith, crediting Andrew Jones, pointed out that even a very minimal deontic logic entails the second half of the equivalence in question, $\mathbf{OB}p \rightarrow p$, which is still enough to make Voltaire grin.

Thm: For any system with PL, $\mathbf{OB}$-D & $\mathbf{OB}$-RE, FD yields $\vdash \mathbf{OB}p \rightarrow p$.

Proof: Assume PL, $\mathbf{OB}$-D, $\mathbf{OB}$-RE and FD. By substitution of $\perp$ for q in FD, we get $\vdash p \,\&\, \mathbf{OB}(p \rightarrow \perp) \rightarrow \mathbf{OB}\perp$. So from that and $\mathbf{OB}$-D, we get $\vdash \sim(p \,\&\, \mathbf{OB}(p \rightarrow \perp))$, that is $\vdash \sim p \vee \sim\mathbf{OB}(p \rightarrow \perp)$. From the latter by $\mathbf{OB}$-RE we get $\vdash \sim p \vee \sim\mathbf{OB} \sim p$, that is $\vdash \mathbf{OB} \sim p \rightarrow \sim p$, which by substitution of $\sim p$ for p, along with $\mathbf{OB}$-RE, yields $\vdash \mathbf{OB}p \rightarrow p$.

Given how minimal $\mathbf{OB} - RE$ and $\mathbf{OB} - D$ are, the friend of factual detachment with conditionals so interpreted cannot shrug this off.

[93]See [Mott, 1973; Chellas, 1974; 1980] for examples, and [DeCew, 1981] for an influential critical evaluation, arguing that although such conditionals are indeed important, there is still a special conditional they overlook at the heart of the Chisholm puzzle.

[94]van Eck [1982].

for action *in the actual world*, where things are often typically quite sub-ideal, as (4) combined with (1) indicates. Thus from the fact that Jones ought to go and he ought to tell if he goes, it doesn't follow that what he ought to *actually* do is tell — that would be so only if it was also a fact that he goes to their aid. At best, we can only say that he ought *ideally* to go.

This suggestion seems a bit more difficult when we change the conditional to something like "If Doe does kill his mother, then it is obligatory that Doe kills her gently". The idea that my obligation to not kill my mother gently (say for an inheritance) merely expresses an "ideal" obligation, but not an actual obligation, given that I will kill her, seems hard to swallow. So this case makes matters a bit harder for those favoring a factual detachment approach for generating actual obligations. Similarly, it would seem that if it is impermissible for me to kill my mother, then it is impermissible for me to do so gently, or to do so while dancing.[95] So *carte blanche* factual detachment seems to allow the mere fact that I will take an action in the future (killing my mother) that is horribly wrong and completely avoidable now to render obligatory another horrible (but slightly less horrible) action in the future (killing my mother gently). The latter action must be completely avoidable if the former is, and the latter action is one that I would *seem* to be equally obligated to not make intuitively.

The main alternative camp represented conditional obligations via dyadic non-composite obligation operators modeled syntactically on conditional probability. They rejected the idea that $\mathbf{OB}(q/p) =_{\mathrm{df}} p \Rightarrow \mathbf{OB}q$, for some independent conditional. In a sense, on this view, deontic conditionals are viewed as idioms: the meaning of the compound is not a straightforward function of the meaning of the parts. The underlying intuition regarding the Chisholm example is that even if it might be true that we will violate some obligation, that doesn't get us off the hook from obligations that derive from the original one that we will violate. If I must go help and I must inform my neighbors that I'm coming, if I do go help, then I must inform them, and the fact that I will in fact violate the primary obligation does not block the derivative obligation anymore than it does the primary one itself.

One early semantic picture for the latter camp was that a sentence of the form $\mathbf{OB}(q/p)$ is true at a world i iff the i-best p-worlds are all q-worlds. $\mathbf{OB}q$ is then true iff $\mathbf{OB}(q/\top)$, and so iff all the unqualifiedly i-best worlds are q-worlds [Hansson, 1969]. Note that this weds preference-based semantic orderings with dyadic conditional obligations.[96] This reflects a widespread trend. Factual detachment does not work in this case, since even if our world is an I-don't-go-help-world, and the best among the I-don't-go-help-worlds are I-don't-call-worlds, it does not follow that the *unqualifiedly* best worlds are I-don't-call-worlds. In fact, in this example, these folks would maintain, the unqualifiedly best worlds are both I-go

---

[95]In Chisholm's example it is easier to accept that telling is merely *ideal*, but not required, since it is easy to interpret Chisholm's example as one where giving advanced notice is what the agent perhaps *ought* to do, but not something the agent *must* do (even assuming the neighborly help is itself a must).

[96]As Makinson [1993] notes, it was also a forerunner of semantics for defeasible conditionals generally (cf. "if p, normally q").

worlds and I-call worlds, and the fact that I won't do what I'm supposed to do won't change that.

But one is compelled to ask those in the Deontic Detachment Camp: what then is the point of such apparent conditionals if we can't ever detach them from their apparent antecedents, and how are these conditionals related to regular ones? This seems to be the central challenge for this camp. Thus they often endorse a restricted form of factual detachment, of which the following is a representative instance:

> *Restricted Factual Detachment*: $\Box p \,\&\, \mathbf{OB}(q/p). \to \mathbf{OB}q.$

Here $\Box p$ might mean various things, for example that p is physically unalterable or necessary as of this moment in history.[97] Only if p is settled true in some sense, can we conclude from $\mathbf{OB}(q/p)$ that $\mathbf{OB}q$. This certainly helps, but it still leaves us with a bit of a puzzle about why this apparent composite of a conditional and a deontic operator is actually some sort of primitive idiom involving a non-stated alethic modal operator.

So it seems like we are left with a dilemma: either (1) you allow factual detachment and get the consequences earlier noted to the effect that simply because someone will act like a louse, he is obligated to do slightly mitigating louse-like things, or (2) instead you claim that "if p, then ought q" is really an idiom, and the meaning of the whole is not a function of the meaning of its conditional and deontic parts. Each seems to be a conclusion one would otherwise prefer to avoid.

There have been many attempts to try to solve Chisholm's problem by carefully distinguishing the times of the obligations.[98] This was fueled in part by shifts in the examples, in particular to examples where the candidate "derived" obligations were clearly things to be done *after* the primary obligation was either fulfilled or violated (called "forward" versions of CTDs). This made the ploy of differentiating the times and doing careful bookkeeping about just which things were obligatory at which times promising. However, Chisholm's own example is most plausibly interpreted as either a case where the obligation to go help and the perhaps-derivable obligation to tell are *simultaneous* (called "parallel" versions), or where telling is even something to be done *before* you go (called "backward" versions).[99] It is easy to imagine that the way to tell the neighbors that you will help might be to phone, and that would typically take place before you left to actually help. (For younger readers: there were no cell phones back in 1964, and phones were attached to boxes in houses by yard-length coiled chords.) Concerns to coordinate

---

[97]The idea is perhaps implicit in [Hansson, 1969]; it is argued for explicitly in [Greenspan, 1975], and adopted by many since.

[98]Thomason [1981b; 1981a] are classics arguing for the general importance of layering deontic logic on top of temporal logic. [Åqvist and Hoepelman, 1981, Thomason, 1981b, van Eck, 1982, Loewer and Belzer, 1983, Feldman, 1986] argue that attention to time is crucial (or at least helpful) in handling the Chisholm puzzle, among other puzzles.

[99][DeCew, 1981]. [Smith, 1994] contains an illuminating discussion of the three different versions (backward, parallel, and forward) in evaluating different approaches to solving the Chisholm paradox.

aid, or to assure those stressed that aid is coming, often favor giving *advanced* notice.

Alternatively, it was suggested that carefully attending to the action or agential components of the example and distinguishing those from the circumstances or propositional components would dissolve the puzzle.[100] However, the phenomena invoked in the Chisholm example appear to be too general for that. Consider the following non-agential minor variant of an example (say of possible norms for a residential neighborhood) introduced in [Prakken and Sergot, 1996]:

1. It ought to be the case that there are no dogs.

2. It ought to be the case that if there are no dogs, then there are no warning signs.

3. If there are dogs, then it ought to be the case that there are warning signs.

4. There are dogs.

Here we seem to have the same essentially puzzling phenomena present in Chisholm's original example, yet there is no apparent reference to actions above at all; instead the reference seems to be to states of affairs only. (Notice also that there is no issue of different times either for the presence/absence of dogs and the presence/absence of signs.)

Thus, it looks like tinkering with the temporal or action aspects of the Chisholm-style examples (however much time and action are important elsewhere to deontic logic) merely postpones the inevitable. So far, this problem appears to be not easy to convincingly solve.

A formal sketch of a sample system favoring factual detachment can be easily found in [Chellas, 1980; Chapter 10], which is widely available. A system that favors deontic detachment over factual detachment is quickly sketched in the following box (see [Goble, Forthcoming-b], and [van Fraassen, 1972] for a similar system).

Here, we assume a classical propositional language now extended with a dyadic construction, $\mathbf{OB}(/)$, taken as primitive. A monadic $\mathbf{OB}$ operator is then defined in the manner mentioned above:

$$\mathbf{OB}p =_{\mathrm{df}} \mathbf{OB}(p/\top).$$

We can define an ordering relation between propositions as follows:

$$p \geq q =_{\mathrm{df}} {\sim}\mathbf{OB}({\sim}p/p \vee q).$$

---

[100][Castañeda, 1981; Meyer, 1988].

This says that p is ranked as at least as high as q iff it is not obligatory that $\sim p$ on the condition that either p or q.[101]

An axiom system that is a natural dyadic correlate to SDL follows:

A1:	All instances of PC tautologies	(TAUT)
A2:	$\mathbf{OB}(p \rightarrow q/r) \rightarrow (\mathbf{OB}(p/r) \rightarrow \mathbf{OB}(q/r))$	($\mathbf{OB}$-CK)
A3:	$\mathbf{OB}(p/q) \rightarrow \sim\mathbf{OB}(\sim p/q)$	($\mathbf{OB}$-CNC)
A4:	$\mathbf{OB}(\top/\top)$	($\mathbf{OB}$-CN)
A5:	$\mathbf{OB}(q/p) \rightarrow \mathbf{OB}(q \,\&\, p/p)$	($\mathbf{OB}$-CO $\&$)
A6:	$(p \geq \mathrm{q} \,\&\, q \geq \mathrm{q}) \rightarrow p \geq \mathrm{q}$	(Trans)
R1:	If $\vdash p$ and $\vdash p \rightarrow q$ then $\vdash q$	(MP)
R2:	If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}(r/p) \leftrightarrow \mathbf{OB}(r/q)$	($\mathbf{OB}$-CRE)
R3:	If $\vdash p \rightarrow q$ then $\vdash \mathbf{OB}(p/r) \rightarrow \mathbf{OB}(q/r)$	($\mathbf{OB}$-CRM)

A1–A4 and R1–R3 are conditional analogues of formulas or rules we have seen before in discussing axiomatizations of SDL itself. A5 and A6 are needed to generate a complete system relative to an ordering semantics of the following sort (merely sketched here).

Assume we have a set of worlds and a set of ordering relations, P_i, for each world, i, where jP_ik is to be interpreted as saying that relative to i's normative standards, j is at least as good as k. Assume also that all of the ordering relations are non-empty: for each world i, there is a world k and a world m such that mP_ik. Call this structure a "preference frame". For any preference relation in a preference frame, let $F(P_i)$ represent the field of that relation: the set of all worlds that appear in some ordered pair constituting the relation, P_i. As usual, a model on a frame is an assignment to each propositional variable of a set of worlds, (those where it will be deemed true). We then define the basic dyadic operators truth-condition as follows:

$M \vDash_i \mathbf{OB}(q/p)$ iff there is a j in $F(P_i)$ such that $M \vDash_j p \,\&\, q$ and for each k such that kP_ij, if $M \vDash_k p$, then $M \vDash_k q$.

That is, at i, it is obligatory that q given p iff there is some world j in the field of i's preference relation where both p and q are true, and for every world ranked at least as high as j, if p is true at that world, then so is q.

Call a preference frame standard iff all the preference relations in it are connected (and thus reflexive), and transitive relative to their fields:

For each i-relative preference relation, P_i,

1. if j and k are in $F(P_i)$, then either jP_ik or kP_ij (connectedness).

2. if j, k, and m are in $F(P_i)$, then if jP_ik and kP_im, then jP_im (transitivity).

---

[101] Compare p is permissible given $p \vee q$, where $\mathbf{PE}(p/q) =_{\mathrm{df}} \sim\mathbf{OB}(\sim p/q)$.

Goble [Forthcoming-b] shows that the axiom system for dyadic obligation above is sound and complete for the set of standard preference frames. It is also easy to derive SDL using the above dyadic axiom system and the definition given for the monadic obligation operator. Goble's paper contains a number of other such results, for both monadic and dyadic systems, including generalizations that allow for conflicting obligations.

A.3   Doing Well Enough (DWE)[102]

A.3.1   DWE Syntax

Assume that we have a language of classical propositional logic with these additional (personal but non-agential) *primitive* unary operators:

OBp: It is *Obligatory* (for S) that p
MAp: The *Maximum* (for S) involves p
MIp: The *Minimum* (for S) involves p
INp: It is *Indifferent* (for S) that p

We might then tentatively analyze some other agential deontic notions as follows:

"*S must* bring it about that p": **OB**BAp
"*S ought* to bring it about that p": **MA**BAp
"The *least S* can do involves bringing it about that p": **MI**BAp
"It is a matter of indifference for S to bring it about that p": **IN**BAp.

Suppose that I am obligated to contact you to conduct some business, and that I can do so by emailing you, calling you, or stopping by. Add that these are the only ways to conduct the business.[103] Now imagine that the morally relevant value of these actions matches the extent to which the response is personal. Assuming you would not let me conduct our business twice, the three alternatives are exclusive. Then it is *obligatory* for me that I contact you in one of the three ways, but no one in particular, since any one of the three will discharge my obligation to contact you. Now if I choose to discharge my obligation in the minimally acceptable way, I will do so by email rather than by telephone or in person. So *doing the minimum involves* emailing you. On the other hand, if I conduct the business in person, I will have discharged my obligation in the optimal way. *Doing the maximum (what morality recommends) involves* stopping by your place. Finally, we can easily imagine that nothing of moral worth hinges on whether I wear my black socks when I contact you. So wearing them is *a matter of moral indifference.* This illustrates one application of the four primitive operators.

We introduce some defined operators, and their intended readings:

---

[102]We presuppose a simple no-conflicts atmosphere.
[103]To minimize complications, we will assume no one else can do these things on my behalf.

$\textbf{PE}p =_{\mathrm{df}}$	$\sim\!\textbf{OB}\sim\!p.$	(It is *Permissible* for S that p.)
$\textbf{IM}p =_{\mathrm{df}}$	$\textbf{OB}\sim\!p.$	(It is *Impermissible* for S that p.)
$\textbf{GR}p =_{\mathrm{df}}$	$\sim\!\textbf{OB}p.$	(It is *Gratuitous* for S that p.)
$\textbf{OP}p =_{\mathrm{df}}$	$\sim\!\textbf{OB}p\,\&\,\sim\!\textbf{OB}\sim\!p.$	(It is *Optional* for S that p.)
$\textbf{SI}p =_{\mathrm{df}}$	$\sim\!\textbf{IN}p.$	(It is *Significant* for S that p.)
$\textbf{SU}p =_{\mathrm{df}}$	$\textbf{PE}p\,\&\,\textbf{MI}\sim\!p.$	(It is *Supererogatory* for S that p.)
$\textbf{PS}p =_{\mathrm{df}}$	$\textbf{PE}p\,\&\,\textbf{MA}\sim\!p.$	(It is *Permissibly Suboptimal* for S that p.)

Continuing with our example, note that although the three alternatives, conducting the business by email, phone, or in person, are not on a par morally speaking, each is still *morally optional*. For each, the agent is permitted to do it or to refrain from doing it. Now we saw that doing the minimum involves emailing you. But suppose that rather than e-mailing you, I either call or stop by. Both of the latter alternatives are *supererogatory*. In each case, I will have done *more than I had to do* — more good than I would have if I had done the minimum permitted.[104] On the other hand, if I do not stop by, I will have done something sup-optimal, but, since emailing you and calling you are each nonetheless permissible, each is *permissibly suboptimal*. Finally, although each of the three ways of contacting you is optional, none is without *moral significance*. For whatever option I take of the three, I will have done something supererogatory or I will have done only the minimum; in either case, I will have done something with *moral significance*.

Where "$*$" ranges over $\textbf{OB}$, $\textbf{MA}$, $\textbf{MI}$, the associated $\textbf{DWE Logic}$ is:

$A0.$ All tautologous DWE-wffs;
$A1.$ $*(p \to q) \to (*p \to *q)$
$A2.$ $\textbf{OB}p \to (\textbf{MI}p\,\&\,\textbf{MA}p)$
$A3.$ $(\textbf{MI}p \vee \textbf{MA}p) \to \textbf{PE}p$
$A4.$ $\textbf{IN}p \to \textbf{IN}\sim\!p$
$A5.$ $\textbf{IN}p \to (\sim\!\textbf{MI}p\,\&\,\sim\!\textbf{MA}p)$
$A6.$ $(\textbf{OB}(p \to q)\,\&\,\textbf{OB}(q \to r)\,\&\,\textbf{IN}p\,\&\,\textbf{IN}r) \to \textbf{IN}q$

$R1.$ If $\vdash p$ and $\vdash p \to q$ then $\vdash q$
$R2.$ If $\vdash p$, then $\vdash \textbf{OB}p$.

It is easily shown that SDL logics for $\textbf{OB}$, $\textbf{MA}$, and $\textbf{MI}$ are derivable from DWE [McNamara, 1996c].

The increased complexity brought on by the enriched expressive power is graphically reflected in analogues to SDL's deontic hexagon and threefold partitions.[105]

---

[104]I slough over subtleties here about different senses of the philosopher's term "supererogatory".

[105]Recall our prior scheme for diagrams:

   <u>Arrowed Lines</u> represent implications   <u>Dotted Lines</u>: connects sub-contraries
   <u>Dashed Lines</u>: connects contraries   <u>Dotted-Dashed Lines</u>: connect contradictories

THE DEONTIC OCTODECAGON — PART I
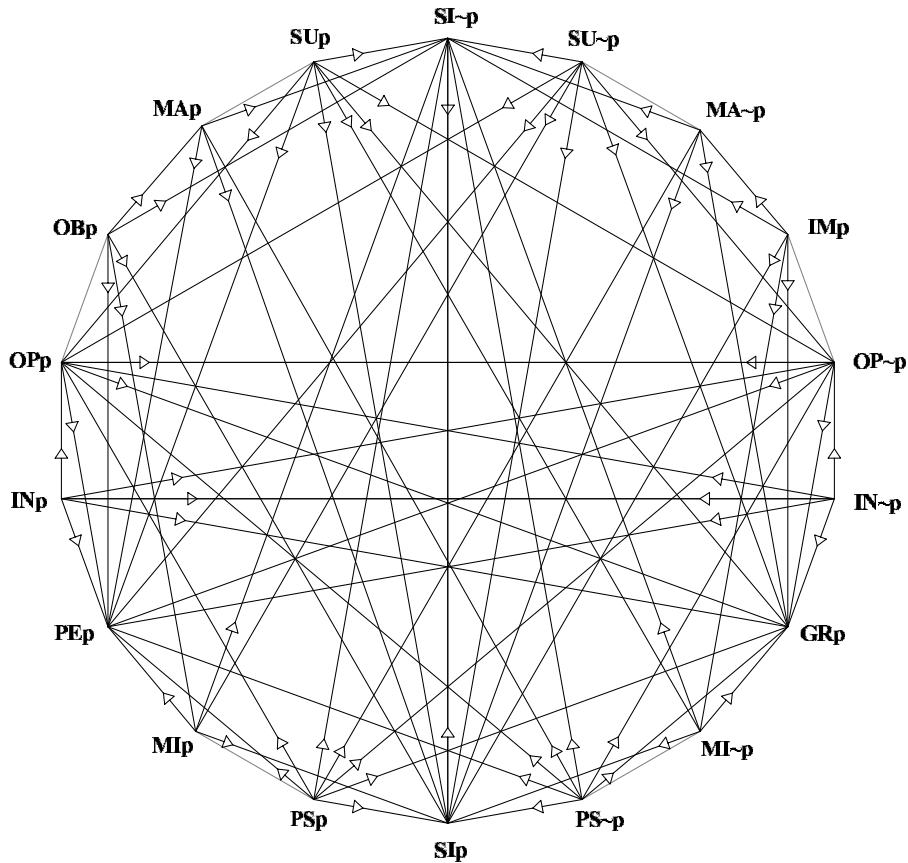
Grayed Plain Border Lines: added for purely aesthetic reasons.

Operator Key:
OBp: it is obligatory that p (cf. "must").
PEp: it is permissible that p (cf. "can").
IMp: it is impermissible that p (cf. "can't").
GRp: it is gratuitous that p.
OPp: it is optional that p.
MAp: doing the maximum involves p (cf. "ought").
MIp: doing the minimum involves p (cf. "the least one can do involves").
SUp: it is supererogatory that p (cf. "exceeding the minimum").
PSp: it is permissibly suboptimal that p (cf. "you can, but ought not").
INp: it is indifferent that p.
SIp: it is significant that p.

THE DEONTIC OCTODECAGON — PART II
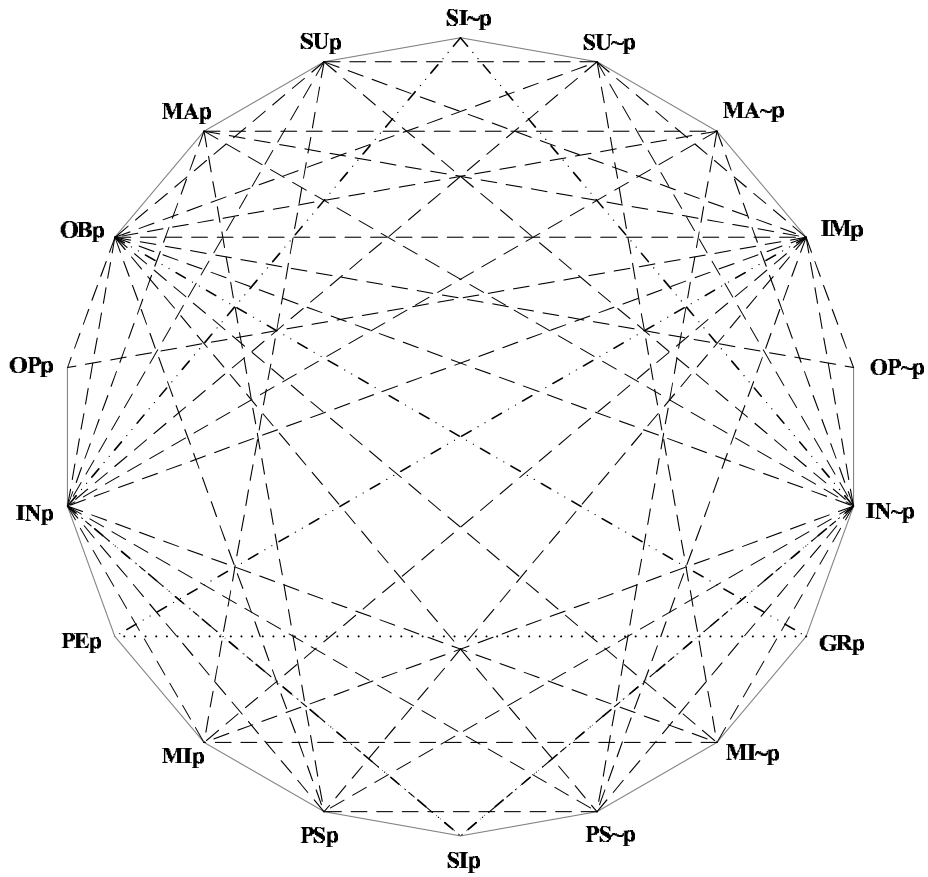
The Deontic Octodecagon is the result of the superimposition of Part II on Part I.[106]

---

[106]Roderick Chisholm brought my attention to the similarity between these diagrams, and those in [Hrushka and Joerden, 1987]. I began creating a series of diagrams expanding on the deontic square in the early 1980s prompted by remarks from Fred Feldman in an ethical theory class with him at the University of Massachusetts.

THE TWELVEFOLD PARTITION

PE

GR

OP

OB (MI & MA)	**MA** MI~ (PS~)	**MA** MI (PS~)	**MA** ~MI & ~MI~ (PS~)	**IN** (~MA & ~MA~) (~MI & ~MI~)	**IM** (MI~ & MA~)
	MA~ MI~ (PS)	**MA~** MI (PS)	**MA~** ~MI & ~MI~ (PS~)		
	~MA & ~MA~ MI~	~MA & ~MA~ MI	~MA & ~MA~ ~MI & ~MI~		
	SU	SU~	~SU & ~SU~		SI

SI
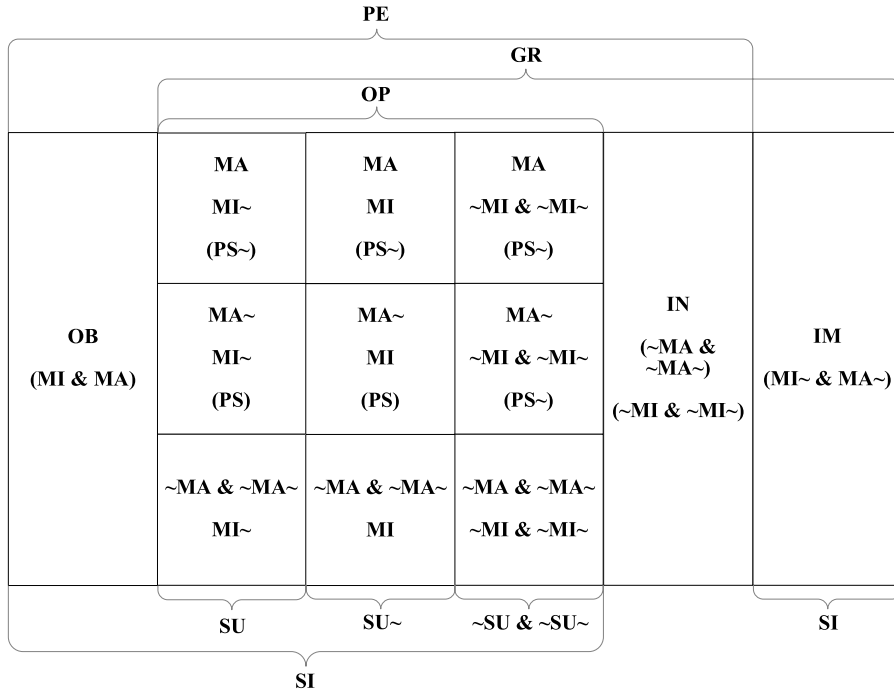
The partition is drawn with the black lines. As with the Traditional Threefold Classification, the twelve cells are mutually exclusive and jointly exhaustive. Parenthetical operators, as well as those tagged to grayed curly brackets outside the partition, highlight the location of various nonfinest classes within the partition. Below, the twelve classes are defined via schemata, using only primitives, without redundancies.

THE TWELVE FINEST CLASSES EXPRESSED VIA SCHEMATA

OB	**MA** ~& ~**MI** & ~**MI** ~
MA & **MI** ~	~**MA** & ~**MA** ~ & **MI** ~
MA & **MI** & ~**OB**	~**MA** & ~**MA** ~ & **MI**
MA & ~**MI** & ~**MI** ~	~**MA** & ~**MA** ~ & ~**MI** & ~**MI** ~ & ~**IN**
MA ~& **MI** ~& ~**O***B* ~	**IN**
MA ~& **MI**	**OB** ~

We turn now to one semantic framework for this logic.

A.3.2   DWE Semantics

To get the semantic structures we need, we simply combine and interpret two familiar ingredients in a convenient way: an accessibility relation and an ordering

relation. We imagine that we have a set of worlds, and an accessibility relation
— interpreted here as relating worlds to their morally *acceptable* alternatives.
We assume that seriality holds: for each world, there is a morally acceptable
alternative. Note that we do not think of these acceptable worlds as morally
ideal or *optimal* alternatives. Rather, we assume that for any world i, there is
a morally relevant i-relative weak ordering of the i-acceptable worlds (i.e. the
i-relative ordering relation is reflexive, connected, and transitive with respect to
the i-acceptable worlds). Thus, although all the acceptable alternatives to a given
world are just that — acceptable, they needn't be on a par morally speaking.
Some may be ranked higher than others, some may be ranked highest or lowest
among the acceptable worlds, and there may be ties throughout (and thus there
may be genuine *levels* of acceptable worlds). We can represent the i-acceptable
worlds and their i-relative ordering as follows:
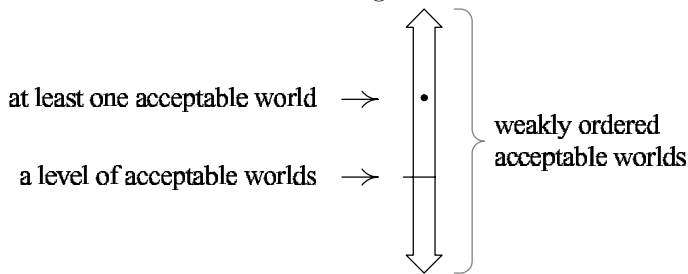
The vertical arrowed bar represents the weakly ordered i-acceptable worlds.
The horizontal line through the bar is a reminder that there can be levels of i-
acceptable worlds (each an equivalence class with respect to equi-rank), as is the
fact that we choose vertical figures with width. The dot indicates there is always
at least one i-acceptable world in these structures.

We can informally represent the truth-conditions (relative to a world i) for the
traditional SDL operators as follows (where a "ˆ" under an operator indicates that
it is primitive in DWE).
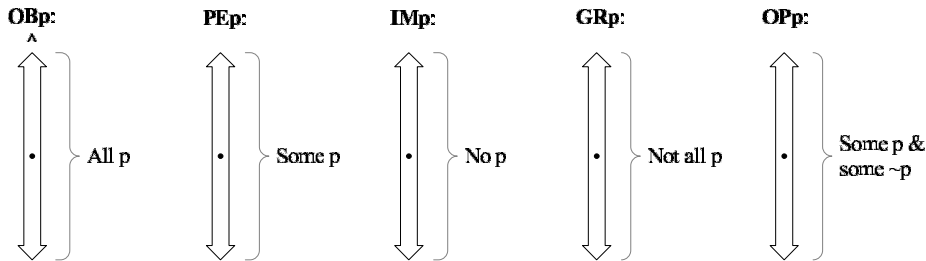
For these operators, the interpretation does not depend on the ordering and
matches that for SDL: p is *obligatory* (for agent S) iff p occurs in all of the i-
acceptable alternatives; p is *permissible* iff it occurs in some, etc. However, the
interpretation of the remaining operators depends crucially on the ordering of the
i-acceptable worlds:
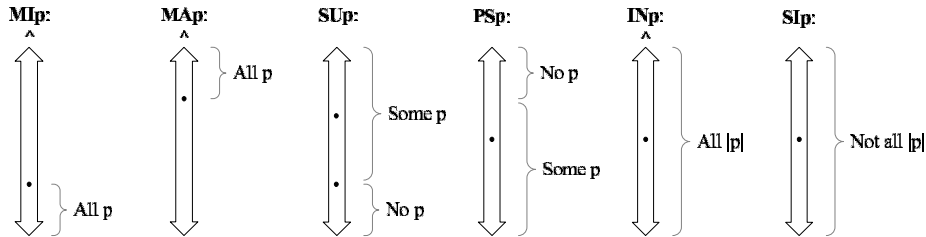
The *minimum* (for S) involves p iff p holds in all the lowest ranked acceptable alternatives; whereas *the maximum (optimum)* involves p iff it holds in all the highest ranked acceptable alternatives.[107] Thus as cast here, the minimum and the maximum are linked to the respective poles of the ranked acceptable alternatives and are mirror images of one another, which effects various symmetries in the logic [McNamara, 1996a]. p will be *supererogatory* if it holds in some acceptable alternative, but fails to hold in any of the lowest ranked acceptable alternatives. Similarly, p will be *permissibly suboptimal* if it holds in some acceptable alternative, but fails to hold in any of the highest ranked acceptable alternatives. Regarding moral indifference and moral significance, since we allow for ties, the ranked acceptable alternatives can be divided into "levels" (equivalence classes with respect to equal rank). An "all $|p|$" indicates that both p-worlds and $\sim p$-worlds occur at each of the associated <u>levels</u>. p will then be a matter of *moral indifference* if at every such level its performance and its non-performance occurs somewhere therein. Conversely, p will be *morally significant* if there is some level of value that uniformly includes it or uniformly excludes it.

---

[107]These are informally cast assuming lower and upper limit assumptions hold. The informal glosses can be easily adapted to discharge these assumptions, and the formal clauses below do not depend on any such boundedness.

DWE Formal Semantics: The following formal semantics is generalized in [Mares and McNamara, 1997].
Frames are defined as follows:

$F = \langle W, A, \leq \text{j} \rangle$ is a <u>DWE-Frame</u>:

1. W is non-empty

2. A is a subset of W^2 and A is serial: (Aij: j is an i-acceptable world)

3. $\leq$ is a subset of W^3:

    (a) ($k \leq_i j$ or $j \leq_i k$) iff ($Aij \,\&\, Aik$), for any i, j, k in W

    (b) if $j \leq_i k$ and $k \leq_i l$ then $j \leq_i l$, for any i, j, k, l in W.

The notions of an assignment and a model are then easily defined:

<u>P is an Assignment on F</u>: $F = \langle W, A, \leq \rangle$ is a DWE-Frame and P is a function from PV to Power(W), defined on PV (Propositional Variables).

<u>$M = \langle F, P \rangle$ is a DWE-Model</u>: $F = \langle W, A, \leq \rangle$ is a DWE-frame and P is an assignment on F.

<u>Truth at an Index in a Model</u>: Let $M = \langle F, P \rangle$ be a DWE-model, where $F = \langle W, A, \leq \rangle$ and $j =_i k =_{\mathrm{df}} j \leq_i k \,\&\, k \leq_i j$; then truth at a world in a model ($M \vDash_i$), truth in a model, and validity are easily defined:

<u>Basic Truth-Conditions at a world, i, in a Model, M</u>:

0. (Conditions for variables and truth functional connectives)

1. $M \vDash_i \mathbf{OB}p : \forall j$ (if Aij then $M \vDash_j p$).

2. $M \vDash_i \mathbf{MA}p : \exists j(Aij \,\&\, \forall k$ (if $j \leq_i k$ then $M \vDash_k p$)).

3. $M \vDash_i \mathbf{MI}p : \exists j(Aij \,\&\, \forall k$ (if $k \leq_i j$ then $M \vDash_k p$)).

4. $M \vDash_i \mathbf{IN}p : \forall j$ [if Aij then $\exists k(k =_i j \,\&\, M \vDash_k p) \,\&\, \exists k(k =_i j \,\&\, M \vDash_k {\sim}p)$]

<u>Derivative Truth Conditions</u>:

5. $M \vDash_i \mathbf{PE}p : \exists j(Aij \,\&\, M \vDash_j p)$.

6. $M \vDash_i \mathbf{IM}p : \forall j($if Aij then $M \vDash_j {\sim}p)$.

7. $M \vDash_i \mathbf{GR}p : \exists j(Aij \,\&\, M \vDash_j {\sim}p)$.

8. $M \vDash_i \mathbf{OP}p : \exists j(Aij \,\&\, M \vDash_j p)$ and $\exists j(Aij \,\&\, M \vDash_j {\sim}p)$.

9. $M \vDash_i \mathbf{SI}p : \exists j[Aij \,\&\, \text{either } \forall k(if k =_i j \text{ then } M \vDash_k p) \text{ or } \forall k(if k =_i j$
   then $M \vDash_k \sim p)]$.

10. $M \vDash_i \mathbf{SU}p : \exists j(Aij \,\&\, M \vDash_j p) \,\&\, \exists j[Aij \,\&\, (k)(\text{if } k \leq_i j \text{ then } M \vDash_k \sim p)]$.

11. $M \vDash_i \mathbf{PS}p : \exists j(Aij \,\&\, M \vDash_j p) \,\&\, \exists j[Aij \,\&\, (k)(\text{if } j \leq_i k \text{ then } M \vDash_k \sim p)]$.

<u>Truth in a DWE-Model</u>: $M \vDash p$ iff $M \vDash_i p$, for every i in W of M.

<u>Validity</u>: $\vDash p$ iff $M \vDash p$, for all M.

In [Mares and McNamara, 1997], the metatheorem below is proven as a special case:

<u>Metatheorem</u>: The DWE-logic is determined by the class of DWE-models.

A.4   A Glimpse at STIT Theory and Deontic Logic

"STIT Theory" is so-called because it is a particular approach to constructions like "Jones sees to it that ___". The following exposition draws from [Horty, 2001] and [McNamara, 2004b]. STIT theory builds on a formal indeterministic "branching time" framework initiated by A. Prior, championed by R. Thomason, and now the basis of a robust research program anchored by N. Belnap, and summarized in [Belnap, 2001].[108] Here I concentrate on only a few elementary aspects of this sort of account of agency and provide just a glimpse of its employment in deontic logic. The reader is encouraged to consult the above two works, which are rich in details we can hardly touch on here. See also the works mentioned in the main essay under agency, especially those by Hilpinen, for further critical exposition of this approach to agency.

A.4.1   The Indeterministic Framework

The basic primitives are a set of 'moments', *Tree*, and a two-place ordering relation, *after*, defined on *Tree*. A *moment* (represented as a node below) is thought of as momentary world state (cf. instantaneous possible world slice). Moments are not to be confused with seconds or instants. One moment is (possibly) *after* a second iff the first is some still possible future moment of the second. Moments can branch forward (upward in the diagrams), toward the future, but not backward. (There can be more than one possible future, but only one past, at a moment.)

---

[108]This is the quintessential tome on STIT theory *per se*, and itself contains chapters on deontic logic in the context of STIT theory.
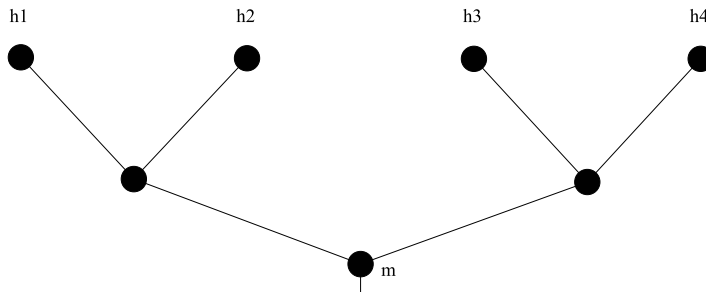
Upper moments are ones that can occur after line-connected lower moments. *A history* (cf. possible world) is construed as a maximal path or branch on a tree (e.g. each of the three-noded paths tracing from $h1$–$h4$ back to m above). In models with two or more histories, some moment (e.g. each top moment above) is not comparable to another (neither is a possible future moment of the other), and some moment is common to distinct histories (e.g. all but the top moments above). *A history passes through a moment* when that moment is part of that history. *The past at a moment* (in a history) is the ordered set of moments before the moment in question. *The future at a moment in a history*: the ordered moments in the history after the moment in question. (There is no actual future at a moment *per se*, since there are many possible such futures, unless determinism is true.)

Since the future is open, contingent future tensed statements, and thus all statements for uniformity, are assigned truth values at a moment-history *pair, m-h*. (I will sometimes ignore histories in formulations where uniformity is the only reason to mention them.) Here is a simple illustration, where "P" is the past tense operator "it was the case that", and "F" is the future tense operator "it will be the case that":
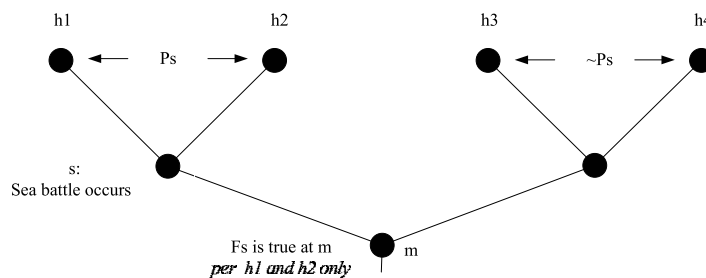
It will be the case that *s* at a moment *in a history* iff at some later moment *in that history*, *s* is true; and it was the case that *s* at a moment in a history iff that moment is after one where *s* is true in that history. Because the past is closed, simple past truths are true at a moment *per se*. The only case where we can say at a moment *simpliciter* that a statement will be true is where its future truth is historically necessary. More generally, possibility and necessity are handled as

follows: it is *(still) possible (POSS) that* s is true at a moment iff there is a history passing through that moment where s is true. It is *(now) necessary/settled* (NEC) that s at a moment iff s is true at every history passing through that moment. We can illustrate via the sea battle again, which is completely open as of m, but settled false in $h3$ and $h4$ just after m.
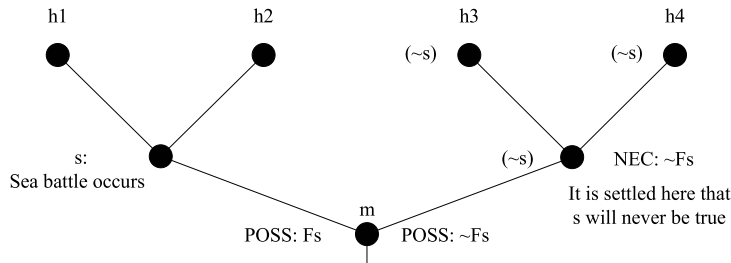
A.4.2 Agency

Next, a set of *Agents*, and a *Choice* function are introduced. The *Choice* function partitions the histories passing through a moment relative to each agent. Thus the agent's possible choices or basic actions (the cells) constitute a mutually exclusive and exhaustive division of the histories passing through that moment. Choices at a moment place instantaneous constraints on the possible futures. Intentions are not represented.[109] Where there is more than one cell, no particular basic action is determined at that moment. If a history is part of a choice cell, then that is *the choice the agent makes at that moment in that history.* Below Choice 1 is the set containing just $h1$ and $h2$; it rules out all that depends on either $h3$ or $h4$ unfolding. Choice 1 is the basic action the agent takes in $h1$ and in $h2$ at m.
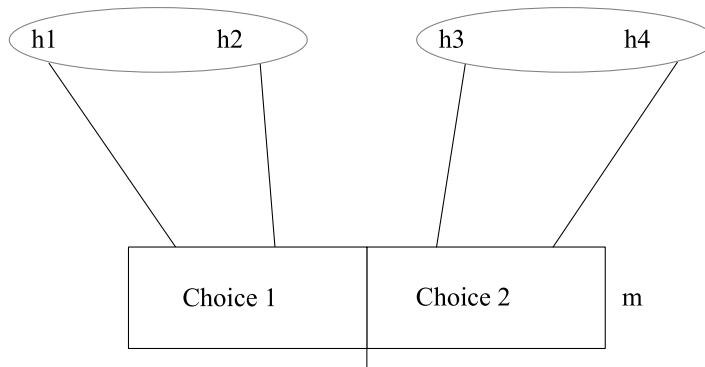
We can now easily distinguish two simple accounts of agency in terms of these basic actions or choices. The first is close to one Chellas gave in his seminal

---

[109]Thus "basic action" is perhaps better than "choice".

[Chellas, 1969]. Jane Doe *c-sees to it that (c-stit)* p at a moment-history pair iff p is guaranteed by the choice Jane takes at that moment in that history (i.e. that choice cell contains only p-histories). Jane Doe is *able to c-see to it that* p at a moment-history pair iff it is possible that she c-stit p at that pair. In the illustration below, Jane c-stit p at m in $h1$ and $h2$. However, she does not c-stit p at m in $h3$ and $h4$ (since p's truth value varies independently of choice 2), nor is she able to c-stit $\sim p$ at m (since no history passing through m involves a choice at m that guarantees $\sim p$).
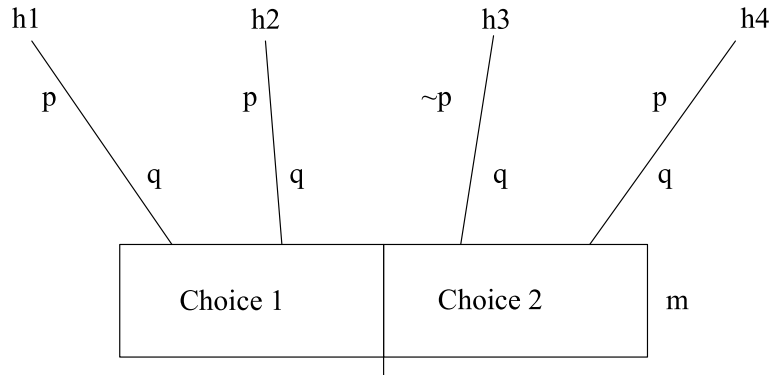
An obvious rub with c-stit is illustrated above: Jane c-stit q at m in $h1$–$h4$. The upshot is that agents see to everything that is historically necessary (e.g. that the sun will rise and that $2 + 2 = 4$). Enter: d-stit, which just adds the exclusion of necessary things for agency. Jane Doe *d-sees to it that (d-stit)* p at a moment-history pair iff she c-stit p at that pair and it is not necessary that p (i.e. $\sim p$ is consistent with some other choice open to her at m). This is called the 'deliberative stit' because the second condition, is meant to assure a real choice. Then, Jane Doe is *able to d-see to it that* p at a moment-history pair iff it is possible that she d-stit p at that pair. In the illustration above Jane d-sees to it that p at m only in $h1$–$h2$, but Jane does not d-see to it that q at m in any of $h1$–$h4$. Jane is able to d-see to it that p at m (for at a history passing through m she does see to it that p) but Jane is not able to d-see to it that $\sim p$ or that q at m.

Belnap's Achievement stit:

Belnap has a more complex alternative formal account of agency, which we can only briefly allude to here. The basic idea is that Jane sees to it that p now iff p now holds and was guaranteed by a *prior* choice of Jane's. Above, with c-stit and d-stit, one sees to it that something is the case at the moment of the choice or basic action; in Belnap's alternative the focus is on something's now holding as a result of a past action, so that the result and the initial instrumentality on the part of the agent that triggers the result are separated in time in this account of agency. Roughly, Belnap introduces the notion of instants as equivalence classes of contemporaneous moments. Intuitively, on a full tree-display, moments on the same level are contemporaneous with one another, and the time or instant is taken to be the set of these moments. Then, *Jane Doe a-stit that p at moment $m1$ in a history, h* iff 1) there is a moment in h, $m0$, that is earlier than $m1$, and p holds at the instant of $m1$ in all histories consistent with the choice Jane makes at $m0$ in h, but 2) there is also a moment that is after $m0$ (and thus was still possible then), that is contemporaneous with $m1$, and at which p is false.
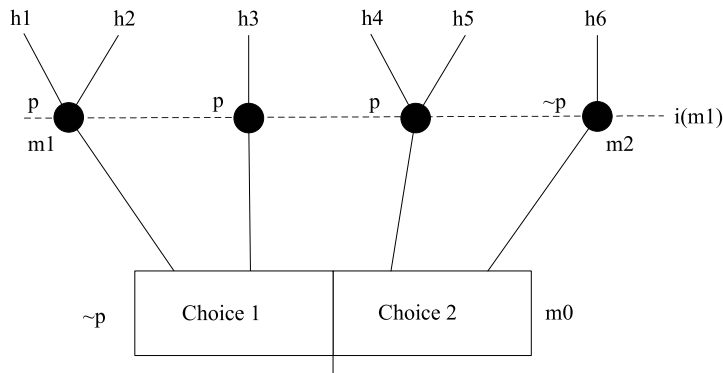
Above, in $h1$–$h3$, Jane a-stit p at $m1$ (but not at $m0$ where the choice resulting in p is made), but not in $h4$–$h6$ at any moment in $i(m1)$, since at $m2$, $\sim p$ holds.

A.4.3   Two Deontic Operators

Let's assume that histories have a rank-reflecting numerical value that does not vary from moment to moment (so histories are weakly ordered and thus mutually comparable). For simplicity, I will assume we always have best histories. Impersonal ought's may then be analyzed as follows: *it ought to be that p* holds at moment-history pair iff the best histories passing through m are histories where p holds. In the model below, since $h1$ and $h2$ are the highest ranked histories still

possible as of m, and it rains at each of these at the last moments listed in those histories, it follows that at m, it ought to be the case that it will rain.

h1: 10   h2: 10   h3: 5   h4: 2

rains   rains

O-be: F(rain)   m
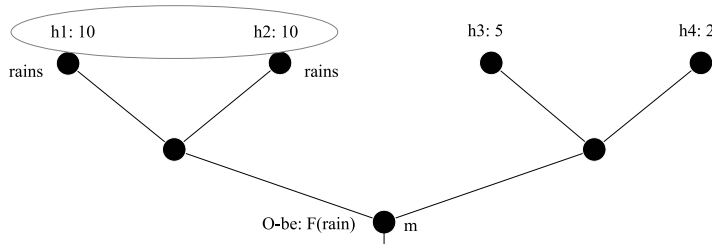
All the principles of Standard Deontic Logic (SDL), including no ought-conflicts, follow. A non-agential version of Kant's Law (it ought to be that p only if p is historically possible) also follows [Horty, 2001].

If we endorse the 'Meinong–Chisholm reduction', then recast in the c-stit framework (for simplicity), this becomes *an agent ought to see to p iff it ought to be the case that the agent c-sees to it that p*. Given the previously proposed semantics, an agent ought to c-see it that p holds at a moment-history pair iff that agent chooses a p-guaranteeing action at the best histories passing through that moment. In the diagram below, Jane ought to see to it that p at m (in $h1$–$h4$) whether we recast the Meinong–Chisholm reduction via c-stit or d-stit, since Choice 1 guarantees p, and Choice 2 is consistent with $\sim p$, and the best worlds are ones where Jane makes choice 1 and thus sees to it that p.

h1: 10   h2: 10   h3: 5   h4: 2

p p p $\sim p$

Choice 1   Choice 2   m
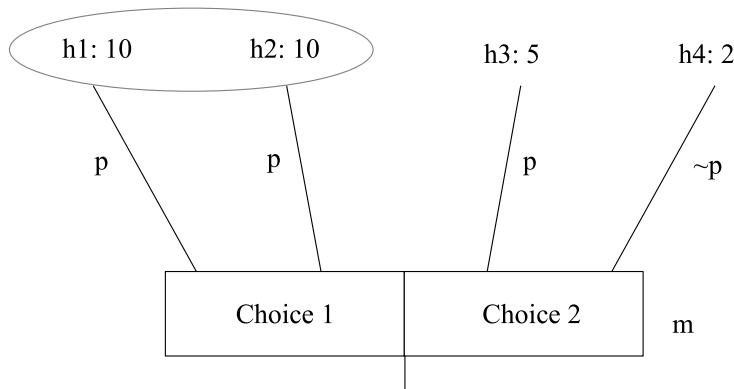
Relativized to c-stit (but not to d-stit) this analysis of agential ought's yields a normal modal operator satisfying the principles of SDL. An agential version of Kant's Law follows: an agent ought to see to p only if she is able to. It also follows that what an agent ought to do, ought to be, but that the converse does not hold

is illustrated by the next diagram. Here, although it ought to be that Jane makes Choice 1 and that p comes about, since both these things hold throughout the best histories (namely $h1$), it is not true that she ought to bring it about that p, since it is not true that it ought to be the case that she does. In the best world, $h1$, Jane does not make a choice that guarantees p's occurrence.
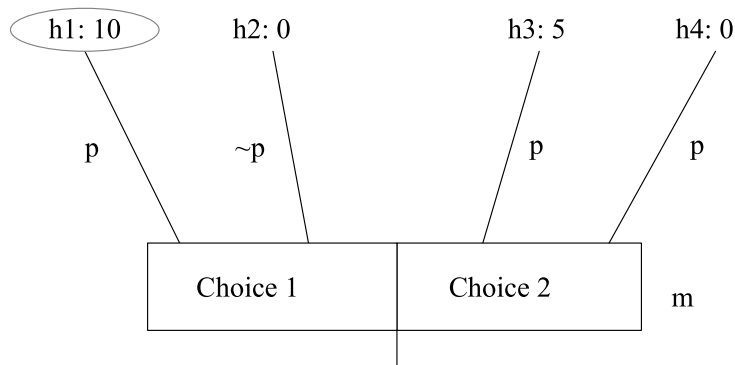
Refraining Again: Following a view championed by Belnap, if we analyze Jane's *refraining* from seeing to it that p as her d-seeing to it that she does not d-see to it that p (Jane d-stit $\sim$(Jane d-stit p)), and distinguish this from Jane's *omitting* p ($\sim$Jane d-stit p) then we can in fact say that it ought to be that Jane refrains from seeing to it that p above. She does d-see to it that p at m in $h3$ and $h4$, but these are suboptimal histories. So by making Choice 1 instead of Choice 2, she d-sees to it that she does not d-see to it that p, and this is what happens in the best history. Roughly, the best history is one where p occurs by luck or by some other agency than Jane's. Recall von Wright's alternative analysis of refraining: Jane refrains from seeing to it that p iff Jane is *able* to see to it that p, but she doesn't. Recast via d-stit this becomes Jane refrains from seeing to it that p iff $\sim$Jane d-stit p and it is possible that Jane d-stit p. It can be shown about Belnap's and von Wright's glosses, when recast via d-stit as indicated, that Jane Belnap-refrains from p iff S von-Wright refrains from p, and that Jane refrains from p iff Jane refrains from refraining from p. For c-stit, omitting and refraining are indistinguishable. (See [Horty, 2001].)

Horty considers some previous objections to the Meinong–Chisholm reduction, and argues that from the standpoint of his framework, these objections are unsound. He then introduces his own objection to the analysis, via the 'Gambling Problem'.[110] Suppose I have two options available to me, gamble \$5 (g) or not.

---

[110]Cf. [Feldman, 1986, pp.194–195].

Now suppose that if I gamble and win, I get \$10; and if I gamble and lose, I get \$0. Suppose the only values at stake are the dollar values, and thus the value of not gambling is \$5 saved. Ignore probabilities. To illustrate:
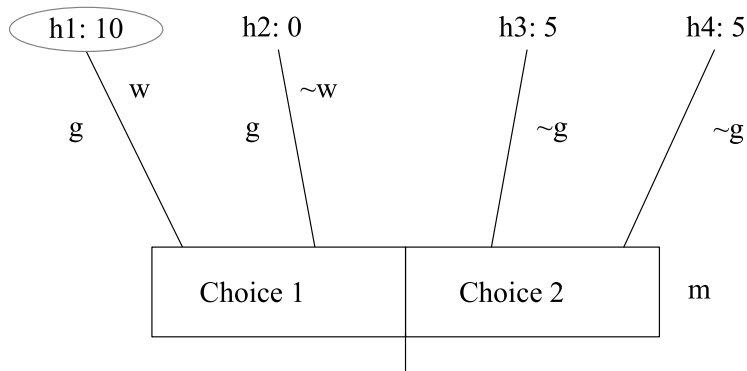
Since I cannot determine whether or not I win (this happens only in $h1$, which I can't guarantee), it is not true in fact that what I ought to do is gamble (or not gamble for that matter). But in the best histories ($h1$), I win, and my gambling is entailed by my winning, so it ought to be that I see to it that I gamble, and hence the Meinong–Chisholm reduction implies that I ought to gamble after all. Horty takes this to decisively defeat the Meinong–Chisholm reduction, arguing that we need an independent analysis, one where we can rank actions, not just whole histories. Horty goes on to develop an alternative analysis of agential ought's, one in which, among other things, he uses the ranking of histories to generate a decision-theoretic dominance ordering account of agential ought's. Here we must pass over this fascinating work, and simply raise a few quick questions about a few elementary matters.

A.4.4   Some Challenges

We saw that in the case of c-stit, an agent sees to all necessary truths. Few who work on agency accept this. D-stit is intended to get around this, but here seeing to it that p requires that it still might be that $\sim p$, thus making agency depend _logically_ on the falsity of compatibilism [Elgesem, 1997]. Thus _nothing_ inevitable can be the result of my agency. But compatibilism is a widely endorsed live option in philosophy. It does not seem that a logic for agency ought to presuppose the falsity of this widely endorsed philosophical view.

Furthermore, the stit framework seems to make it too easy to undermine genuine agency by making the conditions for agent causation too strong. Consider the following "Windy Day Assassin" scenario:
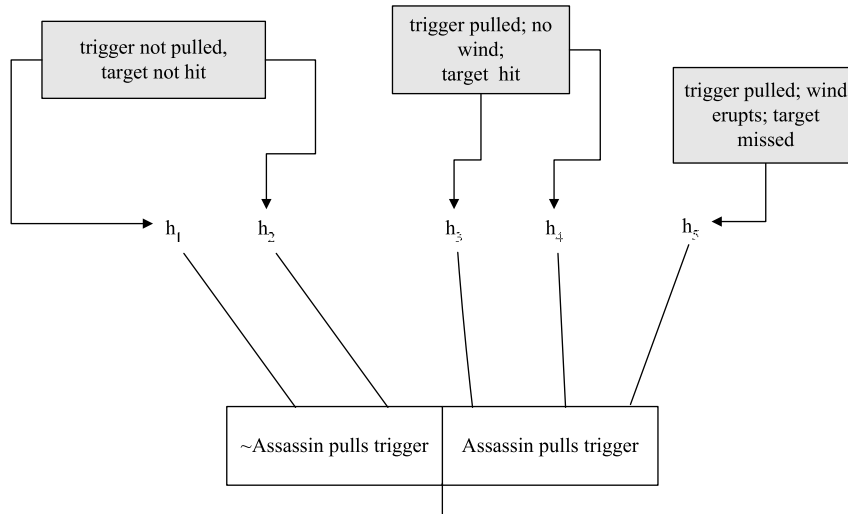
Suppose I pull the trigger of a gun aiming at you intending to kill you, and you are hit by the bullet and die as a result of being hit by that bullet, just as planned. Now add that when I pulled the trigger a random gust of wind *could* have occurred and knocked the bullet off target, though it didn't occur. On the current analyses, it follows that I did not see to it that you were hit, because no choice I made *guaranteed* that you were hit. The mere fact that the wind *could have* interfered with the course of the bullet is enough to undermine the claim that I was the agent of your being hit. Even if I aim, pull the trigger, and the wind doesn't blow, as in $h3$ and $h4$, I still don't see to it that the target is hit on stit theory. This smacks of getting away with murder. (This problem applies to c-stit (the Chellas-inspired stit operator), d-stit (the Horty–Belnap deliberative stit operator), and a-stit (Belnap's achievement stit operator).

Finally, there is a general problem with the analysis of impersonal ought's in terms of what is still possible. It is now tragically settled that some children will die of starvation tomorrow, but that ought to not now be the case. It seems false that everything that ought to be the case still could be the case. Talk of what would be ideal is not constrained by what is still possible, so why should talk of what ought to be the case, which involves, after all, just the evaluation of states of affairs, not of agent's actions, be any different?

RELEVANT AND SUBSTRUCTURAL LOGICS

Greg Restall

1   INTRODUCTION

Logics tend to be viewed of in one of two ways — with an eye to *proofs*, or with an eye to *models*.[1] Relevant and substructural logics are no different: you can focus on notions of proof, inference rules and structural features of deduction in these logics, or you can focus on interpretations of the language in other structures.

This essay is structured around the bifurcation between proofs and models: The first section discusses *Proof Theory* of relevant and substructural logics, and the second covers the *Model Theory* of these logics. This order is a natural one for a history of relevant and substructural logics, because much of the initial work — especially in the Anderson–Belnap tradition of relevant logics — *started* by developing proof theory. The model theory of relevant logic came some time later. As we will see, Dunn's algebraic models [1970; 1971] Urquhart's operational semantics [1972c; 1972d] and Routley and Meyer's relational semantics [1972a; 1972b; 1973] arrived many years after the initial burst of activity from Alan Anderson and Nuel Belnap. The same goes for work on the Lambek calculus: although inspired by a very particular application in linguistic typing, it was developed first proof-theoretically, and only later did model theory come to the fore. Girard's linear logic is a different story: it was discovered through considerations of the categorical models of coherence spaces. However, as linear logic appears on the scene much later than relevant logic or the Lambek calculus, starting with proof theory does not result in *too* much temporal reversal.

I will end with one smaller section *Loose Ends*, sketching avenues for further work. The major sections, then, are structured thematically, and inside these sections I will endeavour to sketch the core historical lines of development in substructural logics. This, then, will be a *conceptual* history, indicating the linkages, dependencies and development of the content itself. I will be less concerned with identifying who did what and when.[2]

I take it that logic is best learned by *doing* it, and so, I have taken the liberty to sketch the proofs of major results when the techniques used in the proofs tells us something distinctive about the field. The proofs can be skipped or skimmed without any threat to

---

[1]Sometimes you see this described as the distinction between an emphasis on *syntax* or *semantics*. But this is to cut against the grain. On the face of it, rules of proof have as much to do with the *meaning* of connectives as do model-theoretic conditions. The rules interpreting a formal language in a model pay just as much attention to *syntax* as does any proof theory.

[2]In particular, I will say little about the intellectual ancestry of different results. I will not trace the degree to which researchers in one tradition were influenced by those in another.

the continuity of the story. However, to get the full flavour of the history, you should attempt to savour the proofs at leisure.

Let me end this introduction by situating this essay in its larger context and explaining how it differs from other similar introductory books and essays. Other comprehensive introductions such as Dunn's "Relevance Logic and Entailment" [1986] and its descendant "Relevance Logic" [2001], Read's *Relevant Logic* [1988] and Troelstra's *Lectures on Linear Logic* [1992] are more narrowly focussed than this essay, concentrating on one or other of the many relevant and substructural logics.[3] The Anderson–Belnap two-volume *Entailment* [1975; 1992] is a gold mine of historical detail in the tradition of relevance logic, but it contains little about other important traditions in substructural logics.

My *Introduction to Substructural Logics* [2000a] has a similar scope to this chapter, in that it covers the broad sweep of substructural logics: however, that book is more technical than this essay, as it features many formal results stated and proved in generality. It is also written to introduce the subject purely *thematically* instead of *historically*.

2   PROOFS

The discipline of *relevant* logic grew out of an attempt to understand notions of consequence and conditionality where the conclusion of a valid argument is *relevant* to the premises, and where the consequent of a true conditional is *relevant* to the antecedent.

"*Substructural*" is a newer term, due to Schröder-Heister and Došen. They write:

> Our proposal is to call logics that can be obtained . . . by restricting structural rules, *substructural logics* [Schröder-Heister and Došen, 1993, p. 6].

The *structural rules* mentioned here dictate admissible forms of transformations of premises contained in proofs. Later in this section, we will see how relevant logics are naturally counted as substructural logics, as certain commonly admitted structural rules are responsible for introducing irrelevant consequences into proofs.

Historical priority in the field belongs to the tradition of relevant logic, and it is to the early stirrings of considerations of relevance that we will turn.

2.1   Relevant Implication: Orlov, Moh and Church

Došen has shown us [1992b] that substructural logic dates back at least to 1928 with I. E. Orlov's axiomatisation of a propositional logic weaker than classical logic [1928].[4]

---

[3]You have now been introduced to the geographical bifurcation of terminology. Americans call our topic "relevan*ce*" logic and people of Commonwealth countries (primarily Australia and Scotland) call it "relevan*t*" logic. The split comes down to a disagreement between Nuel Belnap and Robert Meyer. Meyer brought his favoured terminology "relevant" to Australia with him, where it has stuck. I have been taught in this tradition, so I also call what I study *relevant logic*, though nothing of substance hangs on the issue.

[4]Allen Hazen has shown that in Russell's 1906 paper "The Theory of Implication" his propositional logic (without negation) is free of the structural rule of contraction [Hazen, 1997; Russel, 1906]. Only after negation is introduced can contraction be proved. However, there seems to be no real sense in which Russell could be

Orlov axiomatised this logic in order to "represent relevance between propositions in symbolic form" [Došen, 1992b, p. 341]. Orlov's propositional logic has this axiomatisation.[5]

○	$A \rightarrow \sim\sim A$	double negation introduction
○	$\sim\sim A \rightarrow A$	double negation elimination
○	$A \rightarrow \sim(A \rightarrow \sim A)$	contraposed *reductio*
○	$(A \rightarrow B) \rightarrow (\sim B \rightarrow \sim A)$	contraposition
○	$(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$	permutation
○	$(A \rightarrow B) \rightarrow ((C \rightarrow A) \rightarrow (C \rightarrow B))$	prefixing
○	$A, A \rightarrow B \Rightarrow B$	*modus ponens*

The axioms and rule here form a traditional Hilbert system. The rule *modus ponens* is written in the form using a bold arrow to echo the general definition of logical consequence in a Hilbert system. Given a set *X* of formulas, and a single formula *A*, we say that *A* can be proved from *X* (which I write "$X \Rightarrow A$") if and only if there is a proof in the Hilbert system with *A* as the conclusion, and with hypotheses from among the set *X*. A proof from hypotheses is simply a list of formulas, each of which is either a hypothesis, an axiom, or one which follows from earlier formulas in the list by means of a rule. In Orlov's system, the only rule is *modus ponens*. We will see later that this is not necessarily the most useful notion of logical consequence applicable to relevant and substructural logics. In particular, more interesting results can be proven with consequence relations which do not merely relate *sets* of formulas as premises to a conclusion, but rather relate lists, or other forms of structured collections as premises, to a conclusion. This is because lists or other structures can distinguish the *order* or *quantity* of individual premises, while sets cannot. However, this is all that can simply be done to define consequence relations within the confines of a Hilbert system, so here is where our definition of consequence will start.

These axioms and the rule do not *explicitly* represent any notion of relevance — after all, there is no "relevantly" or "is relevant to" operator. Instead, we have an axiomatic system governing the behaviour of *implication* and *negation*. The system tells us about *relevance* in virtue of what it leaves out, rather than what it includes. Neither of the following formulas are provable in Orlov's system:

$$A \rightarrow (B \rightarrow B) \qquad \sim(B \rightarrow B) \rightarrow A.$$

This distinguishes his logic from both classical and intuitionistic propositional logic.[6] If the "→" is read as either the material conditional or the conditional of intuitionistic logic, those formulas are provable. However, both of these formulas commit an obvious failure of *relevance*. The consequent of the main conditional need not have anything to do with

---

pressed in to favour as a proponent of substructural logics, as his aim was not to do without contraction, but to give an axiomatic account of material implication.

[5]The names are mine, and not Orlov's. I have attempted to give each axiom or rule its common name (see for example Anderson and Belnap's *Entailment* [1975] for a list of axioms and their names). In this case, "contraposed *reductio*" is my name, as the axiom $A \rightarrow \sim(A \rightarrow \sim A)$ is a rarely seen axiom, but it is a contraposed form of $(A \rightarrow \sim A) \rightarrow \sim A$, which is commonly known as *reductio*.

[6]Heyting's original text is still a classic introduction to intuitionistic logic, dating from this era [1956].

the antecedent. If when we say "if *A* then *B*" we mean that *B* follows *from A*, then it seems that we have *lied* when we say that "if *A* then $B \to B$", for $B \to B$ (though true enough) need not follow *from A*, if *A* has nothing to do with $B \to B$. Similarly, *A* need not follow *from* $\sim(B \to B)$ (though $\sim(B \to B)$ is false enough) for again, *A* need not have anything to do with $\sim(B \to B)$. If "following *from*" is to respect these intuitions, we need to look further afield than classical or intuitionistic propositional logic, for these logics contain those formulas as tautologies. Excising these fallacies of relevance is no straightforward job, for once they go, so must other tautologies, such as these

$$\circ \quad A \to (B \to A) \qquad \text{weakening}$$
$$\circ \quad B \to (\sim B \to A) \qquad \textit{ex contradictione quodlibet}$$

from which they can be derived.[7] To do without obvious fallacies of relevance, we must do without these formulas too. And this is exactly what Orlov's system manages to do. His system contains none of these "fallacies of relevance", and this makes his system a *relevant logic*. In Orlov's system, a formula $A \to B$ is provable only when *A* and *B* share a propositional atom. There is no way to prove a conditional in which the antecedent and the consequent have nothing to do with one another. Orlov did not prove this result in his paper. It only came to light more than 30 years later, with more recent work in relevant logic. This more recent work is applicable to Orlov's system, because Orlov has axiomatised the implication and negation fragment of the now well-known relevant logic R.

Orlov's work didn't end with the implication and negation fragment of a relevant propositional logic. He looked at the behaviour of other connectives definable in terms of conjunction and negation. In particular, he showed that defining a conjunction connective

$$A \circ B =_{\text{df}} \sim(A \to \sim B)$$

gives you a connective you can prove to be associative, commutative and square increasing[8]

$$
\begin{aligned}
(A \circ B) \circ C &\to A \circ (B \circ C) \\
A \circ (B \circ C) &\to (A \circ B) \circ C \\
A \circ B &\to B \circ A \\
A &\to A \circ A.
\end{aligned}
$$

However, the converse of the "square increasing" postulate

$$A \circ A \to A$$

<hr />

[7]Using substitution and *modus ponens*, and identity. If weakening is an axiom then $(A \to A) \to (B \to (A \to A))$ is an instance, and hence, by *modus ponens*, with $A \to A$, we get $B \to (A \to A)$.

[8]Here, and elsewhere, brackets are minimised by use of binding conventions. The general rules are simple: conditional-like connectives such as $\to$ bind less tightly than other two-place operators such as conjunction and disjunction (and fusion $\circ$ and fission $+$) which in turn bind less tightly than one place operators. So, $\sim A \vee B \to C \wedge D$ is the conditional whose antecedent is the disjunction of $\sim A$ with *B* and whose consequent is the conjunction of *C* with *D*.

is not provable, and neither are the stronger versions $A \circ B \to A$ or $B \circ A \to A$. However, for all of that, the connective Orlov defined is quite like a *conjunction*, because it satisfies the following condition:

$$\Rightarrow A \to (B \to C) \text{ if and only if } \Rightarrow A \circ B \to C.$$

You can prove a nested conditional if and only if you can prove the corresponding conditional with the two antecedents combined together as one. This is a *residuation* property.[9] It renders the connective $\circ$ with properties of *conjunction*, for it stands with the implication $\to$ in the same way that extensional conjunction and the conditional of intuitionistic or classical logic stand together.[10] Residuation properties such as these will feature a great deal in what follows.

It follows from this residuation property that $\circ$ cannot have *all* of the properties of extensional conjunction. $A \circ B \to A$ is not provable because if it were, then the weakening axiom $A \to (B \to A)$ would also be provable. $B \circ A \to A$ is not provable, because if it were, $B \to (A \to A)$ would be.

In the same vein, Orlov defined a *disjunction* connective

$$A + B =_{df} {\sim} A \to B$$

which can be proved to be associative, symmetric and square *decreasing* $(A + A \to A)$ but not square increasing. It follows that these *defined* connectives do not have the full force of the *lattice* disjunction and conjunction present in classical and intuitionistic logic. At the very first example of the study of substructural logics we are that the doorstep of one of the profound insights made clear in this area: the *splitting* of notions identified in stronger logical systems. Had Orlov noticed that one could define conjunction explicitly following the lattice definitions (as is done in intuitionistic logic, where the definitions in terms of negation and implication also fail) then he would have noticed the *split* between the *intensional* notions of conjunction and disjunction, which he defined so clearly, and the *extensional* notions which are distinct. We will see this distinction in more detail and in different contexts as we continue our story through the decades. In what follows, we will refer to $\circ$ and $+$ so much that we need to give them names. I will follow the literature of *relevant* logic and call them *fusion* and *fission*.

Good ideas have a feature of being independently discovered and rediscovered. The logic R is no different. Moh [1950] and Church [1951b], independently formulated the implication fragment of R in the early 1950's. Moh formulated an axiom system

$\circ$	$A \to A$	identity
$\circ$	$(A \to (A \to B)) \to (A \to B)$	contraction
$\circ$	$A \to ((A \to B) \to B)$	assertion
$\circ$	$(A \to B) \to ((B \to C) \to (A \to C))$	suffixing.

---

[9] It ought to remind you of simple arithmetic results: $x = z/y$ if and only if $x \times y = z$; $x = z - y$ if and only if $x + y = z$.

[10] Namely, that $A \wedge B \supset C$ is provable if and only if $A \supset (B \supset C)$.

Whereas Church's axiom system replaces the *assertion* and *suffixing* with *permutation* and *prefixing*

○ $(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$ permutation
○ $(A \rightarrow B) \rightarrow ((C \rightarrow A) \rightarrow (C \rightarrow B))$ prefixing.

Showing that these two axiomatisations are equivalent is an enjoyable (but lengthy) exercise in axiom chopping. It is a well-known result that in the presence of either prefixing or suffixing, permutation is equivalent to assertion. Similarly, in the presence of either permutation or assertion, prefixing is equivalent to suffixing. (These facts will be more perspicuous when we show how the presence of these axioms correspond to particular structural rules. But this is to get ahead of the story by a number of decades.)

Note that each of the axioms in either Church's or Moh's presentation of R are tautologies of intuitionistic logic. Orlov's logic of relevant implication extends intuitionistic logic when it comes to *negation* (as double negation elimination is present), but when it comes to implication alone, the logic R is *weaker* than intuitionistic logic. As a corollary, Peirce's law

○ $((A \rightarrow B) \rightarrow A) \rightarrow A$ Peirce's law

is not provable in R, even though it is a classical tautology. The fallacies of relevance are examples of intuitionistic tautologies which are not present in relevant logic. Nothing so far has shown us that adding negation *conservatively extends* the implication fragment of R (in the sense that there is no implicational formula which can be proved with negation which cannot also be proved without it). However, as we will see later, this is indeed the case. Adding negation does not lead to new implicational theorems.

Church's work on his *weak implication* system closely paralleled his work on the lambda calculus. (As we will see later, the tautologies of this system are exactly the types of the terms in his λI calculus.)[11] Church's work extends that of Orlov by proving a *deduction theorem*. Church showed that if there is a proof with hypotheses A_1 to A_n with conclusion B, then there is either a proof of B from hypotheses A_1 to A_{n-1} (in which case A_n was irrelevant as a hypothesis) or there is a proof of $A_n \rightarrow B$ from $A_1, \ldots, A_{n-1}$.

FACT 1 (Church's Deduction Theorem). In the implicational fragment of the relevant logic R, if $A_1, \ldots, A_n \Rightarrow B$ can be proved in the Hilbert system then at least one of the following two consequences can also be proved in that system.

○ $A_1, \ldots, A_{n-1} \Rightarrow B$,
○ $A_1, \ldots, A_{n-1} \Rightarrow A_n \rightarrow B$.

Proof. The proof follows the traditional proof of the Deduction Theorem for the implicational fragment of either classical or intuitionistic logic. A proof for $A_1, \ldots, A_n \Rightarrow B$ is transformed into a proof for $A_1, \ldots, A_{n-1} \Rightarrow A_n \rightarrow B$ by prefixing each step of the

---

[11]In which λx can abstract a variable from only those terms *in which the variable x occurs*. As a result, the λ-term $\lambda x.\lambda y.x$, of type $A \rightarrow (B \rightarrow A)$, is a term of the traditional λ-calculus, but not of the λI calculus.

proof by "$A_n \rightarrow$". The weakening axiom $A \rightarrow (B \rightarrow A)$ is needed in the traditional result for the step showing that if a hypothesis is not used in the proof, it can be introduced as an antecedent anyway. Weakening is not present in R, and this step is not needed in the proof of Church's result, because he allows a special clause, exempting us from proving $A_n \rightarrow B$ when A_n is not actually used in the proof.   ■

We will see other forms of the deduction theorem later on in our story. This deduction theorem lays some claim to helping explain the way in which the logic R can be said to be *relevant*. The conditional of R respects *use* in proof. To say that $A \rightarrow B$ is true is to say not merely that B is true whenever A is true (keeping open the option that A might have nothing to do with B): To say that $A \rightarrow B$ is true is to say that B follows *from A*. This is not the only kind of deduction theorem applicable to relevant logics. In fact, it is probably not the most satisfactory one, as it fails once the logic is extended to include extensional conjunction. After all, we would like $A, B \Rightarrow A \wedge B$ but we can have neither $A \Rightarrow B \rightarrow A \wedge B$ (since that would give the fallacy of relevance $A \Rightarrow B \rightarrow A$, in the presence of $A \wedge B \rightarrow A$) nor $A \Rightarrow A \wedge B$ (which is classically invalid, and so, relevantly invalid). So, another characterisation of relevance must be found in the presence of conjunction. In just the same way, combining conjunction-like pairing operations in the λI calculus has proved quite difficult [Pottinger, 1979]. Avron has argued that this difficulty should make us conclude that relevance and extensional connectives cannot live together [1986; 1992].

Meredith and Prior were also aware of the possibility of looking for logics weaker than classical propositional logic, and that different axioms corresponded to different principles of the λ-calculus (or in Meredith and Prior's case, combinatory logic). Following on from work of Curry and Feys [1958; 1972], they formalised subsystems of classical logic including what they called BCK (logic without contraction) and BCI (logic without contraction or weakening: which is now known as *linear logic*) [Meredith and Prior, 1963]. They, with Curry, are the first to explicitly chart the correspondence of propositional axioms with the behaviour of combinators which allow the rearrangement of premises or antecedents.[12]

For a number of years following their pioneering work, Anderson and Belnap continued in this vein, using techniques from other branches of proof theory to explain how the logic R and its cousins respected conditions of relevance and necessity. We will shift our attention now to another of the precursors of Anderson and Belnap's work, one which pays attention to conditions of *necessity* as well as *relevance*.

2.2   Entailment: Ackermann

Ackermann formulated a logic of *entailment* in the late 1950s [1956]. He extended C. I. Lewis' work on systems of entailment to respect relevance and to avoid the paradoxes of strict implication. Ackermann's favoured system of entailment is a weakening

---

[12]It is in their honour that I use Curry's original terminology for the structural rules we will see later: W for contraction, K for weakening, C for commutativity, etc.

of the system S4 of strict implication designed to avoid the paradoxes. Unlike earlier work on relevant implication, Ackermann's system includes the full complement of sentential connectives.

To motivate the departures that Ackermann's system takes from R, note that the arrow of R cannot be used to model *entailment*. If we want to say that A entails that B, the arrow of R is significantly too strong. Specifically, axioms such as *permutation* and *assertion* must be rejected for the arrow of entailment. To take an example, suppose that A is contingently true. It is an instance of assertion that

$$A \rightarrow ((A \rightarrow A) \rightarrow A).$$

However, even if A is true, it ought not be true that $A \rightarrow A$ *entails* A. For $A \rightarrow A$ is presumably *necessarily* true. We cannot not have this necessity transferring to the contingent claim A.[13] *Permutation* must go too, as assertion follows from permuting the identity $(A \rightarrow B) \rightarrow (A \rightarrow B)$. So, a logic of entailment must be weaker than R. However, it need not be *too* much weaker. It is clear that prefixing, suffixing and contraction are not prone to any sort of counterexample along these lines: they can survive into a logic of entailment.

Ackermann's original paper features two different presentations of the system of entailment. The first, Σ', is an ingenious consecution calculus, which is unlike any proof theory which has survived into common use, so unfortunately, I must skim over it here in one paragraph.[14] The system manipulates consecutions of the form $A, B \vdash C$ (to be understood as $A \wedge B \rightarrow C$) and $A^*, B \vdash C$ (to be understood as as $A \rightarrow (B \rightarrow C)$). Note that the comma in the antecedent place has no uniform interpretation: In effect, there are two different premise combining operations. This is, in embryonic form at least, the first explicit case of a dual treatment of both intensional and extensional conjunction in a proof theory that I have found.

Ackermann's other presentation of the logic of entailment is a Hilbert system. The axioms and rules are presented in Figure 1. You can see that many of the axioms considered have already occurred in the study of relevant implication. The innovations appear in both what is omitted (assertion and permutation, as we have seen) and in the full complement of rules for conjunction and disjunction.[15]

To make up for the absence of assertion and permutation, Ackermann adds *restricted permutation*. This rule is not a *permutation* rule (it doesn't permute anything) but it is a *restriction* of the permutation rule to infer $B \rightarrow (A \rightarrow C)$ from $A \rightarrow (B \rightarrow C)$. For the restricted rule we conclude $A \rightarrow C$ from $A \rightarrow (B \rightarrow C)$ and B. Clearly this follows from permutation. This restriction allows a restricted form of *assertion* too.

$$\circ \quad (A \rightarrow A') \rightarrow (((A \rightarrow A') \rightarrow B) \rightarrow B) \quad \text{restricted assertion}$$

---

[13]If something is *entailed* by a necessity, it too is necessary. If A entails B then if we cannot have A false, we cannot have B false either.

[14]The interested reader is referred to Ackermann's paper (in German) [1956] or to Anderson, Belnap and Dunn's sympathetic summary [1992, §44–46] (in English).

[15]The choice of *counterexample* as a thesis connecting implication and negation in place of *reductio* (as in Orlov [1928]) is of no matter. The two are equivalent in the presence of contraposition and double negation rules. Showing this is a gentle exercise in axiom-chopping.

<div align="center">Axioms</div>

○	$A \rightarrow A$	identity
○	$(A \rightarrow B) \rightarrow ((C \rightarrow B) \rightarrow (C \rightarrow A))$	prefixing
○	$(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$	suffixing
○	$(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$	contraction
○	$A \wedge B \rightarrow A, A \wedge B \rightarrow B$	conjunction elimination
○	$(A \rightarrow B) \wedge (A \rightarrow C) \rightarrow (A \rightarrow B \wedge C)$	conjunction introduction
○	$A \rightarrow A \vee B, B \rightarrow A \vee B$	disjunction introduction
○	$(A \rightarrow C) \wedge (B \rightarrow C) \rightarrow (A \vee B \rightarrow C)$	disjunction elimination
○	$A \wedge (B \vee C) \rightarrow B \vee (A \wedge C)$	distribution
○	$(A \rightarrow B) \rightarrow (\sim B \rightarrow \sim A)$	contraposition
○	$A \wedge \sim B \rightarrow \sim (A \rightarrow B)$	counterexample
○	$A \rightarrow \sim\sim A$	double negation introduction
○	$\sim\sim A \rightarrow A$	double negation elimination

<div align="center">Rules</div>

(α)	$A, A \rightarrow B \Rightarrow B$	*modus ponens*
(β)	$A, B \Rightarrow A \wedge B$	adjunction
(γ)	$A, \sim A \vee B \Rightarrow B$	disjunctive syllogism
(δ)	$A \rightarrow (B \rightarrow C), B \Rightarrow A \rightarrow C$	restricted permutation rule

<div align="center">Figure 1. Ackermann's axiomatisation Π'</div>

This is an instance of the assertion where the first position A is replaced by the entailment $A \rightarrow A'$. While assertion might not be valid for the logic of entailment, it is valid when the proposition in the first position is itself an entailment.

As Anderson and Belnap point out [1992, §8.2], (δ) is not a particularly satisfactory rule. Its status is akin to that of the rule of necessitation in modal logic (from $\Rightarrow A$ to infer $\Rightarrow \Box A$). It does not extend to an *entailment* ($A \rightarrow \Box A$). If it is possible to do without a rule like this, it seems preferable, as it licences transitions in proofs which do not correspond to valid entailments. Anderson and Belnap showed that you can indeed do without (δ) to no ill effect. The system is unchanged when you replace *restricted permutation* by *restricted assertion*.

This is not the only rule of Ackermann's entailment which provokes comment. The rule (γ) (called *disjunctive syllogism*) has had more than its fair share of ink spilled. It suffers the same failing in this system of entailment as does (δ): it does not correspond to a valid entailment. The corresponding entailment $A \wedge (\sim A \vee B) \rightarrow B$ is not provable. I will defer its discussion to Section 2.4, by which time I will be able to prove *theorems* about disjunctive syllogism as well as arguing about its significance.

Ackermann's remaining innovations with this system are at least twofold. First, we have a thorough treatment of extensional disjunction and conjunction. Ackermann noticed that you need to add distribution of conjunction over disjunction as a separate axiom.[16] The conjunction and disjunction elimination and introduction rules are sufficient to show that conjunction and disjunction are lattice join and meet on propositions ordered by provable entailment. (It is a useful exercise to show that in this system of entailment, you can prove $A \vee \sim A$, $\sim(A \wedge \sim A)$, and that all De Morgan laws connecting negation, conjunction and disjunction hold.)

The second innovation is the treatment of *modality*. Ackermann notes that as in other systems of modal logic which take entailment as primary, it is possible to define the one-place modal operators of necessity, possibility and others in terms of entailment. A traditional choice is to take *impossibility* "U"[17] defined by setting UA to be $A \rightarrow B \wedge \sim B$ for some choice of a contradiction. Clearly this will not do in the case of a *relevant* logic as even though it makes sense to say that if A entails the contradictory $B \wedge \sim B$ then A is impossible, we might have A entailing *some* contradiction (and so, being impossible) without entailing *that* contradiction. It is a fallacy of relevance to take all contradictions to be provably equivalent. Instead, Ackermann takes another tack, by introducing a new constant f, with some special properties.[18] The intent is to take f to mean "some contradiction is true". Ackermann then posits the following axioms and rules.

$$
\begin{array}{ll}
\circ & A \wedge \sim A \rightarrow f \\
\circ & (A \rightarrow f) \rightarrow \sim A \\
(\epsilon) & A \rightarrow B, (A \rightarrow B) \wedge C \rightarrow f \Rightarrow C \rightarrow f
\end{array}
$$

Clearly the first two are true, if we interpret f as the disjunction of all contradictions. The last we will not tarry with. It is an idiosyncratic rule, distinctive to Ackermann. More important for our concern is the definition of f. It is a new constant, with new properties which open up once we enter the substructural context. Classically (or intuitionistically) f would behave as $\perp$, a proposition which entails all others. In a substructural logic like R or Ackermann's entailment, f does no such thing. It is true that f is provably false (we can prove $\sim f$, from the axiom $(f \rightarrow f) \rightarrow \sim f$) but it does not follow that f entails *everything*. Again, a classical notion *splits*: there are two different kinds of *falsehood*. There is the *Ackermann* false constant f, which is the *weakest* provably false proposition, and there is the *Church* false constant $\perp$, which is the *falsest* false proposition, which entails every proposition whatsoever. Classically and intuitionistically, both are equivalent. Here, they come apart.

The two false constants are mirrored by their negations: two *true* constants. The Ackermann true constant t (which is $\sim f$) is the conjunction of all tautologies. The Church true constant $\top$ (which is $\sim\perp$) is the weakest proposition of all, such that $A \rightarrow \top$ is true for each A. If we are to define *necessity* by means of a propositional constant, then $t \rightarrow A$

---

[16]If we have the residuation of conjunction by $\supset$ (intuitionistic or classical material implication) then distribution follows. The algebraic analogue of this result is the thesis that a residuated lattice is distributive.

[17]For *unmöglich*.

[18]Actually, Ackermann uses the symbol "Λ", but it now appears in the literature as "f".

is the appropriate choice. For $t \to A$ will be provable whenever A is provable. Choosing $\top \to A$ would be much too restrictive, as we would only allow as "necessary" propositions which were entailed by all others. Since we do not have $A \vee \sim A \to B \vee \sim B$, if we want *both* to be necessary, we must be happy with the weaker condition, of being entailed by t.

This choice of true constant to define necessity motivates the choice that Anderson and Belnap used. t must entail each proposition of the form $A \to A$ (as each is a tautology). Anderson and Belnap showed that $t \to A$ in Ackermann's system is equivalent to $(A \to A) \to A$, and so they use $(A \to A) \to A$ as a definition of $\Box A$, and in this way, they showed that it was possible to define the one-place modal operators in the original language alone, without the use of propositional constants at all.[19] It is instructive to work out the details of the behaviour of $\Box$ as we have defined it. Necessity here has properties roughly of S4. In particular, you can prove $\Box A \to \Box\Box A$ but not $\Diamond A \to \Box\Diamond A$ in Ackermann's system.[20] (You will note that using this definition of necessity and without (δ) you need to add an axiom to the effect that $\Box A \wedge \Box B \to \Box(A \wedge B)$,[21] as it cannot be proved from the system as it stands. Defining $\Box A$ as $t \to A$ does not have this problem.)

2.3   Anderson and Belnap

We have well-and-truly reached beyond Ackermann's work on entailment to that of Alan Anderson and Nuel Belnap. Anderson and Belnap started their exploration of relevance and entailment with Ackermann's work [1959; 1962], but very soon it became an independent enterprise with a wealth of innovations and techniques from their own hands, and from their student, colleagues and collaborators (chiefly J. Michael Dunn, Robert K. Meyer, Alasdair Urquhart, Richard Routley (later known as Richard Sylvan) and Kit Fine). Much of this research is reported in the two-volume *Entailment* [Anderson and Belnap, 1975; 1992], and in the papers cited therein. There is no way that I can adequately summarise this work in a few pages. However, I can sketch what I take to be some of the most important and enduring themes of this tradition.

Fitch Systems

Hilbert systems are not the only way to present proofs. Other proof theories give us us different insights into a logical system by isolating rules relevant to each different connective. Hilbert systems, with many axioms and few rules, are not so suited to a project of understanding the *internal structure* of a family of logical systems. It is no surprise that in the relevant logic tradition, a great deal of work was invested toward providing different proof theories which model directly the relationship between premises and conclusions.

The first natural deduction system for R and E (Anderson and Belnap's system of *entailment*) was inspired by Fitch's natural deduction system, in widespread use in undergraduate and postgraduate logic instruction in the 1950s in the United States [Fitch,

---

[19]Impossibility UA is then $A \to \sim(A \to A)$.

[20]Defining $\Diamond A$ as $\sim\Box\sim A$, i. e., as $\sim((\sim A \to \sim A) \to \sim A)$.

[21]The axiom is ungainly when it is written out in full: $((A \to A) \to A) \wedge ((B \to B) \to B) \to ((A \wedge B \to A \wedge B) \to A \wedge B)$.

1952].[22] A Fitch system is a *linear* presentation of a natural deduction proof, with introduction and elimination rules for each connective, and the use of vertical lines to present subproofs — parts of proofs under hypotheses. Here, for example, is a proof of the relevantly unacceptable *weakening* axiom in a Fitch system for classical (or intuitionistic) logic:

$$
\begin{array}{lll}
1 & A & \text{hyp} \\
2 & \quad B & \text{hyp} \\
3 & \quad A & 1 \text{ reit} \\
4 & B \to A & 2\text{--}3 \to\text{I} \\
5 & A \to (B \to A) & 1\text{--}4 \to\text{I}
\end{array}
$$

Each line is numbered to the left, and the annotation to the right indicates the provenance of each formula. A line marked with "hyp" is a hypothesis, and its introduction increases the level of nesting of the proof. In line 4 we have the application of *conditional proof*, or as it is indicated here, *implication introduction* ($\to$I). Since A has been proved under the hypothesis of B, we deduce $B \to A$, discharging that hypothesis. The other distinctive feature of Fitch proofs is the necessity to *reiterate* formulas. If a formula appears outside a nested subproof, it is possible to reiterate it under the assumption, for use inside the subproof.

Now, this proof is defective, if we take $\to$ to indicate relevant implication. There are two possible points of disagreement. One is to question the proof at the point of line 3: perhaps something has gone wrong at the point of *reiterating A* in the subproof. This is not where Anderson and Belnap modify Fitch's system in order to model R.[23] As you can see in the proof of the (relevantly acceptable) *assertion* axiom, reiteration of a formula from outside a subproof is unproblematic.

$$
\begin{array}{lll}
1 & A & \text{hyp} \\
2 & \quad A \to B & \text{hyp} \\
3 & \quad A & 1 \text{ reit} \\
4 & \quad B & 2\text{--}3 \to\text{E} \\
5 & (A \to B) \to B & 2\text{--}4 \to\text{I} \\
6 & A \to ((A \to B) \to B) & 1\text{--}5 \to\text{I}
\end{array}
$$

The difference between the two proofs indicates what has gone wrong in the proof of the weakening axiom. In this proof, we have indeed *used $A \to B$* in the proof of B from lines 1 to 4. In the earlier "proof", we indeed proved A *under* the assumption of B but

---

[22]That Fitch systems would be used by Anderson and Belnap is to be expected. It is also to be expected that Read [1988] and Slaney [1990] (from the U. K.) use Lemmon-style natural deduction [1965], modelled after Lemmon's textbook, used in the U. K. for many years. Logicians on continental Europe are much more likely to use Prawitz [1965] or Gentzen-style [1934; 1969] natural deduction systems. This geographic distribution of pedagogical techniques (and its resulting influence on the way *research* is directed, as well as teaching) is remarkably resilient across the decades. The recent publication of Barwise and Etchemendy's popular textbook introducing logic still uses a Fitch system [2000]. As far as I am aware, instruction in logic in none of the major centres in Europe or Australia centres on Fitch-style presentation of natural deduction.

[23]Restricting reiteration is the way to give hypothesis generation and conditional introduction *modal* force, as we shall see soon.

we did not *use B* in that proof. The implication introduction in line 4 is fallacious. If I am to pay attention to *use* in proof, I must keep track of it in some way. Anderson and Belnap's innovation is to add *labels* to formulas in proofs. The label is a set of indices, indicating the hypotheses upon which the formula depends. If I introduce a hypothesis A in a proof, I add a new label, a singleton of a new index standing for that hypothesis. The implication introduction and elimination rules must be amended to take account of labels. For implication elimination, given A_a and $A \to B_b$, I conclude $B_{a \cup b}$, for this instance of B in the proof depends upon everything we needed for A and for $A \to B$. For implication introduction, given a proof of B_a under the hypothesis $A_{\{i\}}$, I can conclude $A \to B_{a - \{i\}}$, *provided that $i \in a$*. With these amended rules, we can annotate the original proof of *assertion* with labels, as follows.

$$
\begin{array}{lll}
1 & A_{\{1\}} & \text{hyp} \\
2 & \quad A \to B_{\{2\}} & \text{hyp} \\
3 & \quad A_{\{1\}} & \text{1 reit} \\
4 & \quad B_{\{1,2\}} & \text{2–3 } \to\text{E} \\
5 & (A \to B) \to B_{\{1\}} & \text{2– 4 } \to\text{I} \\
6 & A \to ((A \to B) \to B) & \text{1–5 } \to\text{I}
\end{array}
$$

The proof of *weakening*, on the other hand, cannot be annotated with labels satisfying the rules for implication.

$$
\begin{array}{lll}
1 & A_{\{1\}} & \text{hyp} \\
2 & \quad B_{\{2\}} & \text{hyp} \\
3 & \quad A_{\{1\}} & \text{1 reit} \\
4 & B \to A_{???} & \text{2–3 } \to\text{I} \\
5 & A \to (B \to A)_{???} & \text{1–4 } \to\text{I}
\end{array}
$$

Modifying the system to model entailment is straightforward. As I hinted earlier, if the arrow has a modal force, we do not want unrestricted reiteration. Instead of allowing an arbitrary formula to be reiterated into a subproof, since entertaining a hypothesis now has the force of considering an alternate possibility, we must only allow for reiteration formulas which might indeed hold in those alternate possibilities. Here, the requisite formulas are entailments. Entailments are not only true, but true of necessity, and so, we can reiterate an entailment under the context of a hypothesis, but we cannot reiterate atomic formulas. So, the proof above of *assertion* breaks down at the point at which we wished to reiterate A into the second subproof. The proof of *restricted assertion* will succeed.

$$
\begin{array}{lll}
1 & A \to A'_{\{1\}} & \text{hyp} \\
2 & \quad (A \to A') \to B_{\{2\}} & \text{hyp} \\
3 & \quad A \to A'_{\{1\}} & \text{1 reit} \\
4 & \quad B_{\{1,2\}} & \text{2–3 } \to\text{E} \\
5 & ((A \to A') \to B) \to B_{\{1\}} & \text{2–4 } \to\text{I} \\
6 & (A \to A') \to (((A \to A') \to B) \to B) & \text{1–5 } \to\text{I}
\end{array}
$$

This is a permissible proof because we are entitled to reiterate $A \to A'$ at line 3. Even given the assumption that $(A \to A') \to B$, the prior assumption of $A \to A'$ holds in the new context.

Here is a slightly more complex proof in this Fitch system for entailment. (Recall that $\Box A$ is shorthand for $(A \to A) \to A$, for Anderson and Belnap's system of entailment.) This proof shows that in E, the truth of an entailment (here $B \to C$) entails that anything *entailed* by that entailment (here A) is itself necessary too. The reiterations on lines 4 and 5 are permissible, because $B \to C$ and $(B \to C) \to A$ are both entailments.

$$
\begin{array}{llll}
1 & B \to C_{\{1\}} & & \text{hyp} \\
2 & \quad (B \to C) \to A_{\{2\}} & & \text{hyp} \\
3 & \quad\quad A \to A_{\{3\}} & & \text{hyp} \\
4 & \quad\quad B \to C_{\{1\}} & & 1 \text{ reit} \\
5 & \quad\quad (B \to C) \to A_{\{2\}} & & 2 \text{ reit} \\
6 & \quad\quad A_{\{1,2\}} & & 4,5 \to\text{E} \\
7 & \quad\quad A_{\{1,2,3\}} & & 3,6 \to\text{E} \\
8 & \quad (A \to A) \to A_{\{1,2\}} & & 3\text{--}7 \to\text{I} \\
9 & ((B \to C) \to A) \to \Box A_{\{1\}} & & 2\text{--}8 \to\text{I} \\
10 & (B \to C) \to (((B \to C) \to A) \to \Box A) & & 1\text{--}9 \to\text{I}
\end{array}
$$

We say that A follows *relevantly* from B when a proof with hypothesis $A_{\{i\}}$ concludes in $B_{\{i\}}$. This is written "$A \vdash B$". We say that A is provable by itself when there is a proof of A with no label at all. Then the Hilbert system and the natural deduction system match in the following two senses.

FACT 2 (Hilbert and Fitch Equivalence). $\Rightarrow A \to B$ if and only if $A \vdash B$. $\Rightarrow A$ if and only if $\vdash A$.

(Note that this fact claims equivalence only between the Hilbert and Fitch systems for the implicational fragment of R and not any fuller language, as we have considered natural deduction rules for implication only, thus far.)

Proof. The proof is by an induction on the complexity of proofs in both directions. To convert a Fitch proof to a Hilbert proof, we replace the hypotheses $A_{\{i\}}$ by the identity $A \to A$, and the arbitrary formula $B_{\{i_1, i_2, \ldots, i_n\}}$ by $A_1 \to (A_2 \to \cdots \to (A_n \to B)\cdots)$ (where A_j is the formula introduced with label A_j). Then you show that the steps between *these* formulas can be justified in the Hilbert system. Conversely, you simply need to show that each Hilbert axiom is provable in the Fitch system, and that *modus ponens* preserves provability. Neither proof is difficult. ∎

Other restrictions on reiteration can be made in this Fitch system in order to model weaker logics. In particular, Anderson and Belnap examine a system T of *ticket* entailment, with the underlying idea that statements of the form $A \to B$ are *rules* but not *facts*. They are to be used as major premises of implication eliminations, but not as minor premises. The restriction on reiteration to get this effect allows you to conclude $B_{a \cup b}$ from A_a and $A \to B_b$, *provided that* $\max(b) \leqslant \max(a)$. The effect of this is to render restricted assertion unprovable, while *identity*, *prefixing*, *suffixing* and *contraction* remain provable

(and these axiomatise the calculus T of ticket entailment).[24] (It is an open problem to this day whether the implicational fragment of T is decidable.)

Before considering the extension of this proof theory to deal with the extensional connectives, let me note one curious result in the vicinity of T. The logic TW you get by removing *contraction* from T has a surprising property. Errol Martin has shown that if $A \to B$ and $B \to A$ are provable in TW, then A and B must be the same formula [Martin and Meyer, 1982].[25]

First Degree Entailment

It is one thing to provide a proof theory for implication or entailment. It is another to combine it with a theory of the other propositional connectives: conjunction, disjunction and negation. Anderson and Belnap's strategy was to first decide the behaviour of conjunction, disjunction and negation, and then combine this theory with the theory of entailment or implication. This gives the structure of the first volume of *Entailment* [Anderson and Belnap, 1975]. The first 100 pages deals with implication alone, the next 50 with implication and negation, the next 80 with the *first degree* fragment (entailments between formulas not including implication) and only at page 231 do we find the formulation of the full system E of entailment.

Anderson and Belnap's work on entailments between truth functions (or what they call *first degree entailment*) dates back to a paper in [1962]. There are many different ways to carve out first degree entailments which are *relevant* from those which are not. For example, *filter* techniques due to von Wright [1957], Lewy [1958], Geach [1958] and Smiley [1959] tell us that statements like

$$A \to B \vee \sim B \qquad A \wedge \sim A \to B$$

fail as entailments because there is no atom shared between antecedent and consequent. So far, so good, and their account follows Anderson and Belnap's. However, if this is the *only* criterion to add to classical entailment, we allow their analogues

$$A \to A \wedge (B \vee \sim B) \qquad (A \wedge \sim A) \vee B \to B$$

for the propositional atom A is shared in the first case, and B in the second. Since both of the following classical entailments

$$A \wedge (B \vee \sim B) \to B \vee \sim B \qquad A \wedge \sim A \to (A \wedge \sim A) \vee B$$

---

[24]This is as good a place as any to note that the axiom of *self distribution* $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$ will do instead of *contraction* in any of these axiomatisations.

[25]Martin's proof proceeds via a result showing that the logic given by *prefixing* and *suffixing* (without identity) has *no* instances of identity provable at all. This is required, for $(A \to B) \to ((B \to A) \to (A \to A))$ is an instance of *suffixing*. The system S (for *syllogism*) has interesting properties in its own right, modelling noncircular (non "question begging") logic [Martin, 1984].

also satisfy the atom-sharing requirement, using variable sharing as the *only* criterion makes us reject the transitivity of entailment. After all, given $A \to A \wedge (B \vee \sim B)$ and given $A \wedge (B \vee \sim B) \to B \vee \sim B$, if $\to$ is transitive, we get $A \to B \vee \sim B$.[26]

Anderson and Belnap respond by noting that if $A \to B \vee \sim B$ is problematic because of relevance, then $A \to A \wedge (B \vee \sim B)$ is at least 50% problematic [Anderson and Belnap, 1975, p. 155]. Putting things another way, if to say that A entails $B \wedge C$ is at least to say that A entails B and that A entails C, then we cannot just add a blanket atom-sharing criterion to filter out failures of relevance, for it might apply to one conjunct and not the other. Filter techniques do not work.

Anderson and Belnap characterise valid first degree entailments in a number of ways. The simplest way which does not use any *model* theory is a normal form theorem for first degree entailments. We will use a process of *reduction* to transform arbitrary entailments into *primitive* entailments, which we can determine on sight. The first part of the process is to drive negations inside other operators, leaving them only on atoms. We use the De Morgan equivalences and the double negation equivalence to do this.[27]

$$\sim(A \vee B) \leftrightarrow \sim A \wedge \sim B \qquad \sim(A \wedge B) \leftrightarrow \sim A \vee \sim B \qquad \sim\sim A \leftrightarrow A$$

(I write "$A \leftrightarrow B$" here a shorthand for "both $A \to B$ and $B \to A$".)

The next process involves pushing conjunctions and disjunctions around. The aim is to make the antecedent of our putative entailment a disjunction of conjunctions, and the consequent a conjunction of disjunctions. We use these distribution facts to this effect.[28]

$$(A \vee B) \wedge C \leftrightarrow (A \wedge C) \vee (B \wedge C) \qquad (A \wedge B) \vee C \leftrightarrow (A \vee C) \wedge (B \vee C)$$

With that transformation done, we break the entailment up into primitive entailments in these two kinds of steps:

$$A \vee B \to C \text{ if and only if } A \to C \text{ and } B \to C$$

$$A \to B \wedge C \text{ if and only if } A \to B \text{ and } A \to C.$$

All of the transformation rules in this process are intended to be unproblematically valid when it comes to relevant entailment. The first batch (the negation conditions) seem unproblematic if negation is truth functional. The second batch (the distribution conditions, together with the associativity, commutativity and *idempotence* of both disjunction and

---

[26]Nontransitive accounts of entailment have survived to this day, with more sophistication. Neil Tennant has an idiosyncratic approach to normalisation in logics, arguing for a "relevant logic" which differs from our substructural logics by allowing the validity of $A \vee B, \sim A \vdash B$ and $A \vdash A \vee B$, while rejecting $A, \sim A \vdash B$ [1992; 1994]. Tennant's system rejects the unrestricted transitivity of proofs: the 'Cut' which would allow $A, \sim A \vdash B$ from the proofs of $A \vee B, \sim A \vdash B$ and $A \vdash A \vee B$ is not admissible. Tennant uses normalisation to motivate this system.

[27]We also lean on the fact that we can replace provable equivalents *ad libitum* in formulas. Formally, if we can prove $A \to B$ and $B \to A$ then we can prove $C \to C'$ and $C' \to C$, where C' results from C by changing as many instances of A to B in C as you please. All substructural logics satisfy this condition.

[28]Together with the associativity, commutativity and *idempotence* of both disjunction and conjunction, which I will not bother to write out formally.

conjunction) are sometimes questioned[29] but we have been given no reason *yet* to quibble with these as relevant entailments. Finally, the steps to break down entailments from disjunctions and entailments to conjunctions are fundamental to the behaviour of conjunction and disjunction as lattice connectives. They are also fundamental to the inferential properties of these connectives. $A \vee B$ licences an inference to C (and a *relevant* one, presumably!) if and only if A and B both licence that inference. $B \wedge C$ follows from A (and *relevantly* presumably!) if and only if B and C both follow from A.

The result of the completed transformation will then be a collection of primitive entailments: each of which is a conjunction of atoms and negated atoms in the antecedent, and a disjunction of atoms and negated atoms in the consequent. Here are some examples of primitive entailments:

$$p \wedge \sim p \to q \vee \sim q \qquad p \to p \vee \sim p \qquad p \wedge \sim p \wedge \sim q \wedge r \to s \vee \sim s \vee q \vee \sim r.$$

Anderson and Belnap's criterion for deciding a primitive entailment is simple. A primitive entailment $A \to B$ is valid if and only if one of the conjuncts in the antecedent also features as a disjunct in the consequent. If there is such an atom, clearly the consequent follows *from* the antecedent. If there is no such atom, the consequent may well be true (and perhaps even necessarily so, if an atom and its negation both appear as disjuncts) but its truth does not follow from the truth of the antecedent. This makes some kind of sense: what is it for the consequent to be true? It's for at least one of B_1, B_2, B_3 ... to be true. (And that's all, as that's all that the consequent says.) If none of these things are given by the antecedent, then the consequent as a whole doesn't follow from the antecedent either.[30]

We can then decide an *arbitrary* first degree entailment by this reduction process. Given an entailment, reduce it to a collection of primitive entailments, and then the original entailment is valid if and only if *each* of the primitive entailments is valid. Let's apply this to the inference of disjunctive syllogism: $(A \vee B) \wedge \sim A \to B$. Distributing the disjunction over the conjunction in the antecedent, we get $(A \wedge \sim A) \vee (B \wedge \sim A) \to B$. This is a valid entailment if and only if $A \wedge \sim A \to B$ and $B \wedge \sim A \to B$ both are. The second is, but the first is not. Disjunctive syllogism is therefore rejected by Anderson and Belnap. To accept it as a valid entailment is to accept $A \wedge \sim A \to B$ as valid. Since this is a fallacy of relevance, so is disjunctive syllogism.

This is one simple characterisation of first degree entailments. Once we start looking at *models* we will see some different models for first degree entailment which give us other straightforward characterisations of the first-degree fragment of R and E. Now, however,

---

[29]We will see later that linear logic rejects the distribution of conjunction over disjunction.

[30]I am not here applying the fallacious condition that $B_1 \vee B_2$ follows from A if and only if B_1 follows from A or B_2 follows from A, which is invalid in general. Let A be $B_1 \vee B_2$, for example. But in *that* case we note that B_1 follows from some disjunct of A and B_2 also follows from other disjunct of A. In the atomic case, A can no longer be split up.

To demonstrate the entailment $p \to q \vee \sim q$ *classically*, the idea would be to import the *tautologous* $q \vee \sim q$ into the antecedent, to get $p \wedge (q \vee \sim q) \to q \vee \sim q$, distribute to get $(p \wedge q) \vee (p \wedge \sim q) \to q \vee \sim q$, and split to get both $p \wedge q \to q \vee \sim q$ (which is valid, by means of q) and $p \wedge \sim q \to q \vee \sim q$ (which is valid, by means of $\sim q$). With eyes of *relevance* there's no reason to see the appeal for importing $q \vee \sim q$ in the first place.

we must consider how to graft this account together with the account of implicational logics we have already seen.

Putting Them Together

To add the truth functional connectives to a Hilbert system for R or E, Anderson and Belnap used the axioms due to Ackermann for his system Π'. The *conjunction introduction* and *elimination*, *disjunction introduction* and *elimination* axioms, together with *distribution* and the rule of *adjunction* is sufficient to add the distributive lattice connectives. To add negation, you add the *double negation* axioms and *contraposition*, and *counterexample* (or equivalently, *reductio*). Adding the truth functions to a Hilbert system is straightforward.

It is more interesting to see how to add the connectives to the natural deduction system, because these systems usually afford a degree of separation between different connectives, and they provide a context in which you can see the distinctive behaviour of those connectives. Let's start with negation. Here are the negation rules proposed by Anderson and Belnap:

○ ($\sim I$) From $\sim A_a$ proved under the hypothesis $A_{\{k\}}$, deduce $\sim A_{a-\{k\}}$ (if $k \in a$). (This discharges the hypothesis.)

○ (Contraposition) From B_a and $\sim B_b$ proved under the hypothesis $A_{\{k\}}$, deduce $\sim A_{a \cup b - \{k\}}$ (if $k \in b$). (This discharges the hypothesis.)

○ ($\sim\sim E$) From $\sim\sim A_a$ to deduce A_a.

These rules follow directly the axioms of *reductio*, *contraposition* and *double negation elimination*. They are sufficient to derive all of the desired negation properties of E and R. Here, for example, is a proof of the *reductio* axiom.

$$
\begin{array}{lll}
1 & A \to \sim A_{\{1\}} & \text{hyp} \\
2 & \quad A_{\{2\}} & \text{hyp} \\
3 & \quad A \to \sim A_{\{1\}} & 1 \text{ reit} \\
4 & \quad \sim A_{\{1,2\}} & 2\text{--}3 \ \to\text{E} \\
5 & \sim A_{\{1\}} & 2\text{--}4 \ \sim\text{I} \\
6 & (A \to \sim A) \to \sim A & 1\text{--}5 \ \to\text{I}
\end{array}
$$

The rules for conjunction are also straightforward.

○ ($\wedge E_1$) From $A \wedge B_a$ to deduce A_a.

○ ($\wedge E_2$) From $A \wedge B_a$ to deduce B_a.

○ ($\wedge I$) From A_a and B_a to deduce $A \wedge B_a$.

These rules mirror the Hilbert axiom conditions (which make $\wedge$ a lattice join). The conjunction entails both conjuncts, and the conjunction is the weakest thing which entails both conjuncts.

We do not have a rule which says that if *A* depends on *something* and *B* depends on *something else* then $A \wedge B$ depends on *those things together*, because that would allow us

to do too much. If we *did* have a connective (use "&" for this connective for the moment) which satisfied the same *elimination* clause as conjunction, and which satisfied that liberal *introduction* rule, it would allow us to prove the positive paradox in the following way.

$$
\begin{array}{lll}
1 & A_{\{1\}} & \text{hyp} \\
2 & \quad B_{\{2\}} & \text{hyp} \\
3 & \quad A_{\{1\}} & \text{1 reit} \\
4 & \quad A\&B_{\{1,2\}} & \text{2, 3 \&I} \\
5 & \quad A_{\{1,2\}} & \text{4 \&E} \\
6 & B \to A_{\{1\}} & \text{2–5 } \to\text{I} \\
7 & A \to (B \to A) & \text{1–6 } \to\text{I}
\end{array}
$$

If we have a connective with the elimination rules of conjunction (which we surely require, if that connective is to be "and" in the traditional sense) then the liberal rules are too strong. They would allow us to take vacuous excursions through conjunction introductions and elimination, picking up irrelevant indices along the way.

No, the appropriate introduction rule for a conjunction is the restricted one which requires that both conjuncts *already* have the same relevance label. This, somewhat surprisingly, suffices to prove everything we can prove in the Hilbert system. Here, for an example, is the proof of the *conjunction introduction* Hilbert axiom.

$$
\begin{array}{lll}
1 & (A \to B) \wedge (A \to C)_{\{1\}} & \text{hyp} \\
2 & A \to B_{\{1\}} & \text{1 } \wedge\text{E} \\
3 & A \to C_{\{1\}} & \text{1 } \wedge\text{E} \\
4 & \quad A_{\{2\}} & \text{hyp} \\
5 & \quad A \to B_{\{1\}} & \text{2 reit} \\
6 & \quad B_{\{1,2\}} & \text{4, 5 } \to\text{E} \\
7 & \quad A \to C_{\{1\}} & \text{3 reit} \\
8 & \quad C_{\{1,2\}} & \text{4, 7 } \to\text{E} \\
9 & \quad B \wedge C_{\{1,2\}} & \text{6, 8 } \wedge\text{I} \\
10 & A \to B \wedge C_{\{1\}} & \text{4–9 } \sim\text{I} \\
11 & (A \to B) \wedge (A \to C) \to (A \to B \wedge C) & \text{1–10 } \to\text{I}
\end{array}
$$

From these rules, using the De Morgan equivalence between $A \vee B$ and $\sim(\sim A \wedge \sim B)$, it is possible to derive the following two rules for disjunction.[31] Unfortunately, these rules essentially involve the conditional. There seems to be no way to isolate rules which involve disjunction alone.

○ ($\vee I_1$) From A_a to deduce $A \vee B_a$.

○ ($\vee I_2$) From B_a to deduce $A \vee B_a$.

○ ($\vee E$) From $A \to C_a$ and $B \to C_a$ and from $A \vee B_b$ to deduce $C_{a \cup b}$.

The most disheartening thing about these rules for disjunction (and about the natural deduction system itself) is that they do not suffice. They do not prove the distribution of conjunction over disjunction. Anderson and Belnap had to posit an extra rule.

---

[31] See Anderson and Belnap's *Entailment* [1975, §23.2] for the details.

○ (Dist) From $A \wedge (B \vee C)_a$ to deduce $(A \wedge B) \vee C_a$

It follows that this Fitch-style proof theory, while useful for proving things in R or E, and while giving *some* separation of the distinct behaviours of the logical connectives, does not provide pure introduction and elimination rules for each connective. For a proof theory which does *that*, the world would have to wait until the 1970s, and for some independent work of Grigori Minc [1972; 1977][32] and J. Michael Dunn [1973].[33]

The *fusion* connective ○ plays a minor role in early work in the Anderson–Belnap tradition.[34] They noted that it has some interesting properties in R, but that the residuation connection fails in E if we take $A \circ B$ to be defined as $\sim(A \to \sim B)$. Residuation fails because $\sim(A \to \sim B) \to C$ does not entail $A \to (B \to C)$ if we cannot permute antecedents of arbitrary conditionals. Since E was their focus, fusion played a little role in their early work. Later, with Dunn's development of natural algebraic semantics, and with the shift of focus to R, fusion began to play a more central role.

The topic of finding a natural proof theory for relevant implication — and in particular, the place of distribution in such a proof theory — was a recurring theme in logical research in this tradition. The problem is not restricted to Fitch-style systems. Dag Prawitz's [1965] monograph *Natural Deduction*, launched Gentzen-style natural deduction systems on to centre stage. At the end of the book, Prawitz remarked that modifying the rules of his system would give you a system of relevant implication. Indeed they do. Almost.

Rules in Prawitz's system are simple. Proofs take the form of a tree. Some rules simply extend trees downward, from one conclusion to another. Others, take two trees and join them into a new tree with a single conclusion.

$$\frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B} \qquad \frac{A \quad B}{A \wedge B} \qquad \frac{A \to B \quad A}{B}$$

These rules have as *assumptions* any undischarged premises at the top of the tree. To prove things on the basis of no assumptions, you need to use rules which *discharge* them. For example, the implication introduction rule is of this form:

$$\frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array}}{A \to B.}$$

This indicates that at the node for B there is a collection of open assumptions A, and we can derive $A \to B$, closing those assumptions. Prawitz hypothesised that if you modified his rules to only allow the discharge of assumptions which were actually *used* in a proof, as opposed to allowing vacuous discharge (which is required in the proof of $A \to (B \to A)$, for example), you would get a system of relevant logic in the style of

---

[32]Then in Russia, and now at Stanford. He publishes now under the name "Grigori Mints".

[33]A graduate student of Nuel Belnap's.

[34]They call ○ "fusion" after trying out names such as "cotenability" or "compossibility", connected with the definition as $\sim(A \to \sim B)$.

Anderson and Belnap. Keeping our attention to implication alone, the answer is correct. His rule modification gives us a simple natural deduction system for R.

However, Prawitz's rules for relevant logic are less straightforward once we attempt to add conjunction. If we keep the rules as stated, then the conjunction rules allow us to prove the positive paradox in exactly the same way as in the case with & in the Fitch system.[35]

$$
\frac{\dfrac{A^2 \qquad B^1}{A \wedge B} \, (\wedge\mathrm{I})}{\dfrac{\underline{\quad A \quad}}{\dfrac{B \to A}{A \to (B \to A)} \, (2, \to \mathrm{I})} \, (1, \to \mathrm{I})} (\wedge\mathrm{E})
$$

We must do something to the rule for conjunction introduction to ban this proof. The required amendment is to only allow conjunction introduction when the two subproofs have exactly the same open assumptions. A similar amendment is required for disjunction elimination. And *then*, once those patches are applied, it turns out that distribution is no longer provable in the system. (The intuitionistic or classical proof of distribution in Prawitz's system requires either a weakening in or an irrelevant assumption or a banned conjunction or disjunction move.) Prawitz's system is no friendlier to distribution than is Fitch's.

Logics *without* distribution, such as linear logic, are popular, in part, because of the difficulty of presenting straightforward proof systems for logics with distribution. In general, proof theories seem more natural or straightforward doing without it. The absence of distribution has also sparked debate. The *naturalness* or otherwise of a proof theory is no argument in and of itself for the *failure* of distribution. See Belnap's "Life in the Undistributed Middle" [1993] for more on this point.

Embeddings

One of the most beautiful results of early work on relevant logic is the *embedding* results showing how intuitionistic logic, classical logic and S4 find their home inside R and E [Anderson and Belnap, 1961; Meyer, 1970a; 1973b]. The idea is that we can move to an *irrelevant* conditional by recognising that such conditionals might be *enthymemes*. When I say that $A \supset B$ holds ($\supset$ is the intuitionistic conditional), I am not saying that B follows *from A*, I am saying that B follows from A together perhaps with some truth or other. One simple way to say this is to lean on the addition of the Ackermann constant t. We can easily add t to R by way of the following equivalences

$$
A \to (t \to A) \qquad (t \to A) \to A.
$$

---

[35]The superscripts and the line numbers pair together assumptions and the points in the proof at which they were discharged.

These state that a claim is *true* just in case it follows from t.[36] Given t we can define the enthymematic conditional $A \supset B$ as follows. $A \supset B$ is

$$A \wedge t \rightarrow B$$

which states that B follows from A together with some truth or other. Now, $A \supset (B \supset A)$ is provable: in fact, the stronger claim $A \rightarrow (B \supset A)$ is provable, since it follows directly from the axiom $A \supset (t \supset A)$. But this is no longer paradoxical, since $B \supset A$ does not state that A follows from B. (The proof that you get precisely intuitionistic logic through this embedding is a little trickier than it might seem. You need to revisit the definition of intuitionistic negation (write it "$\neg$" for the moment) in order to show that $A \wedge \neg A \supset B$ holds.[37] The subtleties are to be found in a paper by Meyer [1973b].

The same kind of process will help us embed the *strict* conditional of S4 into E. In E, t is not only true but *necessary* (as the necessary propositions are those entailed by t) so the effect of the enthymematic definition in E is to get a strict conditional. If we define $A \Rightarrow B$ as $A \wedge t \rightarrow B$ in E, then the $\wedge, \vee, \Rightarrow$ fragment of E is exactly the $\wedge, \vee, \Rightarrow$ fragment of S4 [Anderson and Belnap, 1992, §35].

We can extend the modelling of intuitionistic logic into E if we step further afield. We require not only the propositional atom t, but some more machinery: the machinery of propositional *quantification*. If we add propositional quantifiers $\forall p$ and $\exists p$ to E[38] then intuitionistic and strict implication are defined as follows:

$$
\begin{aligned}
A \supset B \quad &=_{df} \quad \exists p(p \wedge (p \wedge A \rightarrow B)) \\
A \Rightarrow B \quad &=_{df} \quad \exists p(\Box p \wedge (p \wedge A \rightarrow B)).
\end{aligned}
$$

An intuitionistic conditional asserts that there is some *truth*, such that it conjoined with *A entails B*. A strict conditional asserts that there is some *necessary truth*, such that it conjoined with *A entails B*.

Embedding the classical conditional into relevant logic is also possible. The amendment is that not only do we need to show that *weakening* is possible, but contradictions must entail everything: and we want to attempt this without introducing a new negation. The innovation comes from noticing that we can *dualise* the enthymematic construction. Instead of just requiring an extra truth as a conjunct in the antecedent, we can admit an extra *falsehood* as a *disjunct* in the *consequent*. The classical conditional (also written "$A \supset B$") can be defined like this

$$A \wedge t \rightarrow B \vee f$$

where $f = \sim t$. Now we will get $A \wedge \sim A \supset B$ since $A \wedge \sim A \rightarrow f$.[39] Anderson and Belnap make some sport of material "implication" on the basis of this definition. Note

---

[36]The first axiom here is too strong to govern t in the logic E, in which case we replace it by the permuted form $t \rightarrow (A \rightarrow A)$. The claim t doesn't *entail* all truths. (If it did, then all truths would be provable, since t is provable.) Instead, t entails all *identities*.

[37]You can't just use the negation of relevant logic, because of course we get $A \supset B \vee \sim B$, since $t \rightarrow B \vee \sim B$.

[38]And the proof theory for propositional quantifiers is not difficult [Anderson and Belnap, 1992, §30–32].

[39]The result can be extended to embed the whole of S4 into E (rather than only its positive fragment of S4) by setting $A \Rightarrow B =_{df} \exists p(\Box p \wedge (p \wedge A \rightarrow B \vee \sim p))$.

that constructive implication is still genuinely an *implication* with the consequent being what we expect to conclude. A "material" implication is genuinely an implication, but you cannot conclude the consequent of the original conditional by modus ponens with the antecedent. No, you can only conclude the consequent with a disjoined $\vee f$.[40]

Arguments about disjoined fs lead quite well into arguments over the law of *disjunctive syllogism*, and these are the focus of our next section.

2.4   Disjunctive Syllogism

We have already seen that Ackermann's system Π' differs from Anderson and Belnap's system E by the presence of the rule (γ). In Ackermann's Π', we can directly infer B from the premises $A \vee B$ and $\sim A$. In E, this is not possible: for E a rule of inference from X to B is admitted only when there is some corresponding *entailment* from X (or the conjunction of formulas in X) to B. As disjunctive syllogism in an *entailment*

$$(A \vee B) \wedge \sim A \rightarrow B$$

is not present, Anderson and Belnap decided to do without the rule too. This motivates a question. Does dropping the rule (γ) change the set of theorems? Is there anything you can prove with (γ) that you cannot prove without it? Of course there are things you can prove from *hypotheses*, using (γ) which cannot be proved without it. In Ackermann's system Π' there is a straightforward proof for $A, \sim A \Rightarrow B$. In Anderson and Belnap's E there is no such proof. However, this leaves the special case of proofs from no hypotheses. Is it the case that in E, if $\vdash A \vee B$ and $\vdash \sim A$ that $\vdash B$ too? This is the question of the *admissibility* of disjunctive syllogism. If disjunctive syllogism is admissible in E then its *theorems* do not differ from the theorems of Ackermann's Π'.

A Proof of the Admissibility of Disjunctive Syllogism

There are four different proofs of the admissibility of disjunctive syllogism for logics such as E and R. The first three proofs are due to Meyer [1973](with help from Dunn on the first [1969], and help from Routley on the second [1976a]). They all depend on the same first step, which we will describe here as the *way up lemma*. The last proof was obtained by Kripke in 1978. In this section I will sketch the third of Meyer's proofs, because it will illustrate two techniques which have proved fruitful in the study of relevant and substructural logics. It is worth examining this result in some detail because it shows some of the distinctive techniques in the metatheory of relevant logics.

FACT 3 (Disjunctive Syllogism is Admissible in E and R).   In both E and R, if $\vdash A \vee B$ and $\vdash \sim A$ then $\vdash B$.

---

[40]I suspect that the situation is not quite so bad for material implication. If one treats acceptance and rejection, assertion and denial with equal priority, and if you take the role of implication as not only warranting the acceptance of the consequent, given the acceptance of the antecedent but also the rejection of the antecedent on the basis of the rejection of the consequent, then the enthymematic definition of the material conditional seems not so bad [Restall, 2000c].

To present the bare bones of the proof of this result, we need some definitions.

DEFINITION 4 (Theories). A set T of formulas is a *theory* if whenever $A, B \in T$ then $A \wedge B \in T$, and if $A \vdash B$ then if $A \in T$ we also have $B \in T$. Theories are closed under conjunction and provable consequence.

Note that theories in relevant logics are rather special. Nonempty theories in *irrelevant* logics contain all theorems, since if $A \in T$ and if B is a *theorem* then so is $A \to B$ in an irrelevant logic. In relevant logics this is not the case, so theories need not contain all theorems.

Furthermore, since $A \wedge \sim A \to B$ is not a theorem of relevant logics, theories may be inconsistent without being trivial. A theory might contain an inconsistent pair A and $\sim A$, and contain its logical consequences, without the theory containing any formula whatsoever.

Finally, consistent and complete theories in classical propositional logic respect all logical connectives. In particular, if $A \vee B$ is a member of a consistent and complete theory, then one of A and B is a member of that theory. For if neither are, then $\sim A$ and $\sim B$ are members of the theory, and so is $\sim(A \vee B)$ (by logical consequence) contradicting $A \vee B$'s membership of the theory. In a logic like R or E it is quite possible for $A \vee B$ and $\sim(A \vee B)$ to be members of our theory without the theory becoming trivial. A theory can be complete without respecting disjunction. It turns out that theories which respect disjunction play a very important role, not only in our proof of the admissibility of disjunctive syllogism, but also in the theory of models for substructural logics. So, they deserve their own definition.

DEFINITION 5 (Special Theories). A theory T is said to be *prime* if whenever $A \vee B \in T$ then either $A \in T$ or $B \in T$. A theory T is said to be *regular* (with respect to a particular logic) whenever it contains all of the theorems of that logic.

Now we can sketch the proof of the admissibility of (γ).

Proof. We will argue by *reductio*, showing that there cannot be a case where $A \vee B$ and $\sim A$ are provable but B is not. Focus on B first. If B is not provable, we will show first that there is a *prime theory* containing all of the theorems of the logic but which still avoids B. This stage is the *Way Up*. We may have overshot our mark on the Way Up, as a prime theory containing all theorems will certainly be *complete* (as $C \vee \sim C$ is a theorem in E or R so one of C and $\sim C$ will be present in our theory) but it may not be *consistent*. If we can have a *consistent* complete prime theory containing all theorems but still missing out B we will have found our contradiction, for since this new theory contains all theorems, it contains $A \vee B$ and $\sim A$. By *primeness* it contains either A or it contains B. Containing A is ruled out since it already contains $\sim A$, so containing B is the remaining option.[41] So, the *Way Down* cuts down our original theory into a consistent and complete one. Given the way up and the way down, we will have our result. Disjunctive syllogism is admissible. ■

---

[41]Note here that disjunctive syllogism was used in the language used to present the proof. Much has been made of this in the literature on the significance of disjunctive syllogism [Belnap and Dunn, 1981; Meyer, 1978].

All that remains is to prove both Way Up and Way Down lemmas.

FACT 6 (Way Up Lemma). If $\nvdash A$, then there is a regular prime theory T such that $A \notin T$.

This is a special case of the general *pair extension theorem*, which is so useful in relevant and substructural logics that it deserves a separate statement and a sketch of its proof. To introduce this proof, we need a new definition to keep track of formulas which are to appear *in* our theory, and those which are to be kept *out*.

DEFINITION 7 ($\vdash$-pairs). An ordered pair $\langle L, R \rangle$ of sets of formulae is said to be a $\vdash$-*pair* if and only if there are *no* formulas $A_1, \ldots, A_n \in L$ and $B_1, \ldots, B_m \in R$ where $A_1 \wedge \cdots \wedge A_n \vdash B_1 \vee \cdots \vee B_m$.

A helpful shorthand will be to write '$\bigwedge A_i \vdash \bigvee B_j$' for the extended conjunctions and disjunctions. A $\vdash$-pair is represents a set of formulas we wish to take to be *true* (those in the left) and those we wish to take to be *false* (those in the right). The process of constructing a prime theory will involve enumerating the entire language and building up a pair, taking as many formulas as possible to be true, but adding some as false whenever we need to. So, we say that a $\vdash$-pair $\langle L', R' \rangle$ *extends* $\langle L, R \rangle$ if and only if $L \supseteq L'$ and $R \supseteq R'$. We write this as "$\langle L, R \rangle \subseteq \langle L', R' \rangle$." The end point of this process will be a *full* pair.

DEFINITION 8 (Full $\vdash$-Pairs). A $\vdash$-pair $\langle L, R \rangle$ is a *full $\vdash$-pair* if and only if $L \cup R$ is the entire language.

Full $\vdash$-pairs are important, as they give us prime theories.

FACT 9 (Prime Theories from Full $\vdash$-Pairs). If $\langle L, R \rangle$ is a full $\vdash$-pair, L is a prime theory.

Proof. We need to verify that L is closed under consequence and conjunction, and that it is prime. First, consequence. Suppose $A \in L$ and that $A \vdash B$. If $B \notin L$, then since $\langle L, R \rangle$ is full, $B \in R$. But then $A \vdash B$ contradicts the condition that $\langle L, R \rangle$ is a $\vdash$-pair.

Second, conjunction. If $A_1, A_2 \in x$, then since $A_1 \wedge A_2 \vdash A_1 \wedge A_2$, and $\langle L, R \rangle$ is a $\vdash$-pair, we must have $A_1 \wedge A_2 \notin y$, and since $\langle L, R \rangle$ is full, $A_1 \wedge A_2 \in L$ as desired.

Third, primeness. If $A_1 \vee A_2 \in L$, then if A_1 and A_2 are both not in L, by fullness, they are both in R, and since $A_1 \vee A_2 \vdash A_1 \vee A_2$, we have another contradiction to the claim that $\langle L, R \rangle$ is a $\vdash$-pair. Hence, one of A_1 and A_2 is in L, as we wished. ∎

FACT 10 (Pair Extension Theorem). If $\vdash$ is the logical consequence relation of a logic including all distributive lattice properties, then any $\vdash$-pair $\langle L, R \rangle$ is extended by some full $\vdash$-pair $\langle L', R' \rangle$.

To prove this theorem, we will assume that we have enumerated the language so that every formula in the language is in the list $C_1, C_2, \ldots, C_n, \ldots$ We will consider each formula one by one, to check to see whether we should throw it in L or in R instead. We assume, in doing this, that our language is *countable*.[42]

Proof. First we show that if $\langle L, R \rangle$ is a $\vdash$-pair, then so is at least one of $\langle L \cup \{C\}, R \rangle$ and $\langle L, R \cup \{C\} \rangle$, for any formula C. Equivalently, we show that if $\langle L \cup \{C\}, R \rangle$ is not a $\vdash$-pair,

---

[42]The general kind of proof works for *well-ordered* languages as well as countable languages.

then the alternative, $\langle L, R \cup \{C\}\rangle$, is. If this were not a $\vdash$-pair either, then there would be some $A \in \bigwedge L$ (the set of all conjunctions of formulae from L) and $B \in \bigvee R$ where $A \vdash B \vee C$. Since $\langle L \cup \{C\}, R\rangle$ is not a $\vdash$-pair, there are also $A' \in \bigwedge L$ and $B' \in \bigvee R$ such that $A' \wedge C \vdash B'$. But then, $A \wedge A' \vdash B \vee C$. But this means that $A \wedge A' \vdash (B \vee C) \wedge A'$. Now by distributive lattice properties, we then get $A \wedge A' \vdash B \vee (A' \wedge C)$. But $A' \wedge C \vdash B'$, so disjunction properties, and the transitivity of $\vdash$ together give us $A \wedge A' \vdash B \vee B'$, contrary to the fact that $\langle L, R\rangle$ is a $\vdash$-pair.

With that fact in hand, we can create our full pair. Define the series of $\vdash$-pairs $\langle L_n, R_n\rangle$ as follows. Let $\langle L_0, R_0\rangle = \langle L, R\rangle$, and given $\langle L_n, R_n\rangle$ define $\langle L_{n+1}, R_{n+1}\rangle$ in this way.

$$\langle L_{n+1}, R_{n+1}\rangle = \begin{cases} \langle L_n \cup \{C_n\}, R_n\rangle & \text{if } \langle L_n \cup \{C_n\}, R_n\rangle \text{ is a } \vdash\text{-pair,} \\ \langle L_n, R_n \cup \{C_n\}\rangle & \text{otherwise.} \end{cases}$$

Each $\langle L_{n+1}, R_{n+1}\rangle$ is a $\vdash$-pair if its predecessor $\langle L_n, R_n\rangle$ is, for there is always a choice for placing C_n while keeping the result a $\vdash$-pair. So, by induction on n, each $\langle L_n, R_n\rangle$ is a $\vdash$-pair. It follow then that $\langle \bigcup_{n\in\omega} L_n, \bigcup_{n\in\omega} R_n\rangle$, the limit of this process, is also a $\vdash$-pair, and it covers the whole language. (If $\langle \bigcup_{n\in\omega} L_n, \bigcup_{n\in\omega} R_n\rangle$ is not a $\vdash$-pair, then we have some $A_i \in \bigcup L_n$ and some $B_j \in \bigcup R_n$ such that $A_1 \wedge \cdots \wedge A_l \vdash B_1 \vee \cdots \vee B_m$, but if this is the case, then there is some number n where each A_i is in L_n and each B_j is in R_n. It would follow that $\langle L_n, R_n\rangle$ is not a $\vdash$-pair.) So, we are done. ∎

Belnap proved the Pair Extension Theorem in the early 1970s. Dunn circulated a write-up of it in about 1975, and cited it in some detail in 1976 [1976b]. Gabbay independently used the result for first-order intuitionistic logic, also in 1976 [1976]. The theorem gives us the Way Up Lemma, because if $\nvdash B$, then $\langle Th, \{B\}\rangle$ is a $\vdash$-pair, where Th is the set of theorems. Then this pair is extended by a full pair, the left part of which is a regular prime theory, avoiding B.

Now we can complete our story with the proof of the Way Down Lemma.

Proof. We must move from our regular prime theory T to a consistent regular prime theory $T^* \subseteq T$. We need the concept of a "metavaluation." The concept and its use in proving the admissibility (γ) is first found in Meyer's paper from 1976 [1976a]. A metavaluation is a set of formulas T^* on formulas defined inductively on the construction of formulas as follows:

○ For a propositional atom p, $p \in T^*$ if and only if $p \in T$;

○ $\sim A \in T^*$ iff (A) $A \notin T^*$, and (B) $\sim A \in T$;

○ $A \wedge B \in T^*$ iff both $A \in T^*$ and $B \in T^*$;

○ $A \vee B \in T^*$ iff either $A \in T^*$ or $B \in T^*$;

○ $A \rightarrow B \in T^*$ iff (A) if $A \in T^*$ then $B \in T^*$ and (B) $A \rightarrow B \in T$.

Note the difference between the clauses for the *extensional* connectives $\wedge$ and $\vee$ and the *intensional* connectives $\rightarrow$ and $\sim$. The extensional connectives have "one-punch" rules which match their evaluation with respect to truth tables. The *intensional* connectives are more complicated. They require *both* that the formula is in the original theory and that the extensional condition holds in the new set T^*.

We will prove that T^* is a regular theory. Its *primeness* and *consistency* are already delivered by fiat, from the clauses for $\vee$ and $\sim$. The first step on the way is a simple lemma.

FACT 11 (Completeness Lemma). If $A \in T^*$ then $A \in T$, and if $A \notin T^*$ then $\sim A \in T$.

It is simplest to prove both parts *together* by induction on the construction of A. As an example, consider the case for implication. The positive part is straightforward: if $A \to B \in T^*$ then $A \to B \in T$ by fiat. Now suppose $A \to B \notin T^*$. Then it follows that either $A \to B \notin T$ or $A \in T^*$ and $B \notin T^*$. In the first case, by the completeness of T, $\sim(A \to B) \in T$ follows immediately. In the second case, $A \in T^*$ (so by the induction hypothesis, $A \in T$) and $B \notin T^*$ (so by the induction hypothesis, $\sim B \in T$). Since $A, \sim B \vdash \sim(A \to B)$ in both R and E, we have $\sim(A \to B) \in T$, as desired.

It is also not too difficult to check that T^* is a regular theory. First, T^* is closed under conjunction (by the conjunction clause) and it is *detached* (closed under *modus ponens*, by the implication clause). To show that it is a regular theory, then, it suffices to show that every axiom of the Hilbert system for R is a member. To give you an idea of how it goes, I shall consider two typical cases.

First we check *suffixing*: $(A \to B) \to ((B \to C) \to (A \to C))$. Suppose it isn't in T^*. Since it is a theorem of the logic and thus a member of T, it satisfies the intensional condition and so must fail to satisfy the extensional condition. So $A \to B \in T^*$ and $(B \to C) \to (A \to C) \notin T^*$. By the Completeness Lemma, then $A \to B \in T$, and so by *modus ponens* from the suffixing axiom itself, we have that $(B \to C) \to (A \to C) \in T$. So $(B \to C) \to (A \to C)$ satisfies the intensional condition, and so must fail to satisfy the extensional condition: $B \to C \in T^*$ and $A \to C \notin T^*$. By similar reasoning we derive that $A \to C$ must finally fail to satisfy the extensional condition, i.e. $A \in T^*$ and $C \notin T^*$. But since of $A \to B \in T^*$, $B \to C \in T^*$, $A \in T^*$, by the extensional condition, $C \in T^*$, and we have a contradiction.

Second, check *double negation elimination*: $\sim\sim A \to A$. Suppose it isn't in T^*. Again, since it's a theorem of the logic and thus a member of T, if it fails it must fail the extensional condition. So, $\sim\sim A \in T^*$ but $A \notin T^*$. Since $\sim\sim A \in T^*$, by the negation clause, we have both $\sim A \notin T^*$ and $\sim\sim A \in T$. From $\sim\sim A \in T$, using double negation elimination, we get $A \in T$. Using the negation clause again, unpacking $\sim A \notin T^*$, we have either $A \in T^*$ or $\sim A \notin T$. The first possibility clashes with our assumption that $A \notin T^*$. The second possibility, $\sim A \notin T$ clashes again with $A \notin T^*$, using the Completeness Lemma.

The same techniques show that each of the other axioms are also present in T^*. Finally T^* is closed under *modus ponens*, and as a result, T^* is a complete, consistent regular theory, and a subset of T. This completes our proof of the Way Down Lemma. ∎

Meyer pioneered the use of metavaluations in relevant logic [1971; 1976]. Metavaluations were also used by Kleene in his study of intuitionistic theories [1962; 1963], who was in turn inspired by Harrop, who used the technique in the 1950s to prove primeness for intuitionistic logic [1956].

There are many different proofs of the admissibility of disjunctive syllogism. Meyer [1998] pioneered the technique using metavaluations, and Meyer and Dunn [1969] have

used other techniques . Friedman and Meyer [1992] showed that disjunctive syllogism fails in first-order relevant Peano arithmetic, but that it holds when you add an infinitary "omega" rule. Meyer and I have used a different style of metavaluation argument to construct a complete "true" relevant arithmetic [1996]. This metavaluation argument treats negation with a "one-punch" clause: $\sim A \in T^*$ if and only if $A \notin T^*$. In this arithmetic, $0 = 1 \rightarrow 0 = 2$ is a theorem, as you can deduce $0 = 2$ from $0 = 1$ by arithmetic means, while $\sim(0 = 2 \rightarrow 0 = 1)$ is a theorem, as there is no way, by using multiplication, addition and identity, to deduce $0 = 1$ from $0 = 2$.

Interpretation

A great deal of the literature *interpreting* relevant logics has focussed on the status of disjunctive syllogism. The *relevantist* of Belnap and Dunn's essay "Entailment and Disjunctive Syllogism" [1981] is a stout-hearted person who rejects all use of disjunctive syllogism. Belnap and Dunn explain how difficult it is to maintain this line. Once you learn $A \vee B$ and you learn $\sim A$, it is indeed difficult to admit that you have no reason at all to conclude B. Stephen Read is perhaps the most prominent *relevantist* active today [1981; 1988]. Read's way of resisting disjunctive syllogism is to argue that in any circumstance in which there is pressure to conclude B from $A \vee B$ and $\sim A$, we have pressure to admit more than $A \vee B$: we have reason to admit $\sim A \rightarrow B$, which will licence the conclusion to B.

Some proponents of relevantism reject disjunctive syllogism not merely because it leads to fallacies of relevance, but because it renders non-trivial but inconsistent theories impossible [Meyer and Martin, 1986; Routley, 1984]. The *strong* version of this view is that inconsistencies are not only items of non-trivial theories, they are genuine *possibilities* [Priest, 2000]. Such a view is *dialetheism*, the thesis that contradictions are possible. Not all proponents of relevant logics are dialetheists, but dialetheism has provided a strong motivation for research into relevant logics, especially in Australia.[43]

My view on this issue differs from each of the relevantist, the dialetheist and the classicalist (who accepts disjunctive syllogism, and hence rejects relevant logic) by being *pluralistic* [Beall and Restall, 2000; Restall, 1999]. Disjunctive syllogism is indeed inappropriate to apply to the content of inconsistent theories. However, it is impossible that the premises of an instance of disjunctive syllogism be true if at the very same time the conclusion is not true. Relevant entailment is not the *only* constraint under which truth may be regulated. Relevant entailment is one useful criterion for evaluating reasoning, but it is not the only one. If we are given reason to believe $A \vee B$ and reason to believe $\sim A$, then (provided that these reasons do not conflict with one another) we have reason to believe B. This reason is not one licensed by relevant consequence, but relevant consequence is not the only sort of licence to which a good inference might aspire.

---

[43]See the Australian entries in the volume "Paraconsistent Logic: Essays on the Inconsistent" [Priest *et al.*, 1989], for example [Brady, 1989; Brady and Routley, 1989; Meyer and Slaney, 1989; Priest and Sylvan, 1989; Slaney, 1989].

Debate over disjunctive syllogism has motivated important formal work in relevant logics. If you take the lack of disjunctive syllogism to be a fault in relevant logics, you can always add a new negation (say, *Boolean* negation, written '$-$') which satisfies the axioms $A \wedge -A \rightarrow B$ and $A \rightarrow B \vee -B$. Then relevant logics are truly systems of modal logic extending classical propositional logic with two modal or intensional operators, $\sim$ (a one-place operator) and $\rightarrow$ (a two-place operator). Meyer and Routley have presented alternative axiomatisations of relevant logics which contain Boolean negation '$-$', and the material conditional $A \supset B =_{df} -A \vee B$, as the primary connective [1973a; 1973b].

2.5 Lambek Calculus

Lambek worked on his calculus to model the behaviour of syntactic and semantic types in natural languages. He used technique from proof theory [1958; 1961] (as well as techniques from category theory which we will see later [Lambek, 1969]). His techniques built on work of Bar-Hillel [1953] and Ajdukiewicz [1935] who in turn formalised some insights of Husserl.

The logical systems Lambek studied contain implication connectives and a fusion connective. Fusion in this language is not commutative, so it naturally motivates *two* implication connectives $\rightarrow$ and $\leftarrow$.[44] We get two arrow connectives because we may residuate $A \circ B \vdash C$ by isolating A on the antecedent, or equally, by isolating B.

$$\frac{\overline{\overline{A \vdash B \rightarrow C}}}{\overline{\overline{A \circ B \vdash C}}}$$
$$\overline{B \vdash C \leftarrow A}$$

If fusion is commutative — that is, if $A \circ B$ is equivalent to $B \circ A$ — then $B \rightarrow C$ will have the same effect as $C \leftarrow B$. If $B \circ A$ differs from $A \circ B$ then so $\rightarrow$ and $\leftarrow$ will also differ.

One way to view Lambek's innovation is to see him as motivating and developing a substructural logic in which *two* implications have a natural home. To introduce this system, consider the problem of assigning *types* to strings in some language. We might assign types to primitive expressions in the language, and explain how these could be combined to form complex expressions. The result of such a task is a *typing judgement* of the form $x \Vdash A$, indicating that the string x has the type A. Here are some example typing judgements.

---

[44]Lambek wrote the two implication connectives as "/" for $\rightarrow$ and "\" for $\leftarrow$, and fusion as concatenation, but to keep continuity with other sections I will use the notation of arrows and the circle for fusion.

$$
\begin{array}{rcl}
\texttt{John} & \Vdash & n \\
\texttt{poor} & \Vdash & n \to n \\
\texttt{John works} & \Vdash & s \\
\texttt{works} & \Vdash & s \leftarrow n \\
\texttt{must work} & \Vdash & s \leftarrow n \\
\texttt{work} & \Vdash & i \\
\texttt{must} & \Vdash & i \to (s \leftarrow n) \\
\texttt{John work} & \Vdash & n \circ i
\end{array}
$$

Types can be atomic or complex: they form an algebra of formulas just like those in a propositional logic. Here, the judgement "John $\Vdash n$" says that the string John has the type n (for *name*, or *noun*). The next judgement says that poor has a special *compound* type $n \to n$: it converts names to names. It does this by *composition*. The string poor has the property that when you *prefix* it to a string of type n you get another (compound) string of type n. So, poor John has type n. So does poor Jean, and poor Joan of Arc (if Jean and Joan of Arc have the requisite types).[45] Strings can, of course, be concatenated at the end of other strings too. The string works has type $s \leftarrow n$ because whenever you *suffix* a string of type n with works you get a string of type s (a *sentence*). John works, poor Joan works and poor poor Joan of Arc works are all sentences, according to this grammar.

Typing can be nested arbitrarily. We see that must work has type $s \leftarrow n$ (it acts like works). The word work has type i (*intransitive infinitive*) so must has type $i \to (s \leftarrow n)$. When you concatenate it in front of any string of type i you get a string of type $s \leftarrow n$. So must play and must subscribe to *New Scientist* also have type $s \leftarrow n$, as play and subscribe to *New Scientist* have type i.

Finally, compositions have types even if the results do not have a predefined simple type. John work at least has the type $n \circ i$, as it is a concatenation of a string of type n with a string of type i. The string *must work* also has type $(i \to (s \leftarrow n)) \circ i$, because it is a composition of a string of type $i \to (s \leftarrow n)$ with a string of type i. Clearly here fusion is not commutative. John work has type $n \circ i$, but work John does not. As a corollary, $\to$ and $\leftarrow$ differ. Given the associativity of concatenation of strings, fusion is associative too. Any string of type $A \circ (B \circ C)$ is of type $(A \circ B) \circ C$. We can associate freely in any direction.[46]

Once we have a simple type system like this, it is possible to make *inferences* about type assigments, on the basis of the interactions of the type-constructors $\to$, $\leftarrow$ and $\circ$. One of Lambek's innovations was to notice that this type system can be manipulated using a simple Gentzen-style consecution calculus. This calculus manipulates consecutions of the form $A_1, A_2, \ldots, A_n \vdash B$. We read this consecution as asserting that any string which

---

[45]According to this definition, poor poor John and poor poor poor poor Joan of Arc are also strings of type n.

[46]We can associate *fusion* freely, not the conditionals. $A \to (B \to C)$ is not the same type as $(A \to B) \to C$, as you can check.

is a concatenation of strings of type $A_1, A_2, \ldots, A_n$ also has type B.[47] A list of types will be treated as a type in my explanations below.[48]

The system is made up of one axiom and a collection of rules. The elementary type axiom is the *identity*.

$$A \vdash A$$

Any string of type A is of type A. The rules introduce type constructors on the left and the right of the turnstile. Here are the rules for the left-to-right arrow.

$$\frac{X, A \vdash B}{X \vdash A \to B} \, [\to R] \qquad \frac{X \vdash A \quad Y, B, Z \vdash C}{Y, A \to B, X, Z \vdash C} \, [\to L]$$

If you know that any string of type X concatenated with a string of type A is also a string of type B, then this means that any string of type X is also of type $A \to B$. Conversely, if any string of type X is also of type A, and strings of type Y, B, Z are also of type C, then strings of type $Y, A \to B, X, Z$ are also of type C. Why is this? It is because strings of type $A \to B, X$ are also of type B, because they are concatenations of a string of type $A \to B$ to the left of a string of type X (which also has type A). The mirror image of this reasoning motivates the right-to-left conditional rules:

$$\frac{A, X \vdash B}{X \vdash B \leftarrow A} \, [\leftarrow R] \qquad \frac{X \vdash A \quad Y, B, Z \vdash C}{Y, X, B \leftarrow A, Z \vdash C} \, [\leftarrow L]$$

The next rules make fusion the direct object language correlate of the comma in the metalanguage.

$$\frac{X \vdash A \quad Y \vdash B}{X, Y \vdash A \circ B} \, [\circ R] \qquad \frac{X, A, B, Y \vdash C}{X, A \circ B, Y \vdash C} \, [\circ L]$$

Proofs in this system are trees with consecutions at the nodes, and whose leaves are axioms of identity. Each step in the tree is an instance of one or other of the rules. Here is a proof, showing that the *prefixing* axiom holds in rule form.[49]

$$\frac{\dfrac{C \vdash C \quad \dfrac{A \vdash A \quad B \vdash B}{A \to B, A \vdash B} \, (\to L)}{\dfrac{\dfrac{A \to B, C \to A, C \vdash B}{A \to B, C \to A \vdash C \to B} \, (\to R)}{A \to B \vdash (C \to A) \to (C \to B)} \, (\to R)} \, (\to L)}{}$$

Here is another proof, which combines both implication connectives.

$$
\dfrac{
  \dfrac{
    C \vdash C \quad
    \dfrac{A \vdash A \quad B \vdash B}{A \to B, A \vdash B}\ (\to L)
  }{
    \dfrac{C, (A \to B) \leftarrow C, A \vdash B}{
      \dfrac{(A \to B) \leftarrow C, A \vdash B \leftarrow C}{
        (A \to B) \leftarrow C \vdash A \to (B \leftarrow C)
      }\ (\to R)
    }\ (\leftarrow R)
  }\ (\leftarrow L)
}{}
$$

A proof system like this has a number of admirable properties. Most obvious is the clean division of labour in the rules for each connective. Each rule features only the connective being introduced, whether in antecedent (left) or consequent (right) position. Another admirable property is the way that formulas appearing in the premises also appear in the conclusion of a rule (either as entire formulas or as subformulas of other formulas). In *proof search*, there is no need to go looking for other intermediate formulas in the proof of a consecution. These two facts prove simple *conservative extension* results. Adding ∘ to the logic of ← and → would result in no more provable consecutions in the original language, because a proof of a consecution involving no fusions could not involve any fusions at all.

All of this would be for naught if the deduction system were *incomplete*. If it didn't match its intended interpretation, these beautiful properties would be useless. One important step toward proving that the deduction system is complete is proving that the *cut rule* is admissible. (Recall that a rule is admissible if whenever you can prove the premises you can prove the conclusion: adding it as an extra rule does not increase the stock of provable things.)

$$
\dfrac{X \vdash A \quad Y, A, Z \vdash B}{Y, X, Z \vdash B}\ [Cut_A]
$$

This is not a primitive rule in our calculus, because adding it would destroy the subformula property, and make proof search intolerably unbounded. It ought to be *admissible* because of the intended interpretation of ⊢. If $X \vdash A$, every string of type X is also of type A. If $Y, A, Z \vdash B$, then every string which is a concatenation of a Y an A and a Z has type B. So, given a concatenation of a Y and an X and a Z, this is also a type B since the string of type X is a string of type A. The cut rule expresses the transitivity of the "every string of type x is of type y" relation.

FACT 12 (Cut is admissible in the Lambek calculus). If $X \vdash A$ is provable in the Lambek calculus with the aid of the cut rule, it can also be proved without it.

Proof. Lambek's proof of the cut admissibility theorem parallels Gentzen's own [1934; 1969]. You take a proof featuring a *cut* and you push that cut upwards to the top of the tree, where it evaporates. So, given an instance of the cut rule, if the formula featured in the cut is not introduced in the rules above the cut, you *permute* the cut with the other rules. (You show that you could have done the cut *before* applying the other rule, instead of *after*.) Once that is done as much as possible, you have a cut where the cut formula

was *introduced* in both premises of the cut. If the formula is *atomic*, then the only way it was introduced was in an axiom, and the instance of cut is irrelevant (it has *evaporated*: cutting $Y, A, Z \vdash B$ with $A \vdash A$ gives us just $Y, A, Z \vdash B$). If the formula is not atomic, you show that you could trade in the cut on that formula with cuts on smaller formulas. Here is an example cut on the implication formula $A \to B$ introduced in both left and right branches.

$$\frac{\dfrac{W, A \vdash B}{W \vdash A \to B} [\to R] \qquad \dfrac{X \vdash A \quad Y, B, Z \vdash C}{Y, A \to B, X, Z \vdash C} [\to L]}{Y, W, X, Z \vdash C} [Cut_{A \to B}]$$

We can transform it so that cuts occur on the subformulas of $A \to B$.

$$\frac{\dfrac{X \vdash A \quad W, A \vdash B}{W, X \vdash B} [Cut_A] \qquad Y, B, Z \vdash C}{Y, W, X, Z \vdash C} [Cut_B]$$

The cases for the other formulas are just as straightforward. As formulas have only finite complexity, and trees have only finite height, this process terminates. ∎

The result that cut is admissible gives us a *decision procedure* for the calculus.

FACT 13 (Decidability of the Lambek Calculus).   The issue of whether or not a consecution $X \vdash A$ has a proof is decidable.

Proof. To check if $X \vdash A$ is provable, consider its possible ancestors in the Gentzen proof system. There are only finitely many ancestors, each corresponding to the decomposition of one of the formulas inside the consecution. (The complex cases are the implication left rules, which give you the option of many different possible places to split the Y in the antecedent $X, A \to B, Y$ or $Y, B \leftarrow A, X$, and the fusion right rule, which gives you the choice of locations to split X in $X \vdash A \circ B$.) The possible ancestors themselves are *simpler* consecutions, with fewer connectives. Decision of consecutions with *no* connectives is trivial ($X \vdash p$ is provable if and only if X *is* p) so we have our algorithm by a recursion. ∎

This decision procedure for the calculus is *exceedingly* simple. Gentzen's procedure for classical and intuitionistic logic has to deal with the structural rule of *contraction*:[50]

$$\frac{X(Y, Y) \vdash A}{X(Y) \vdash A} [\text{WI}]$$

---

[50]You'll see that the structural rule is stated in generality: contraction operates on arbitrary structures, in arbitrary contexts. This is needed for the cut elimination process. If we could contract only whole formulas, then if we wanted to push a cut past the move from $X(A, A) \vdash B$ to $X(A) \vdash B$, where we are cutting with $C, D \vdash A$, the result would require us to somehow get from $X((C, D), (C, D)) \vdash B$ to $X(C, D) \vdash B$. We cannot do this if cut operates only on *formulas*, and if associativity or commutativity is absent.

which states that if a formula is used twice in a proof, it may as well have been used once. This makes proof search chronically more difficult, as some kind of limit must be found on how many consecutions might have appeared as the premises of the consecution we are trying to prove.

Sometimes people refer to the Lambek calculus as a logic *without* structural rules, but this is not the case. The Lambek calculus presumes the *associativity* of concatenation. A proper generalisation of the calculus treats antecedent structures not as *lists* of formulas but as more general *bunches* for which the comma is a genuine ordered-pairing operation. In this case, the antecedent structure $A, (B, C)$ is not the same structure as $(A, B), C$.[51] Lambek's original calculus is properly called Lambek's *associative* calculus. The *non-associative* calculus can no longer prove the prefixing consecution. (Try to follow through the proof in the absence of associativity. It doesn't work.) Of course, given a non-associative calculus, you must modify the rules for the connectives. Instead of the rules with antecedent $X, A, Y \vdash B$ we can have $X(A) \vdash B$, where "$X(A)$" indicates a structure with a designated *instance* of A. The rule for implication on the left becomes, for example

$$\frac{X \vdash A \quad Y(B) \vdash C}{Y(A \rightarrow B, X) \vdash C.} \, [\rightarrow L]$$

Absence of structural rules also makes other things fail. The structural rule of contraction (W) is required for the *contraction* consecution.[52]

$$\frac{X((Y, Z), Z)) \vdash A}{X(Y, Z) \vdash C} \, [\mathsf{W}] \qquad \frac{\dfrac{A \vdash A \quad \dfrac{A \vdash A \quad B \vdash B}{A \rightarrow B, A \vdash B} \, (\rightarrow L)}{\dfrac{((A \rightarrow (A \rightarrow B), A), A) \vdash B}{\dfrac{A \rightarrow (A \rightarrow B), A \vdash B}{A \rightarrow (A \rightarrow B) \vdash A \rightarrow B} \, (\rightarrow R)} \, (\mathsf{W})} \, (\rightarrow L)$$

The structural rule of weakening (K) is required for the *weakening* axiom,

$$\frac{X(Y) \vdash C}{X(Y, Z) \vdash C} \, [\mathsf{K}] \qquad \frac{\dfrac{A \vdash A}{A, B \vdash A} \, (\mathsf{K})}{A \vdash B \rightarrow A} \, (\rightarrow R)$$

[51]Non-associative combination plays an important role in general grammars, according to Lambek [1961]. The role of some conversions such as wh- constructions (replacing names by "who", to construct questions, etc.) seem to require a finer analysis of the *phrase* structure of sentences, and seem to motivate a rejection of associativity.

Commutative composition may also have a place in linguistic analysis. Composition of different gestures in sign language may run in parallel, with no natural ordering. This kind of composition might be best modelled as distinct from the temporal ordered composition of different sign units. In this case, we have reason to admit *two* forms of composition, a situation we will see more of later.

[52]Sometimes you see it claimed that (WI) is required for the contraction consecution, this is true in the presence of associativity, but can fail outside that context. The rule (WI) corresponds to the validity of $A \wedge (A \rightarrow B) \vdash B$. It does not correspond to the validity of any consecution in the $\rightarrow$ only fragment of the language.

and the structural rule of *permutation* (C) gives the permutation axiom.

$$
\dfrac{X(Y_1, (Y_2, Z)) \vdash D}{X(Y_2, (Y_1, Z)) \vdash D} \text{ [C]}
\qquad
\dfrac{
  \dfrac{
    A \vdash A \quad
    \dfrac{
      \dfrac{B \vdash B \quad C \vdash C}{B \to C, B \vdash C} (\to L)
    }{(A \to (B \to C), A), B \vdash C} (\to L)
  }{
    \dfrac{
      \dfrac{(A \to (B \to C), B), A \vdash C}{A \to (B \to C), B \vdash A \to C} (\to R)
    }{A \to (B \to C) \vdash B \to (A \to C)} (\to R)
  }{} \;(\mathsf{C})
}{}
$$

Finally (for this brief excursus into the effect of structural rules) the *mingle* rule (M) has been of interest to the relevant logic community. It is the converse of WI contraction, and a special instance of weakening (K). It corresponds to the mingle consecution $A \vdash A \to A$, whose addition to R results in the well-behaved system RM. We will consider models of RM in the next section.

$$
\dfrac{X(Y) \vdash C}{X(Y,Y) \vdash C} \text{ [M]}
\qquad
\dfrac{
  \dfrac{
    \dfrac{A \vdash A}{A, A \vdash A}(\mathsf{M})
  }{A \vdash A \to A}(\to R)
}{}
$$

There are many different structural rules which feature in different logics for different purposes. Table 1 contains some prominent structural rules. I use the notation $X \Leftarrow Y$ to stand for the structural rule

$$
\dfrac{Z(X) \vdash A}{Z(Y) \vdash A}
$$

You can replace Y by X (reading the proof upwards) in any context in an antecedent.

This proliferation of options concerning structural rules leaves us with the issue of how to choose between them. In some cases, such as Lambek's analysis of typing regimes on languages, the domain is explicit enough for the appropriate structural rules to be "read off" the objects being modelled. In the case of finding an appropriate logic of *entailment*, the question is more fraught. Anderson and Belnap's considerations in favour of the logic E are by no means the only choices available for a relevantist. Richard Sylvan's depth relevant program [2000; 1982] and Brady's constraints of *concept containment* [1988; 1996] motivate logics much weaker than E. They motivate logics without weakening, commutativity, associativity and contraction.

Let's return to Lambek, after that excursus on structural rules. In one of his early papers, Lambek considered adding conjunction to his calculus with these rules [Lambek, 1961].

$$
\dfrac{X \vdash A \quad X \vdash B}{X \vdash A \wedge B} \,[\wedge R]
\qquad
\dfrac{X(A) \vdash C}{X(A \wedge B) \vdash C} \,[\wedge L_1]
\qquad
\dfrac{X(B) \vdash C}{X(A \wedge B) \vdash C} \,[\wedge L_2]
$$

Adding disjunction with dual rules is also straightforward.

$$
\dfrac{X \vdash A}{X \vdash A \vee B} \,[\vee R_1]
\qquad
\dfrac{X \vdash B}{X \vdash A \vee B} \,[\vee R_2]
\qquad
\dfrac{X(A) \vdash C \quad X(B) \vdash C}{X(A \vee B) \vdash C} \,[\wedge R]
$$

Name	Label	Rule		
Associativity	B	$X,(Y,Z)$	$\Leftarrow$	$(X,Y),Z$
Twisted Associativity	B′	$X,(Y,Z)$	$\Leftarrow$	$(Y,X),Z$
Converse Associativity	B^c	$(X,Y),Z$	$\Leftarrow$	$X,(Y,Z)$
Strong Commutativity	C	$(X,Y),Z$	$\Leftarrow$	$(X,Z),Y$
Weak Commutativity	CI	X,Y	$\Leftarrow$	Y,X
Strong Contraction	W	$(X,Y),Y$	$\Leftarrow$	X,Y
Weak Contraction	WI	X,X	$\Leftarrow$	X
Mingle	M	X	$\Leftarrow$	X,X
Weakening	K	X	$\Leftarrow$	X,Y
Commuted Weakening	K′	X	$\Leftarrow$	Y,X

Table 1. Structural Rules

Conjunctive and disjunctive types have clear interpretations in the calculus of syntactic types. In English, and is promiscuous. It conjoins sentences, names, verbs, and other things. It makes sense to say that it has a *conjunctive type*

$$\text{and} \Vdash ((a_1 \rightarrow a_1) \leftarrow a_1) \wedge \cdots \wedge ((a_n \rightarrow a_n) \leftarrow a_n)$$

for n types a_i.[53] Disjunctive types also have a simple interpretation. Conjunction and disjunction motivate the following type-assignment clauses:

○ $x \Vdash A \wedge B$ if and only if $x \Vdash A$ and $x \Vdash B$.

○ $x \Vdash A \vee B$ if and only if $x \Vdash A$ or $x \Vdash B$.

Lambek's rules for conjunction and disjunction are satisfied under this interpretation of their behaviour. Lambek's rules are *sound* for this interpretation.

Cut is still admissible with the addition of these rules. It is straightforward to permute cuts past these rules, and to eliminate conjunctions introduced simultaneously by both. However, the addition results in the failure of distribution. The traditional proof of distribution (in Figure 2) requires *both* contraction and weakening. This means that the simple rules for conjunction and disjunction (in the context of this proof theory, including its structural rules) are *incomplete* for the intended interpretation.

---

[53]*Actually* it makes sense to think of and as having type $\forall p((p \rightarrow p) \leftarrow p)$. However, propositionally quantified Lambek calculus is a wide-open field. No-one that I know of has explored this topic, at the time of writing.

$$\dfrac{\dfrac{A \vdash A}{A, B \vdash B}\,(\mathsf{K}) \qquad \dfrac{B \vdash B}{A, B \vdash B}\,(\mathsf{K})}{\dfrac{A, B \vdash A \wedge B}{A, B \vdash (A \wedge B) \vee (A \wedge C)}\,(\vee R)}\,(\wedge R) \qquad \dfrac{\dfrac{A \vdash A}{A, C \vdash C}\,(\mathsf{K}) \qquad \dfrac{C \vdash C}{A, C \vdash C}\,(\mathsf{K})}{\dfrac{A, C \vdash A \wedge C}{A, C \vdash (A \wedge B) \vee (A \wedge C)}\,(\vee R)}\,(\wedge R)$$

$$\dfrac{\dfrac{\dfrac{A, B \vee C \vdash (A \wedge B) \vee (A \wedge C)}{A, A \wedge (B \vee C) \vdash (A \wedge B) \vee (A \wedge C)}\,(\wedge L)}{A \wedge (B \vee C), A \wedge (B \vee C) \vdash (A \wedge B) \vee (A \wedge C)}\,(\wedge L)}{A \wedge (B \vee C) \vdash (A \wedge B) \vee (A \wedge C)}\,(\mathsf{WI})$$

Figure 2. Proof of Distribution of $\wedge$ over $\vee$

2.6   *Kripke's Decidability Technique for* $\mathrm{R}[\to, \wedge]$

Lambek's proof theory for the calculus of syntactic types has a close cousin, for the relevant logic R. Within a year of Lambek's publication of his calculus of types, Saul Kripke published a decidability result using a similar Gentzen system for the implication fragments of the relevant logics R and E [1959]. Kripke's results extend without much modification to the implication and conjunction fragments of these logics, and less straightforwardly to the implication, negation fragment [1961; 1967; 1975] or to the whole logic *without distribution* [Meyer, 1966] (Meyer christened the resulting logic LR for *lattice* R). I will sketch the decidability argument for the implication and conjunction fragment $\mathrm{R}[\to, \wedge]$, and then show how LR can be embedded within $\mathrm{R}[\to, \wedge]$, rendering it decidable as well.

The technique uses the Gentzen proof system for $\mathrm{R}[\to, \wedge]$, which is a version of the Gentzen systems seen in the previous section. It uses the same rules for $\to$ and $\wedge$, and it is modified to make it model the logic R. We have the structural rules of associativity and commutativity (which we henceforth ignore, taking antecedents of consecutions to be *multisets* of formulas[54]). We add also the structural rule WI of *contraction*. Cut is eliminable from this system, using standard techniques. However, the decidability of the system is not straightforward, given the presence of the rule WI. WI makes proof-search fiendishly difficult. The main strategy of the decision procedure for $\mathrm{R}[\to, \wedge]$ is to limit applications WI in order to prevent a proof search from running on forever in the following way: "Is $p \vdash q$ derivable? Well it is if $p, p \vdash q$ is derivable. Is $p, p \vdash q$ derivable? Well it is if $p, p, p \vdash q$ is ..."

We need one simple notion before this strategy can be explained. We will say that the consecution $X' \vdash A$ is *a contraction* of $X \vdash A$ just in case $X' \vdash A$ can be derived from $X \vdash A$ by (repeated) applications of the structural rules. (This means *contraction*, in ef-

---

[54]A multiset is a set-like structure, in which entities may appear more than once. A multiset of formulas may represented by a function f from the set of all formulas to the set of natural numbers, $f(A)$ is the number of times A is a member of the multiset represented by f.

fect, if you take the structures X and X' to be multisets, identifying different permutations and associations of the formulas therein.) Kripke's plan is to drop the WI, replacing it by building into the connective rules a limited amount of contraction.

More precisely, the idea is to allow a contraction of the conclusion of an connective rule only in so far as the same result could not be obtained by first contracting the premises. A little thought shows that this means no change for the rules $(\to R)$, $(\wedge L)$ and $(\wedge R)$, and that the following is what is needed to modify $(\to L)$.

$$\frac{X \vdash A \quad Y, B \vdash C}{[X, Y, A \to B] \vdash C} \ [\to L']$$

where $[X, Y, A \to B]$ is any contraction of $X, Y, A \to B$ such that

○ $A \to B$ occurs only 0, 1, or 2 times fewer than in $X, Y, A \to B$;

○ Any formula other than $A \to B$ occurs only 0 or 1 time fewer.

It is clear that after modifying the system $R[\to, \wedge]$ by building some limited contraction into $(\to L')$ in the manner just discussed, the following lemma is provable by an induction on length of derivations:

LEMMA 14 (Curry's Lemma). *If a consecution $X' \vdash A$ is a contraction of a consecution $X \vdash A$, and $X \vdash A$ has a derivation of length n, then $X' \vdash A$ has a derivation of length no greater than n.*[55]

This shows that the modification of the system leaves the same consecutions derivable (since the lemma shows that the effect of contraction is retained). For the rest of this section we will work in the modified proof system.

Curry's Lemma also has the corollary that every derivable consecution has an *irredundant* derivation: that is, a proof containing no branch with a consecution $X' \vdash A$ below a sequent $X \vdash A$ of which it is a contraction.

Now we can describe the decision procedure. Given a consecution $X \vdash A$, you test for provability by building a proof search tree: you place above $X \vdash A$ all possible premises or pairs of premises from which $X \vdash A$ follows by one of the rules. Even though we have built some contraction into one rule, this will be only a finite number of consecutions. This gives a tree. If a proof of the consecution exists, it will be formed as a subtree of this proof search tree. By Curry's Lemma, the proof search tree can be made irredundant. The tree is also finite, by the following lemma.

LEMMA 15 (König's Lemma). *A tree with finitely branching tree with branches of finite length is itself finite.*

We have already proved that the tree is finitely branching (each consecution can have only finitely many possible ancestors). The question of the length of the branches remains open, and this is where Kripke proved an important lemma. To state it we need an idea from Kleene. Two consecutions $X \vdash A$ and $X' \vdash A$ are *cognate* just when exactly the

---

[55]The name comes from Anderson and Belnap [1975], who note that it is a modification of a lemma due to Curry [1950], applicable to classical and intuitionistic Gentzen systems.

same formulas X are in X'. The class of all consecutions cognate to a given consecution is called a *cognation class*. Now we can state and prove Kripke's lemma.

LEMMA 16 (Kripke's Lemma). *There is no infinite sequence of cognate consecutions such that no earlier consecution is a contraction of a later consecution in the sequence.*

This means that the number of cognation classes occurring in any derivation (and hence in each branch) is finite. But Kripke's Lemma also shows that only a finite number of members of each cognation class occur in a branch (this is because we have constructed the complete proof search tree to be irredundant). So every branch is finite, and so both conditions of König's Lemma hold. It follows that the complete proof search tree is finite and so there is a decision procedure. So, a proof of Kripke's Lemma concludes our search for a decision procedure for $R[\wedge, \rightarrow]$.

Proof. This is not a complete proof of Kripke's Lemma. (The literature contains some clear expositions [Anderson and Belnap, 1975; Belnap and Wallace, 1961].) As Dunn showed [1986] kernel idea can be seen in a picture. As a special case, consider consecutions cognate to $X, Y \vdash A$. Each such consecution can be depicted as a point in the upper right-hand quadrant of the plane, marked with the origin at $(1, 1)$ rather than $(0, 0)$ since $X, Y \vdash A$ is the minimal consecution in the cognation class. So, $X, X, Y, Y, Y, Y \vdash A$ is represented as $(2, 4)$: '2 X units' and '4 Y units'. Now given any initial consecution, for example

$$(\Gamma_0) \quad X, X, X, Y, Y \vdash A$$

you might try to build an irredundant sequence by first inflating the number of Ys (for purposes of keeping on the page we let this be to 5 rather than 3088). But then, you have to decrement number of Xs at least by one. The result is depicted in Figure 3 for the first two members of the sequence Γ_0, Γ_1.

The purpose of the intersecting lines at each point is to mark off areas (shaded in the diagram) into which no further points of the sequence may be placed. If Γ_2 were placed at the point $(6, 5)$, it would reduce to Γ_0. This means that each new point must proceed either one unit closer to the X axis or one unit closer to the Y axis. After a finite number of choices the consecutions will arrive at one or other of the two axes, and then after a time, you will arrive at the other. At that time, no more additions can be made, keeping the sequence irredundant.

This proof sketch generalises to n-dimensional space, corresponding to an initial consecution with n different antecedent parts. The only difficulty is in drawing the pictures.[56]

∎

Extending this result to the whole of R (without distribution) is not difficult. You can amend the proof system to manipulate consecutions with structure on the *right* as well as on the left. (I won't present the modification of the rules here because they are the same as the rules for those connectives in *linear logic* which we will see in a few sections

---

[56]Meyer discovered that Kripke's Lemma is equivalent to Dickson's Theorem about primes: Given any set of natural numbers all of which are composed out of the first m primes (that is, every member has the form $p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$) if no member of this set has a proper divisor in the set, then the set is finite.
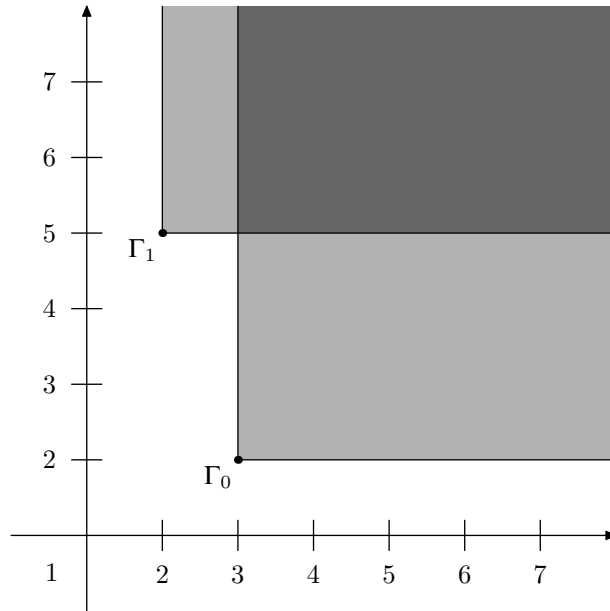
Figure 3. Descending Regions

time.) The system will not prove the distribution of conjunction over disjunction, but an explicit decision procedure for the whole logic can be found. This result is due to Meyer, and can be first found in his dissertation from [1966]. Meyer also showed how LR can be embedded in $R[\rightarrow, \wedge]$ by translation. Meyer's translation is fairly straightforward, but I will not go through the details here.[57] I will sketch a simpler translation which comes from the Vincent Danos' more recent work on linear logic [1990; 1995], and which is a simple consequence of the soundness and completeness of phase space models. We translate formulas in the language of LR into the language of implication and negation by picking a particular distinguished proposition in the target language and designating that as f. Then we define $\sim$ in the language of $R[\rightarrow, \wedge]$ by setting $\sim A$ to be $A \rightarrow f$. Then the rest of the translation goes as follows:

$$
\begin{aligned}
p^t &= \sim\sim p \\
(A \wedge B)^t &= \sim\sim(A^t \wedge B^t) \\
(A \vee B)^t &= \sim(\sim A^t \wedge \sim B^t) \\
(A \circ B)^t &= \sim(A^t \rightarrow \sim B^t) \\
(A \rightarrow B)^t &= A^t \rightarrow B^t \\
(\sim A)^t &= \sim A^t.
\end{aligned}
$$

---

[57]The details of the translation can be found elsewhere [Dunn, 1986; Dunn and Restall, 2001]. The point which makes the translation a little more complex than the translation I use here is the treatment of f and its negation t.

I will not go through the proof of the adequacy of this translation, as we will see it when we come to look at phase spaces. However, a direct demonstration of its adequacy (without an argument taking a detour through models) is possible.[58] Given this translation, any decision procedure for $R[\rightarrow, \wedge]$ transforms into a decision procedure for the whole of LR.

McRobbie and Belnap [1979] have translated the implication negation fragment of the proof theory in an analytic tableau style, and Meyer has extended this to give analytic tableau for linear logic and other systems in the vicinity of R [Meyer *et al.*, 1995]. Neither time nor space allows me to consider tableaux for substructural logics, except for this reference.

Some recent work of Alasdair Urquhart has shown that although $R[\rightarrow, \wedge]$ is decidable, it has *great* complexity [1990; 1997]: There is no primitive recursive bound on either the time or the space taken by a computation deciding any formula. Urquhart follows some work in linear logic [1992] by using the logic to encode the behaviour of a *branching counter machines*. A counter machine has a finite number of *registers* (say, r_i for suitable i) which each hold one non-negative integer, and some finite set of possible *states* (say, q_j for suitable j). Machines are coded with a list of instructions, which enable you to *increment* or *decrement* registers, and test for registers' being zero. A *branching* counter machine dispenses with the test instructions and allows instead for machines to take multiple execution paths, by way of *forking* instructions. The instruction $q_i + r_j q_k$ means "when in q_i, add 1 to register r_j and enter stage q_k," and $q_i - r_j q_k$ means "when in q_i, subtract 1 from register r_j (if it is non-empty) and enter stage q_k," and $q_i f q_j q_k$ is "when in q_i, fork into two paths, one taking state q_j and the other taking q_k."

A machine configuration is a state, together with the values of each register. Urquhart uses the logic LR to simulate the behaviour of a machine. For each register r_i, choose a distinct variable R_i, for each state q_j choose a distinct variable Q_j. The configuration $\langle q_i; n_1, \ldots, n_l \rangle$, where n_i is the value of r_i is the formula

$$Q_i \circ R_1^{n_1} \circ \cdots \circ R_l^{n_l}$$

(where A^n is the *n*-fold self-fusion of *A*) and the instructions are modelled by sequents in the Gentzen system, as follows:

Instruction	Sequent
$q_i + r_j q_k$	$Q_i \vdash Q_k \circ R_j$
$q_i - r_j q_k$	$Q_i, R_j \vdash Q_k$
$q_i f q_j q_k$	$Q_i \vdash Q_j \vee Q_k.$

Given a machine program (a set of instructions) we can consider what is provable from the sequents which code up those instructions. This set of sequents we can call the *theory* of the machine. $Q_i \circ R_1^{n_1} \circ \cdots \circ R_l^{n_l} \vdash Q_j \circ R_1^{m_1} \circ \cdots \circ R_l^{m_l}$ is intended to mean that from

---

[58]The nicest is due to Danos. Take a proof of $X \vdash Y$ in the calculus for LR and translate it step by step into a proof of $X^t, \sim Y^t \vdash f$. (Here $\sim Y^t$ is the collection of the *negations* of the *translations* of each of the elements of *Y*.) The translation here is exactly what you need to make the rules correspond (*modulo* a few applications of Cut).

state configuration $\langle q_i; n_1, \ldots, n_l \rangle$ all paths will go through configuration $\langle q_j; m_1, \ldots, m_l \rangle$ after some number of steps.

A branching counter machine *accepts* an initial configuration if when run on that configuration, all branches terminate at the final state q_f, with all registers taking the value zero. The corresponding condition in LR will be the provability of

$$Q_i \circ R_1^{n_1} \circ \cdots \circ R_l^{n_l} \vdash Q_m.$$

This will *nearly* simulate branching counter machines, except for the fact that in LR we have $A \vdash A \circ A$. This means that each of our registers can be incremented as much as you like, provided that they are non-zero to start with. This means that each of our machines need to be equipped with every instruction of the form $q_i > 0 + r_j q_i$, meaning "if in state q_i, add 1 to r_j, provided that it is already nonzero, and remain in state q_i."

Urquhart is able to prove that a configuration is accepted in a branching counter machine, if and only if the corresponding sequent is provable from the theory of that machine. But this is equivalent to a formula

$$\bigwedge \text{Theory}(M) \wedge t \rightarrow (Q_1 \rightarrow Q_m)$$

in the language of LR. It is then a short step to our complexity result, given the fact that there is no primitive recursive bound on determining acceptability for these machines. Once this is done, the translation of LR into the conjunction and implication fragment of R gives us our complexity result.

Despite this complexity result, Kripke's algorithm has been implemented with quite some success. The theorem prover `Kripke`, written by McRobbie, Thistlewaite and Meyer, implements Kripke's decision procedure, together with some quite intelligent proof-search pruning, by means of finite models. This implementation *works* in many cases [Thistlewaite *et al.*, 1988]. Clearly, work must be done to see whether the horrific complexity of this problem in general can be transferred to results about *average case* complexity.

2.7   Richer Structures: Gentzen Systems for Distribution

Grigori Minc [1972; 1977] and J. Michael Dunn [1973] independently developed a Gentzen-style consecution calculus for relevant logics in the vicinity of R. As we have seen in the Gentzen calculus for R[$\rightarrow$], the distinctive behaviour of implication arises out of the presence or absence of structural rules governing the combination of premises. To find a logic without the paradoxes of implication, we are led to reject the structural rule of *weakening*. However, the structural rule of weakening is *required* to prove distribution.[59] Dunn and Minc's innovations were to see a way around this apparent impasse. One way to think of the problem is this: consider the proof of distribution in Figure 2 on page 325. Focus on the point at which ($\vee L$) is applied. The proof moves from $A, B \vdash \cdots$ and

---

[59]At least, it is required if the proof is going to be anything like the proof of distribution in *standard* Gentzen systems.

$A, C \vdash \cdots$ to $A, B \vee C \vdash \cdots$. It is *this* point at which *some* form of distribution has just been used: we have used the *disjunction* rule inside a *comma* context. This makes disjunction distribute over whatever the comma represents. In the case where comma is the metalinguistic analogue of fusion (as it is in these proof systems) we can prove $A \circ (B \vee C) \vdash (A \circ B) \vee (A \circ C)$. We cannot prove the distribution of extensional conjunction over disjunction simply because there is no structure able to represent conjunction in this proof system.[60] The solution to provide distribution is then to allow a structure to represent *extensional* conjunction, just as there is a structural analogue for *intensional* conjunction.

In a proof system like this, we define structures recursively, allowing both intensional and extensional conjunction.

- ○ A formula is a structure.
- ○ If X and Y are structures, so is $(X; Y)$. This is the *intensional* combination of X and Y.
- ○ If X and Y are structures, so is (X, Y). This is the *extensional* combination of X and Y.

Then *all* of the traditional structural rules (B, C, K, W) are admitted for *extensional* combination, and only a weaker complement (say, omitting K, for relevant logic, or all but associativity, for the Lambek calculus, or some other menu of choices for some other substructural logic) are admitted for *intensional* combination.

The rules for the connectives may remain unchanged (apart from the notational variation ";" for intensional combination, instead of "," which was used up until this point). However, the rules for conjunction may be varied to match those for fusion: we can instead take extensional conjunction to be explicitly paired with extensional combination.

$$\frac{X(A, B) \vdash C}{X(A \wedge B) \vdash C} \, [\wedge \mathrm{L}'] \qquad \frac{X \vdash A \quad Y \vdash B}{X, Y \vdash A \wedge B} \, [\wedge \mathrm{R}']$$

These rules are admissible, given the original structure-free rules, as these demonstrations show.[61]

$$\frac{\dfrac{\dfrac{X(A, B) \vdash C}{X(A \wedge B, B) \vdash C} \, (\wedge \mathrm{L})}{X(A \wedge B, A \wedge B) \vdash C} \, (\wedge \mathrm{L})}{X(A \wedge B) \vdash C} \, (\mathsf{WI}) \qquad \frac{\dfrac{X \vdash A}{X, Y \vdash A} \, (\mathsf{K}) \quad \dfrac{\dfrac{Y \vdash B}{Y, X \vdash B} \, (\mathsf{K})}{X, Y \vdash B} \, (\mathsf{C})}{X, Y \vdash A \wedge B} \, (\wedge \mathrm{R})$$

The modified proof theory is sound and complete for the relevant logic R and its neighbours. The cut elimination proof works as before — even with richer structures, the

[60]That is a simplification. The proof could be dualised, and work in a proof system with single antecedent and multiple consequents, for a dual intuitionistic logic. In this case it is the structure for extensional disjunction which would distribute over the *conjunction* rule. The relevant part of a proof would be the move from $\cdots \vdash A, B$ and $\cdots \vdash A, C$ to $\cdots \vdash A, B \wedge C$, distributing a conjunction over a disjunction again.

[61]The converse proofs, to the effect that $(\wedge \mathrm{L})$ and $(\wedge \mathrm{R})$ are admissible in the presence of $(\wedge \mathrm{L}')$ and $(\wedge \mathrm{R}')$ are just as simple.

conditions of the cut elimination proof (permutability of cut with other rules, eliminability of matching principal formulas) are still satisfied. The subformula property is also satisfied (formulas appearing in a proof of a consecution must be subformulas of those in the consecution proved) and the proof theory is well-behaved.

However, the beneficial consequences of a cut-free Gentzen system for a logic — its decidability — is not always available. The difficulty is the presence of contraction for extensional combination. This is not surprising, because as we will see later, R is undecidable. You cannot extract a decision procedure from its Gentzen calculus. However, in the absence of *expansive* rules such as W and WI, a decision procedure can be found, as Steve Giambrone found in the early 1980s (see [1985]). Giambrone's decidability argument for the negation-free fragment of R without contraction (which, we will see, is equivalent to linear logic with distribution added) and also for positive TW. Ross Brady extended this argument in the early 1990s to show that RW and TW are decidable [1991]. Brady's technique involved extending the Gentzen system with *signed* formulas, to give straightforward rules for negation without resorting to a multiple consequent calculus.

Other extensions to this proof theory are possible for different applications. Belnap, Dunn and Gupta extended Dunn's original work to model R with an S4-style necessity [Belnap *et al.*, 1980]. I have shown how a system like this one can be used to motivate an extension of the Lambek calculus which is sound and complete for its intended interpretation of conjunction and disjunction on frame models [Restall, 1994] (unlike the structure-free rules which Lambek originally proposed).

The natural deduction analogue of the Gentzen system has been the focus of much attention, too. Read uses the natural deduction system as the basis of his presentation of R in his *Relevant Logic* [1988]. Slaney, in an influential article from 1990 [1990] gives a philosophical defence of the two different sorts of bunching operators, characterising *extensional* combination of bodies of information as a monotonic *lumping* of information together, while taking *intensional* combination of X with Y (that's $X; Y$) as the *application* of X to Y. This distinction motivates the rules for implication ($X \vdash A \rightarrow B$ iff $X; A \vdash B$: $A \rightarrow B$ follows from X just when whenever you *apply* X to A, the resulting information gives B).

O'Hearn and Pym call this kind of proof theory the logic of *bunched implications* [1999], and they use it to model computation.

2.8   Display Logic

Nuel Belnap's *Display Logic* [1982] is a neat, uniform method for providing a cut-free consecution calculus for a wide range of formal systems. The central ideas of Belnap's *Display Logic* are simple and elegant. Like other consecution proof theories, the calculus deals with structured collections of formulas, *consecutions*. In display logic, consecutions are of the form $X \vdash Y$, where X and Y are *structures*, made up from formulas. Structures are made up of structure-connectives operating on structures, building up structures from smaller structures, in much the same way as formulas are built up by formula-connectives. The base level of structures are the formulas. So far, display logic is of a piece with

standard Gentzen systems — in traditional systems structures are simply lists, and in the more *avant garde* systems of Dunn and Minc, structures can be made up of two bunching operators — but in Belnap's work, structures can be even richer. This richness is present so that consecutions can support the *display property*: any substructure of a consecution can be *displayed* to be the entire antecedent or consequent of an equivalent consecution. In general, what is wanted is a way to "unravel" a context like so that we can perform equivalences such as this:

$$X(Y) \vdash Z \text{ is equivalent to } Y \vdash X^{-1}(Z)$$

where the Y inside the structure $X(Y)$ is exposed to view, and the surrounding $X(—)$ context is unravelled. Once you can do this, connective rules are simple, because you can assume that each formula is displayed to be the entire antecedent or consequent of a consecution.

Belnap's original work on display logic was motivated by the problem for finding a natural proof theory for relevant logics and their neighbours. As a result, it is illustrative to see the choices he made in constructing rules to allow the display of substructures. Here are some equivalences present in R and weaker relevant logics.

$$\frac{A \circ B \vdash C}{A \vdash {\sim}B + C} \qquad \frac{\dfrac{A \vdash B + C}{A \circ {\sim}B \vdash C}}{A \vdash C + B} \qquad \frac{\dfrac{A \vdash B}{{\sim}B \vdash {\sim}A}}{{\sim}{\sim}A \vdash B}$$

These equivalences allow us to "get under" the connectives in formulas. Here, the equivalences govern fusion, fission and negation. In traditional Gentzen systems, the "comma" is an overloaded operator, signifying *conjunction* in antecedent position and *disjunction* in consequent position. That is, a consecution of the form $X \vdash Y$ is interpreted as saying something like: "if everything in X is true, something in Y must be true." In substructural logics, this comma (the one which also governs the behaviour of implication) is interpreted as *fusion* on the left, and if it appears on the right at all, as *fission*. Belnap noted that we could get the display property if you add a structural connective for *negation*. If you write this connective with an asterisk, you get the following *display postulates* to parallel the facts we have already seen, governing fusion and fission.

$$\frac{X \circ Y \vdash Z}{X \vdash {*}Y \circ Z} \qquad \frac{\dfrac{X \vdash Y \circ Z}{X \circ {*}Y \vdash Z}}{X \vdash Z \circ Y} \qquad \frac{\dfrac{X \vdash Y}{{*}Y \vdash {*}X}}{{*}{*}X \vdash Y}$$

(Belnap uses "∘" for the *structure connective* which is fusion- and fission-like, and I will follow him in this notation.) As before, structures can be interpreted in "antecedent" position or in "consequent" position. However, now we can have "∘" representing fusion on the *right* of the turnstile, or fission on the *left*, because the *negation* operator flips structures from one position to another. Consider the equivalence of $X \circ Y \vdash Z$ with $X \vdash {*}Y \circ Z$. In the first consecution, the structure Y is on the left of the turnstile, but on the second

it is on the right. It must have the same content in both cases[62] which means that the structure connectives inside Y must be interpreted in the same way. With this caveat, the display calculus is a straightforward Gentzen-system with structure connectives allowing both positive and negative information. The rules governing the connectives are straightforward analogs of the traditional rules, with the simplification that we can now assume that principal formulas are the entire antecedent or consequent of the consecutions which introduce them. Here are the conditional rules:

$$\frac{X \circ A \vdash B}{X \vdash A \to B} \qquad \frac{X \vdash A \quad B \vdash Y}{A \to B \vdash *X \circ Y.}$$

The display postulates mean that the cut rule appropriate for a display calculus can be stated exceedingly simply:

$$\frac{X \vdash A \quad A \vdash Y}{X \vdash Y.} \; [Cut]$$

There is no need for a stronger rule placing the cut-formula in a context, because we can always assume that the cut formula has been displayed. This is an advance in the proof theories of substructural logics because some of the various strengthenings of the cut rule, required to prove the cut-elimination theorem, are not valid in some substructural systems.[63] In his original paper, Belnap provides a list of eight easily checked conditions. If a display proof theory satisfies these conditions, then Cut is admissible in the system. We need explore the detail of these conditions here.[64]

Belnap shows that different logical systems can be given by adding different structural rules governing the display connectives — and that furthermore, the one proof system can have more than one *family* of display connectives. This parallels the Dunn-Minc Gentzen system for logics with distribution. Belnap shows how you can construct proof theories for relevant logics, modal logics, intuitionistic logic, and logics which combine connectives from different families.

The idea of using display postulates to provide proof theories for different connectives is not restricted to Belnap's original family featuring a binary operator $\circ$ and a unary $*$. Wansing [1994] extended Belnap's original work showing that a unary structure $\bullet$ with display rules

$$\frac{\bullet X \vdash Y}{X \vdash \bullet Y}$$

would suffice to model normal modal logics. The corresponding connective rules for $\Box$

---

[62]It is justified by the equivalence of $A \circ B \vdash C$ with $A \vdash \sim B + C$, and in this case B "means the same thing" in both cases.

[63]The most generous case of *Mix* — from $X \vdash Y(A)$ and $X'(A) \vdash Y'$ to some conclusion, where both $Y(A)$ and $X'(A)$ involve multiple occurrences of A to be eliminated — seems to have no appropriate valid conclusion in general substructural logics.

[64]I have generalised Belnap's conditions for the admissibility of cut in such a way as to include traditional consecution systems as well as display logics. It remains unclear if this generalisation will prove useful in practice, but it does seem to be an advantage to not have to prove the cut elimination theorem again and again for each proof system you construct [Restall, 2000a, ch. 6].

are

$$\frac{X \vdash \bullet A}{X \vdash \Box A} \ [\Box R] \qquad \frac{A \vdash X}{\Box A \vdash \bullet X} \ [\Box L]$$

This shows that $\Box$ is the object-language correlate of $\bullet$ in *consequent* position.[65] As a result, display logic has been used outside its original substructural setting. Wansing has shown that display logic is a natural home for proof theory for classical modal logics [1994; 1998], Belnap has extended his calculus to model Girard's linear logic [1990],[66] Goré and I have used the display calculus to model substructural logics other than those considered by Belnap [1998; 1995], and I have extracted some decidability results in the vein of Giambrone and Brady [1998].[67]

2.9 *Linear Logic*

Girard, in 1987, introduced *linear logic*, a particular substructural system that allows commuting and reassociating of premises, but no contraction or weakening [Girard, 1987a]. Perhaps Girard's major innovation in linear logic is the introduction of the *modalities* — the *exponentials*[68] which allow the recovery of these structural rules in a limited, controlled fashion. Linear logic has a straightforward *resource* interpretation: when premises and conclusions are taken to be resources to be used in proof, then the absence of contraction indicates that resources cannot be duplicated, and the absence of weakening indicates that resources cannot be simply thrown away. Only particular kinds of resources — those marked off by the exponentials — can be treated in this manner. Linear logic has received a great deal of attention in the literature in theoretical computer science.

Gentzen Systems

The most straightforward proof theory for linear logic is a consecution system where consecutions feature structure in the antecedent and the consequent:

$$A \vdash A \qquad \frac{X \vdash Y, A \quad X', A \vdash Y'}{X', X \vdash Y, Y'} \ [\text{Cut}]$$

$$\frac{X \vdash A, Y}{X, {\sim}A \vdash Y} \ [{\sim}\text{L}] \qquad \frac{X, A \vdash Y}{X \vdash {\sim}A, Y} \ [{\sim}\text{R}]$$

$$\frac{X, A \vdash Y}{X, A \wedge B \vdash Y} \ [\wedge\text{L}_1] \qquad \frac{X, B \vdash Y}{X, A \wedge B \vdash Y} \ [\wedge\text{L}_2] \qquad \frac{X \vdash Y, A \quad X \vdash Y, B}{X \vdash Y, A \wedge B} \ [\wedge\text{R}]$$

---

[65]Its partner in *antecedent* position is a possibility operator, but the *dual* possibility operator which looks *backwards* down the accessibility relation for necessity. It is tied together with $\Box$ by the display postulates $A \vdash \Box B$ if and only if $\blacklozenge A \vdash B$.

[66]The issue is the treatment of the exponentials.

[67]However, Kracht has shown that in general, decidability results from a display calculus are not to be expected. He has shown that it is *undecidable* whether a given displayed modal logic is decidable [1996].

[68]So called because of the equivalence between $!(A \wedge B)$ and $!A \circ !B$, and dually, between $?(A \vee B)$ and $?A + ?B$. This also explains why $\wedge$ and $\vee$ are the *additives* and $\circ$ and $+$ are the *multiplicatives* in the parlance of linear logic: in numbers, $x^{y+z} = x^y \times x^z$.

$$\frac{X, A, B \vdash Y}{X, A \circ B \vdash Y} \; [\circ\text{L}] \qquad \frac{X \vdash Y, A \quad X' \vdash B, Y'}{X, X' \vdash Y, A \circ B, Y'} \; [\circ\text{R}]$$

$$\frac{X, A \vdash Y \quad X, B \vdash Y}{X, A \vee B \vdash Y} \; [\vee\text{L}] \qquad \frac{X \vdash Y, A}{X \vdash Y, A \vee B} \; [\vee\text{R}_1] \qquad \frac{X \vdash Y, B}{X \vdash Y, A \vee B} \; [\vee\text{R}_2]$$

$$\frac{X, A \vdash Y \quad B, X' \vdash Y'}{X, A + B, X' \vdash Y, Y'} \; [+\text{L}] \qquad \frac{X \vdash Y, A, B}{X \vdash Y, A + B} \; [+\text{R}]$$

$$\frac{X \vdash Y, A \quad X', B \vdash Y'}{X', X, A \rightarrow B \vdash Y, Y'} \; [\rightarrow\text{L}] \qquad \frac{X, A \vdash B, Y}{X \vdash A \rightarrow B, Y} \; [\rightarrow\text{R}]$$

$$\frac{X \vdash Y}{X, t \vdash Y} \; [t\text{L}] \qquad \vdash t \; [t\text{R}]$$

$$f \vdash \; [f\text{L}] \qquad \frac{X \vdash Y}{X \vdash Y, f} \; [f\text{R}]$$

$$X, \bot \vdash Y \; [\bot\text{L}] \qquad X \vdash Y, \top \; [\top\text{R}]$$

$$\frac{X \vdash Y}{X, !A \vdash Y} \; [\text{K!}] \quad \frac{X, A \vdash Y}{X, !A \vdash Y} \; [\text{L!}] \quad \frac{!X \vdash A, ?Y}{!X \vdash !A, ?Y} \; [\text{R!}] \quad \frac{X, !A, !A \vdash Y}{X, !A \vdash Y} \; [\text{WI!}]$$

$$\frac{X \vdash Y}{X \vdash ?A, Y} \; [\text{K?}] \quad \frac{X \vdash Y, A}{X \vdash Y, ?A} \; [\text{R?}] \quad \frac{!X, A \vdash ?Y}{!X, ?A \vdash ?Y} \; [\text{L?}] \quad \frac{X \vdash Y, ?A, ?A}{X \vdash Y, ?A} \; [\text{WI?}].$$

Girard's notation for the connectives differs from the one we have chosen here. Figure 4 contains a translation manual between the three traditions we have seen so far.

Connective	Here	Lambek	Girard
Implication	$A \rightarrow B$	$A \,/\, B$	$A \multimap B$
Converse Implication	$B \leftarrow A$	$B \setminus A$	
Fusion	$A \circ B$	$A \bullet B$	$A \otimes B$
Fission	$A + B$		$A \,\invamp\, B$
Conjunction	$A \wedge B$		$A \,\&\, B$
Disjunction	$A \vee B$		$A \oplus B$
Negation	$\sim A$		$A^{\perp}$
Ackermann Truth	t		1
Ackermann Falsehood	f		$\perp$
Church Truth	$\top$		$\top$
Church Falsehood	$\bot$		0
Of course	$!$		$!$
Why not	$?$		$?$

Figure 4. Translation between notations

Linear logic has two distinctive features. First, the exponentials, which allow the recovery of structural rules. Girard in fact *discovered* linear logic as a decomposition of the intuitionistic conditional $A \supset B$ into $!A \to B$ in the models of *coherence spaces*, which we shall see in the next part of this essay. For now, it is enough to get a taste of this decomposition. The linear implication $A \to B$ indicates that one use of A is sufficient to get one instance of B. The exponential is the operator which licences arbitrary *re-use* of resources. So, an intuitionistic conditional says that the consequent B can be found, using as many instances of A as we need. Here are the proofs of the equivalence between $!(A \wedge B)$ and $!A \circ !B$.

$$\cfrac{\cfrac{\cfrac{\cfrac{A \vdash A}{A \wedge B \vdash A}(L\wedge)}{!(A \wedge B) \vdash A}(L!)}{!(A \wedge B) \vdash !A}(R!) \quad \cfrac{\cfrac{\cfrac{B \vdash B}{A \wedge B \vdash A}(L\wedge)}{!(A \wedge B) \vdash B}(L!)}{!(A \wedge B) \vdash !B}(R!)}{\cfrac{!(A \wedge B), !(A \wedge B) \vdash !A \circ !B}{!(A \wedge B) \vdash !A \circ !B}(\text{WI!})}(\circ R)$$

$$\cfrac{\cfrac{\cfrac{\cfrac{A \vdash A}{!A \vdash A}(L!)}{!A, !B \vdash A}(K!) \quad \cfrac{\cfrac{B \vdash B}{!B \vdash B}(L!)}{!A, !B \vdash B}(K!)}{\cfrac{!A, !B \vdash A \wedge B}{\cfrac{!A, !B \vdash !(A \wedge B)}{!A \circ !B \vdash !(A \wedge B)}(L\circ)}(R!)}(\wedge R)}.$$

Vincent Danos has shown that this modelling of intuitionistic logic can be made *very* intimate [1990; 1995]. It is possible to translate intuitionistic logic into linear logic in such a way that *all* intuitionistic Gentzen proofs have step-by-step equivalent linear logic proofs of their translations.

Another distinctive feature of linear logic is the pervasive presence of *duality* in the system. The presence of negation means that other connectives can be easily defined in terms of their duals. On the other hand, it is also possible to take negation as the defined connective in the following way: for each atomic formula p pick out a distinguished atomic formula to be $\sim p$. Then define $\sim A$ for complex formulas as follows:

$$
\begin{array}{rclcrcl}
\sim\sim A & \text{is} & A \\
\sim(A \wedge B) & \text{is} & \sim A \vee \sim B & \quad \sim(A \vee B) & \text{is} & \sim A \wedge \sim B \\
\sim\top & \text{is} & \bot & \quad \sim\bot & \text{is} & \top \\
\sim(A \circ B) & \text{is} & \sim A + \sim B & \quad \sim(A + B) & \text{is} & \sim A \circ \sim B \\
\sim t & \text{is} & f & \quad \sim f & \text{is} & t \\
\sim !A & \text{is} & ?\sim A & \quad \sim ?A & \text{is} & !\sim A.
\end{array}
$$

We also take $A \to B$ to be defined as $\sim A + B$ (or if you like, $\sim(A \circ \sim B)$, which is *literally* the same formula under this new regime). Together with this aspect of duality, we can also transpose consecutions from the multiple left-right variety, to a conclusion only system. We replace the consecution $X \vdash Y$ with the consecution $\vdash \sim X, Y$, where $\sim X$ is the structure containing the negations of all of the formulas in X. Then formulas are introduced only in the right, and we get a much simpler system, with one rule for every connective, as opposed to two.

$$\vdash A, \sim A \qquad \cfrac{\vdash X, A \quad \vdash \sim A, Y}{\vdash X, Y}\text{[Cut]}$$

$$\frac{\vdash X, A \quad \vdash X, B}{\vdash X, A \wedge B} \, [\wedge] \qquad \frac{\vdash X, A}{\vdash X, A \vee B} \, [\vee_1] \qquad \frac{\vdash X, B}{\vdash X, A \vee B} \, [\vee_2]$$

$$\frac{\vdash X, A \quad \vdash B, Y}{\vdash X, A \circ B, Y} \, [\circ] \qquad \frac{\vdash X, A, B}{\vdash X, A + B} \, [+]$$

$$\vdash t \, [t] \qquad \frac{\vdash X}{\vdash X, f} \, [f] \qquad \vdash X, \top \, [\top]$$

$$\frac{\vdash ?X, A}{\vdash ?X, !A} \, [!] \qquad \frac{\vdash X, A}{\vdash X, ?A} \, [?] \qquad \frac{\vdash X}{\vdash X, ?A} \, [\text{K?}] \qquad \frac{\vdash X, ?A, ?A}{\vdash X, ?A} \, [\text{WI?}]$$

Proof Nets

Consider the following two single-sided Gentzen proofs of $\vdash (A \circ B) \circ C, \sim A + \sim B, \sim C$.

$$\frac{\dfrac{\overline{} \, [\text{AX}]}{A, \sim A} \quad \dfrac{\overline{} \, [\text{AX}]}{B, \sim B}}{\dfrac{A \circ B, \sim A, \sim B}{\dfrac{A \circ B, \sim A + \sim B}{(A \circ B) \circ C, \sim A + \sim B, \sim C} \, [\circ]}} \, [+] \quad \overline{C, \sim C} \, [\text{AX}]$$

$$\frac{\dfrac{\overline{} \, [\text{AX}]}{A, \sim A} \quad \dfrac{\overline{} \, [\text{AX}]}{B, \sim B}}{\dfrac{A \circ B, \sim A, \sim B}{\dfrac{(A \circ B) \circ C, \sim A, \sim B, \sim C}{(A \circ B) \circ C, \sim A + \sim B, \sim C} \, [+]}} \quad \overline{C, \sim C} \, [\text{AX}]$$

(I elide the leading "$\vdash$" on each sequent to save space.) Notice that the two proofs here involve three axioms, two applications of the fusion rule and one application of a fission rule. They differ merely in the ordering of the rules in use. Reading the proofs from conclusion upward to the axioms, it is clear that one could decompose the fusion in $(A \circ B) \circ C$ first, as in the proof on the left; or one could decompose the fission in $\sim A + \sim B$ first, as in the proof on the right. The two steps are completely independent of one another, and it is an artefact of the Gentzen system that one must do one "before" the other. We can think of a Gentzen proof as a *serialisation* of what could be a *parallel* process. Girard invented *proof nets* as a way to characterise a parallel notion of proof, in which redundancies like these do not occur.

We will focus on the single-sided sequent calculus for the fragment of linear logic with fusion and fission. So, our rules are [AX], [$\circ$], [+], and [CUT]. We can think of each of these rules as telling us how to to construct a *proof*. Different ways to order these instructions may well end in the same target.

Consider the axiom sequent $A, \sim A$. We can think of this as a proof with two output formulas A and $\sim A$. One way to represent this is by the *graph* A—$\sim A$ with two nodes and one arc. This graph, and the corresponding graph B—$\sim B$ are combined (with an application of [$\circ$]) to form a proof with the conclusion $A \circ B, \sim A, \sim B$. One way to represent this is by connecting the A in the first graph and the B in the second to the newly introduced $A \circ B$. We construct, then, the following graph:

$$\sim A \relbar\joinrel\relbar A \qquad \sim B \relbar\joinrel\relbar B$$
$$A \circ B$$
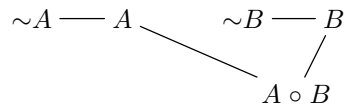
where the conclusions are $\sim A$, $\sim B$ and $A \circ B$. We will call these nodes the *ports* of this graph. The remaining formulas are intermediaries, utilised in the deduction in much the same way that the B in inference from $A \wedge B$ to $B \vee C$ via B in a familiar natural deduction system is an intermediary between premise and conclusion.

Considering the first proof of $\vdash (A \circ B) \circ C, \sim A + \sim B, \sim C$, we see that we next apply a fission rule to conclude $\vdash A \circ B, \sim A + \sim B$. This can be represented by taking our graph and connecting the two ports $\sim A$ and $\sim B$ (closing them) to form a new port $\sim A + \sim B$. So we have

$$
\begin{array}{ccccc}
\sim A \!-\! A & & \sim B \!-\! B \\[2pt]
\big\backslash & \times & & \diagup \\[2pt]
\sim A + \sim B & & A \circ B
\end{array}
$$
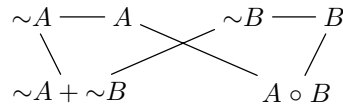
where this graph has the ports $A \circ B$ and $\sim A + \sim B$. Continuing with our Gentzen proof we have an axiom $\vdash C, \sim C$ (represented by the graph $C\!-\!\sim\! C$ with two ports) and we combine them with the application of a fusion rule, joining up the $A \circ B$ from the first graph and the C from the second, to form a new port $(A \circ B) \circ C$. The result is the graph

$$
\begin{array}{cccccc}
\sim A \!-\! A & & \sim B \!-\! B & & C \!-\! \sim C \\[2pt]
\big\backslash & \times & & \diagup & & \diagup \\[2pt]
\sim A + \sim B & & A \circ B & & \big\diagup \\[2pt]
& & & \diagdown & \diagup \\[2pt]
& & & (A \circ B) \circ C
\end{array}
$$
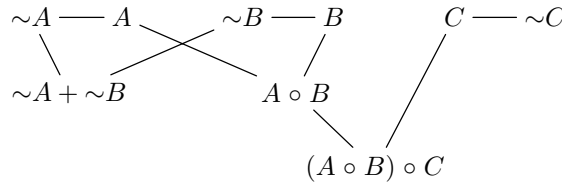
Notice that this graph does not bear the marks of the application of the last fusion rule occuring *after* the application of the fission rule. It could just as well have been constructed using the recipe of the second Gentzen proof. This proof takes the sequent $\vdash A \circ B, \sim A, \sim B$ and introduces the fission $\sim A, \sim B$ next. This intermediate step is the graph:

$$
\begin{array}{cccccc}
\sim A \!-\! A & & \sim B \!-\! B & & C \!-\! \sim C \\[2pt]
& \diagdown & & \diagup & & \diagup \\[2pt]
& & A \circ B & & \big\diagup \\[2pt]
& & & \diagdown & \diagup \\[2pt]
& & & (A \circ B) \circ C
\end{array}
$$
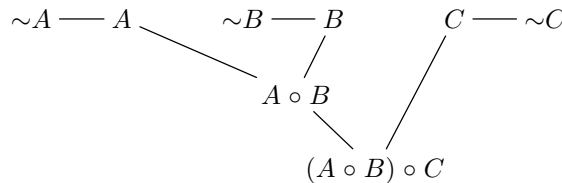
with ports $\sim A, \sim B, (A \circ B) \circ C, \sim C$, which is then completed with the final fission step, to construct

$$
\begin{array}{cccccc}
\sim A \!-\! A & & \sim B \!-\! B & & C \!-\! \sim C \\[2pt]
\big\backslash & \times & & \diagup & & \diagup \\[2pt]
\sim A + \sim B & & A \circ B & & \big\diagup \\[2pt]
& & & \diagdown & \diagup \\[2pt]
& & & (A \circ B) \circ C
\end{array}
$$
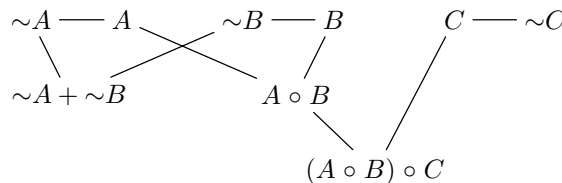
which is exactly the graph constructed by way of the other Gentzen proof. These graphs, or *proof nets* are "parallel" representation of proofs.

We have seen an example of the following definition of proof nets, the *inductive* definition.

- [AX] A graph A—$\sim A$ is an inductive proof net with ports A and $\sim A$.

- [+] If π is an inductive proof net with ports that include A and B, then the graph constructed by adding a node $A + B$, together with links between A and $A + B$ and B and $A + B$ is an inductive proof net whose ports are $A + B$ and the ports of π other than the indicated A and B.

- [○] If π and σ are inductive proof nets, which include ports A and B respectively, then the graph constructed by adding a node $A \circ B$, connecting the indicated A to $A \circ B$ by one arc, and B to $A \circ B$ by another, is an inductive proof net with the ports $A \circ B$ together with those of π and those of σ except for the indicated A and B.

- [CUT] If π is an inductive proof net with a port A and σ is an inductive proof net with a port $\sim A$, then the graph found by adding a link from the indicated A to the indicated $\sim A$ is an inductive proof net whose ports are those of π, except for the indicated A and those of σ, except for the indicated $\sim A$.

Clearly, for each Gentzen proof (in the vocabulary $\circ, +$) of a sequent, there is a corresponding inductive proof net whose ports are exactly the formulas occuring in that sequent.

Proof nets are graphs with distinctive restrictions on links: cut links and axiom links always connect a formula with its negation. Fission links (and fusion links) always come in pairs: an A to an $A + B$ and a B to an $A \circ B$. However, not every graph whose links are structured in this way is an inductive proof net. For example, the graph

$$A \circ B$$

$$A \qquad\qquad B$$

$$\sim A \qquad\qquad \sim B$$

$$\sim A \circ \sim B$$
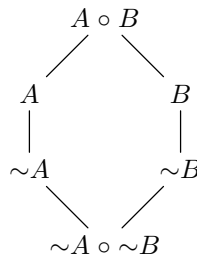
with ports $A \circ B$ and $\sim A \circ \sim B$ is not an inductive proof net, while the graph with the $\sim A \circ \sim B$ replaced by the fission $\sim A + \sim B$ *is* an inductive proof net.

The central theorem in the characterisation of proof nets gives an account of which graphs (of the type loosely characterised above) are genuine proof nets. An elegant criterion, provided by Danos and Regnier [1989] is the *switching* criterion. It can be simply explained. Notice that any proof net generated using the inductive definition — *without using the [+] condition* — is a tree. That is, proof nets generated using [AX], [CUT] and [○] are connected, but they contain no loops. (This shows that the graph with ports $A \circ B$ and $\sim A \circ \sim B$ with a loop is not an inductive proof net.) The only way loops may be

introduced in an inductive proof net is by way of the [+] rule. Given a pair A—$A + B$—B of links in a graph, we will call two different configurations

$$A\text{—}A + B \quad B$$

$$A \quad A + B\text{—}B$$

the two different *switchings* of this pair of fission links. Given a graph with pairs of fission links, a *switching* of a graph is found by replacing each pair of fission links by one of its switchings. Danos and Regnier's theorem is that a graph (of this structure) is an inductive proof net if and only if each switching of that graph is a tree. (As an example, you can see that our example proof net
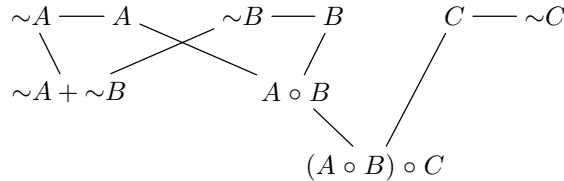
is not a tree, but its two switchings (selecting one link from $\sim\!A$—$\sim\!A + \sim\!B$—$\sim\!B$) are both trees.) The proof of the general fact is simple in one direction and difficult in the other. It is straightforward that every switching of an inductive proof net is a tree. This is a straightforward proof by induction on its construction. (The interesting case is [+]. If every switching of π (with ports A and B) is a tree, then so is every switching of the graph found by linking A and B in π to $A + B$. Each switching of this new graph is a switching of π together with a single link from either A or B to the new node $A + B$. This is also a tree.)

For the converse, we must show that every graph (of the right kind) for which every switching is a tree is an inductive proof net. This theorem has a number of different proofs (see Danos and Reginer [1989] for details), each of which use the switching criterion to show how a proof net may be "unwound" into a Gentzen proof.

The literature on proof nets for linear logic and related systems is growing. See the references for details [Bellin, 1991; Blute *et al.*, 1996; Cockett and Seely, 1997; Galmiche, 2000; Girard, 1987a; 1995].

2.10   *Curry-Howard*

Some logicians have found that it is possible to analyse proofs more closely by giving them *names*. After all, if proofs are first-class entities, we will be better-off if we can distinguish different proofs. I can illustrate this by looking at an example from intuitionistic logic. The language for describing proofs in the intuitionistic logic of the conditional and conjunction is given by the λ-calculus with pairing. A *term* of this calculus is built up from variables $x, y, \ldots$ using the *constructors* $\langle -, - \rangle$, fst $(-)$, snd $(-)$, $\lambda x.M$ and application (which we write as juxtaposition). A *judgement* is a pair $M{:}A$ of a term M and a formula A. Then in proofs in this system we keep tabs on what is going on by building

terms up to represent the ongoing proof. We start with the identity rule $x{:}A \vdash x{:}A$. Then for conjunction, we reason as follows:

$$\frac{\Gamma \vdash M{:}A \quad \Gamma \vdash N{:}B}{\Gamma \vdash \langle M, N \rangle{:}A \wedge B} \qquad \frac{\Gamma \vdash M{:}A \wedge B}{\Gamma \vdash \mathsf{fst}\,(M){:}A} \qquad \frac{\Gamma \vdash M{:}A \wedge B}{\Gamma \vdash \mathsf{snd}\,(M){:}B.}$$

If M is the proof of A from Γ, and N is the proof of B from Γ, then the pair $\langle M, N \rangle$ is the proof of $A \wedge B$ from Γ. Similarly, if M is a proof of $A \wedge B$ (from Γ) then $\mathsf{fst}\,(M)$ (the "first part" of M) is the proof of A from Γ. Similarly, $\mathsf{snd}\,(M)$ is the proof of B from Γ. For implication, we have these rules:

$$\frac{\Gamma \vdash M{:}A \rightarrow B \quad \Delta \vdash N{:}A}{\Gamma, \Delta \vdash (MN){:}B} \qquad \frac{\Gamma, x{:}A \vdash M{:}B}{\Gamma \vdash \lambda x.M{:}A \rightarrow B.}$$

If M is a proof of $A \rightarrow B$, and N is a proof of A, then you get a proof of B by *applying M to N*. So, this proof is (MN). Similarly, if M is a proof of B from $\Gamma, x{:}A$, then a proof of $A \rightarrow B$ is a function from proofs of A to the proof of B. It is of type $\lambda x.M$. We put these together to get names for more complex proofs

$$\frac{\dfrac{x{:}A \rightarrow B \vdash x{:}A \rightarrow B \quad y{:}A \vdash y{:}A}{\dfrac{x{:}A \rightarrow B, y{:}A \vdash (xy){:}B}{\dfrac{y{:}A \vdash \lambda x.(xy){:}(A \rightarrow B) \rightarrow B}{0 \vdash \lambda y.\lambda x.(xy){:}A \rightarrow ((A \rightarrow B) \rightarrow B).}}}}{}$$

The term $\lambda y.\lambda x.(xy)$ encodes the shape of the proof. The first step was an application of one assumption on another (the term (xy)). The second was the abstraction of the first assumption (λx), and the last step was the abstraction of the second assumption (λy). The term encodes the proof. There are a number of important features of these terms.

○ Terms encoding proofs with no premises are *closed*. They have no free variables.

○ More generally, if $\Gamma \vdash M{:}A$ is provable and x is free in M then x appears free in Γ too.

○ Proofs encode connective steps, not structural rules. For example, the rule CI or C was used in the proof of $A \rightarrow ((A \rightarrow B) \rightarrow B)$. It is not encoded in the term explicitly. Its presence can be seen implicitly by noting that the variables x and y are bound in the opposite order to their appearance.

Once we have a term system, we have *contracting rules*, which give us the behaviour of *proof reduction*.

$$\begin{aligned} \mathsf{fst}\,\langle M, N \rangle &\rightsquigarrow M \\ \mathsf{snd}\,\langle M, N \rangle &\rightsquigarrow N \\ (\lambda x.M)N &\rightsquigarrow M[x := N] \end{aligned}$$

These correspond to cutting the detours out of proofs. For example, consider the reduction

$$\frac{\dfrac{\Gamma \vdash M{:}A \quad \Gamma \vdash N{:}B}{\Gamma \vdash \langle M, N \rangle{:}A \wedge B}}{\Gamma \vdash \mathsf{fst}\,\langle M, N \rangle{:}A} \quad \rightsquigarrow \quad \Gamma \vdash M{:}A.$$

Or a slightly more complex case:

$$\frac{\dfrac{\Gamma, x{:}A \vdash M{:}B}{\Gamma \vdash \lambda x.M{:}A \to B \quad \Delta \vdash N{:}A}}{\Gamma, \Delta \vdash (\lambda x.M)N{:}B} \quad \rightsquigarrow \quad \Gamma, \Delta \vdash M[x := N]{:}B.$$

The term $M[x := N]$ indicates that the assumption(s) marked x in M are replaced by N. This matches the assumption(s) A marked x in $\Gamma, x{:}A$ which are replaced by the Δ in the transformation.

An explanation of the Curry–Howard isomorphism between intuitionistic logic and the types of terms in the λ-calculus is found in Howard's original paper [1980]. As we've already heard, Church's original calculus, the λI-calculus, was actually a model for the implicational fragment of R and not intuitionistic logic, as Church's calculus did not allow the binding of variables which were not free in the term in question [1941]. You eliminate contraction if you do not allow a λ term to bind more than one instance of a variable at once. Similarly, you eliminate C if you allow variables to be bound only in the order in which they are introduced. Structural rules correspond to restrictions on binding. A helpful account of more recent general work in types and logic is found in Girard, Lafont and Taylor's *Proofs and Types* [1989], and Girard's monograph *Proof Theory and Logical Complexity* [1987b].

Work on the application of the term calculus to substructural logics, focuses on three aspects. First, on encoding the normalisation results (that cutting detours out of proofs ends, and ends in a canonical "normal" proof). Second, on the appropriate term encoding of the exponentials of linear logic. Work in this area has not yet reached stability. The work of Benton, Bierman, Hyland and de Paiva [Benton *et al.*, 1992; 1992; 1993] shows the difficulty present in the area. Third, on showing that the restrictions on λ-abstraction in substructural logics has useful parallels in computation where resources may be consumed by computation. Wadler and colleagues show that this kind of term system has connections with functional programming [Maraist *et al.*, 1995] and [1990; 1991; 1992b; 1992a; 1993a; 1993b].

2.11 Structurally Free Logic

A very recent innovation in the proof theory of substructural logics is the advent of *structurally free logic*. The idea is not new — it comes from a 1976 essay by Bob Meyer [1976b]. However, the detailed exposition is new, dating from 1997 [Bimbó and Dunn, 1998; Bimbó, 2001; Dunn and Meyer, 1997]. The motivating idea is simple. Just as *free logic* is free from existential commitments and any existence claims can be explicitly examined and questioned, so in a *structurally free logic*, *no* structural rules are present in and of themselves, but structural rules, if applied, are *marked* in a proof as explicit premises. So, structural rules are tagged with a *combinator*, such as these examples:

$$\frac{W(X, (Y, Z)) \vdash A}{W(((\mathsf{B}, X), Y), Z) \vdash A} \; [\mathsf{B}] \qquad \frac{W((X, Z), Y)) \vdash A}{W(((\mathsf{C}, X), Y), Z) \vdash A} \; [\mathsf{C}] \qquad \frac{W((X, Y), Y)) \vdash A}{W((\mathsf{W}, X), Y) \vdash A.} \; [\mathsf{W}]$$

These are the combinator versions of the structural rules B (association) C (commutativity) and W (contraction). Now the conclusions not only feature the structures as rearranged: they also feature a combinator marking the action of the structural rule. Proofs in this kind of system then come with "tickets" indicating which kinds of structural rules licence the conclusion:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{B \vdash B \quad A \vdash A}{(A \to B), A \vdash B}\,(\to L) \quad C \vdash C}{(A \to B), (C \to A, C) \vdash B}\,(\to L)}{((\mathsf{B}, A \to B), C \to A), C \vdash B}\,(\mathsf{B})}{(\mathsf{B}, A \to B), C \to A \vdash C \to B}\,(\to R)}{\mathsf{B}, A \to B \vdash (C \to A) \to (C \to B)}\,(\to R).}{\mathsf{B} \vdash (A \to B) \to ((C \to A) \to (C \to B))}\,(\to R)$$

Any further explanation the workings of this proof system for structurally free logic brings us perilously close to looking at models for combinatory logic and the λ-calculus [Barendregt *et al.*, 1983; Meyer, 1991]. I will defer this discussion to the next section, where we broach the question in a broader setting. What on earth counts as a *model* of a substructural logic?

3   MODELS

Our focus so far has been *syntax* and *proof*. Now we turn our gaze to *interpretation*. Clearly we have not been unconcerned with matters of interpretation thus far. We have paid some attention to the meanings of the connectives when we have examined the kinds of inferential steps appropriate for sentences formed out of these connectives. According to some views, in giving these rules for a connective we have thereby explicated their meanings. According to other views, we have merely cashed out a *consequence* of the meanings of the connectives, meanings which are to be found in some other way.[69] Thankfully, we have no need to adjudicate such a debate here. It is not our place to clarify the ultimate source of meaning. It is, however, our place to consider some of the different kinds of interpretations open to logical systems, and particular, substructural logics.

An interpretation of a language is a map from the sentences of the language into some kind of structure. There are many possible kinds of interpretations. Some propositions are true and others are not true. We can interpret a language in the structure $\{\mathsf{t}, \mathsf{f}\}$ of *truth values* by setting the interpretation $[\![A]\!]$ of A to be t if A is true, and f otherwise.[70] This interpretation is helpful in the study of logical consequence because of the way it interacts

---

[69]This debate is between *truth conditional* [Tarski, 1956] versus *inferentialist* [Brandom, 1994] accounts of meaning in philosophy of language, *proof theorists* [Girard, 1987b; Dragalin, 1987] and *model theorists* [Bell and Slomson, 1969; Hodges, 1993] in mathematical logic, and *operational* and *denotational* semantics in computer science [Mitchell, 1996].

[70]Note: $\{\mathsf{t}$ and $\mathsf{f}\}$ are the two truth values *true* and *false*: not necessarily the Ackermann constants t and f.

with the traditional propositional connectives. A conjunction is true if and only if both of the conjuncts are true. A disjunction is true if and only if one of the disjuncts is true. A negation is true if and only if the negand is not true. It follows that $[\![A \wedge B]\!]$, $[\![A \vee B]\!]$ are functions of $[\![A]\!]$ and $[\![B]\!]$, in the sense that once the values $[\![A]\!]$ and $[\![B]\!]$ are fixed, the values $[\![A \wedge B]\!]$, $[\![A \vee B]\!]$ are also fixed. The behaviour of the operations of conjunction, disjunction and negation on the set $\{t, f\}$ of truth values goes some way towards telling us the meanings of those connectives. More than that, it gives us an account of the behaviour of logical consequence, as the set of truth values has a natural *order*. We can order the set by saying that $f < t$, in the sense that t is "more true" than f. An argument is $\{t, f\}$-valid if no matter how you interpret the propositions in the argument, the conclusion is never any less true than the premises. Or in this case, you never can interpret the premises as true and the conclusion as false. This is the traditional truth-table conception of validity.

The simple set $\{t, f\}$ of truth values is not the only domain in which a language can be interpreted. For example, we might think that not all propositions or sentences in the language are truth-valued. We might interpret the language in the structure $\{t, n, f\}$, where the true claims are interpreted as t, the false ones as f, and the non-truth-valued sentences are interpreted as n. This path leads one to *many valued logics* [Dunn and Epstein, 1977; Urquhart, 1986].

However, one need not interpret the domain of values as *truth values*. For one early example of an alternative sort of domain in which sentences can be interpreted, consider Frege's later philosophy of language. For the Frege of the *Grundgesetze* [1993] declarative sentences had a *reference* (*Bedeutung*) and a *sense* (*Sinn*). We can interpret sentences by mapping them onto a domain of senses and by interpreting the connectives as functions on senses. This is another "denotational" semantics for declarative sentences.[71]

Different applications will motivate different sorts of models and domains of semantic values. In the Lambek calculus for syntactic types, the formulas can be mapped onto sets of syntactic strings. In this interpretation, a sentence will be modelled by the set of strings (in the analysed language) which have the type denoted by the sentence.

These last two examples — of possible worlds and of syntactic strings — have similar structures. Formulas are interpreted as *sets* of objects of one kind or other. These are especially interesting models, which we will discuss in detail soon. For now, however, I will focus on the general idea of interpreting logics in structures, for simple *algebras* are the first port of call when it comes to models of substructural logics.

3.1 Algebras

The most direct way to interpret a logic is by a map from the language of the logic into some structure. Such structures are usually equipped with operations to match the connec-

---

[71]For a modern interpretation of Frege's ideas, one could consider a sense of a claim to be the set of possible worlds in which it is true. Now for each sentence you have its interpretation as some set of possible worlds. For an account of how this approach might be philosophically productive, see Robert Stalnaker's *Inquiry* [1984], David Lewis' *On the Plurality of Worlds* [1986].

For recent work which takes Frege's talk of senses at face value (and which motivates a weak substructural logic, to boot) consider the paper "Sense, Entailment and *Modus Ponens*" by Graham Priest [1980].

tives in the language. The interpretation of a complex formula is then defined recursively in terms of the operations on the interpretations of the atomic subformulas. All of this is standard. In this section, I will examine a few structures which have proved to be useful in the study of substructural logics. Then in the next section, I will explain just a few of the theorems which can be proved about substructural logics by using these structures.

Example Algebras

EXAMPLE 17 (BN4). Perhaps the most simple, yet rich, finite structure used to interpret substructural logics is the four-valued lattice BN4 [Dunn, 1976a; Belnap, 1977a; 1977b]. It first came to fame as a simple lattice sufficient to interpret first degree entailments. Any valid first degree entailment is valid in this structure (in a sense to be explained soon) and any invalid first degree entailment is invalid in this structure. It is also the source of intuitions in its own right. The behaviour of BN4 is presented in the diagram and tables in Figure 5. The diagram can present the behaviour of conjunction, disjunction, $\top$ and $\bot$. The conjunction of two elements is their greatest lower bound, their disjunction, the least upper bound, $\top$ is the top element and $\bot$ is the bottom element. The operations of negation and implication and fusion are read off the tables.

	$\sim$		$\rightarrow$	t	b	n	f		$\circ$	t	b	n	f
t	f		t	t	f	n	f		t	t	t	n	f
b	b		b	t	b	n	f		b	t	b	n	f
n	n		n	t	n	t	n		n	n	n	f	f
f	t		f	t	t	t	t		f	f	f	f	f
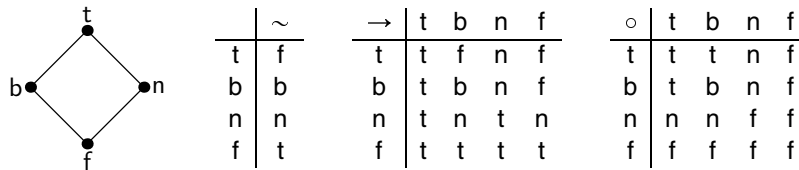
Figure 5. The Algebra BN4

If you think of the values t, b, n, f as the values "true only", "both true and false", "neither true nor false" and "false only" then the negation of a set values is simply the set of the negations of values in that set. Implication is similarly defined. The value "true" is in the set $a \rightarrow b$ just when if a is at least "true" then b is at least "true", and if b is at least "false" then so is a. On the other hand, a conditional $a \rightarrow b$ is at least "false" if a is at least "true" and b is at least "false." This gives the implication table. The values in the fusion table are given by setting $a \circ b$ to be $\sim(a \rightarrow \sim b)$.

Given this definition, fusion is commutative and associative, with an identity b. Negation is definable in terms of implication by setting $\sim a$ to be $a \rightarrow$ b. So in this algebra, the false constant f is modelled by b, as is the true constant t. Fusion is residuated by $\rightarrow$, and the lattice is distributive.

In this algebra, the order in the diagram (read from bottom to top, and written "$\leqslant$") models *entailment*. You can see that $a \wedge b$ always entails a, as the greatest lower bound of a and b (whatever a and b might be) is always lower than, or equal to, a. In just the same way, you can show that all of the entailments of a distributive lattice hold for $\wedge$ and $\vee$, that $a = \sim\sim a$ (and so, double negation elimination and introduction hold) and that the De Morgan laws, such as $\sim(a \vee b) = \sim a \wedge \sim b$ also hold in this structure.

∘	⊥	~c	~b	~a	a	b	c	⊤
⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
~c	⊥	~c	⊤	⊤	~c	⊤	~c	⊤
~b	⊥	⊤	~b	⊤	~b	~b	⊤	⊤
~a	⊥	⊤	⊤	⊤	~a	⊤	⊤	⊤
a	⊥	~c	~b	~a	a	b	c	⊤
b	⊥	⊤	~b	⊤	b	b	⊤	⊤
c	⊥	~c	⊤	⊤	c	⊤	c	⊤
⊤	⊥	⊤	⊤	⊤	⊤	⊤	⊤	⊤
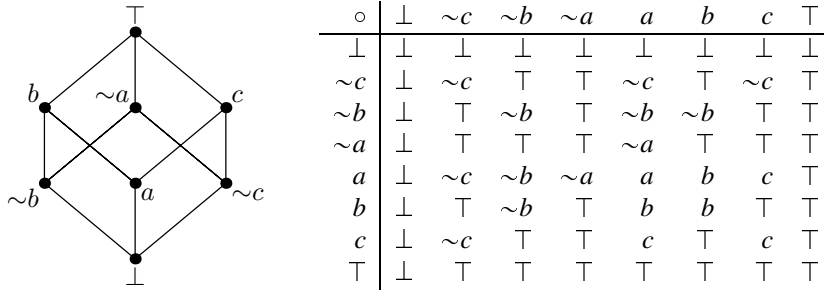
Figure 6. An Eight-Point Lattice for R

In this lattice, some structural rules fail: WI is not satisfied, as n $\not\leqslant$ n ∘ n. The K rule also fails, as t ∘ b $\leqslant$ b, and hence we do not have b $\leqslant$ t → b. So BN4 is a model for linear logic (with the addition of distribution), in the sense that if $A \vdash B$ holds in linear logic plus distribution, then for any interpretation $[\![\cdot]\!]$ into BN4, we must have $[\![A]\!] \leqslant [\![B]\!]$.

However, BN4 is not a model of R, for some contraction related principles fail in BN4. For example, there is an interpretation in which $[\![A \wedge (A \rightarrow B)]\!] \not\leqslant [\![B]\!]$.[72]

This lattice has also been used in the semantics of programming [Fitting, 1989]. Interpreting the four values as epistemic states of no information, positive and negative information, and conflicting information, may be of some help in modelling states of information-bearing devices.

EXAMPLE 18 (An Eight Point Model). Consider the structure with the order and fusion table shown in Figure 6. This is a model of R: Fusion is commutative (the table is symmetric about the diagonal), and associative. We have $x \leqslant x \circ x$, so WI holds. The element a is an identity for fusion. Negation is defined by the names of the elements and the fact that $\sim$ is a De Morgan negation. Setting $x \rightarrow y = \sim(x \circ \sim y)$ makes $\rightarrow$ residuate fusion.

We can use this structure to show that R has the relevance property. Suppose we have two propositions A and B, in the language $\wedge$, $\vee$, $\sim$, $\circ$ and $\rightarrow$, such that there is no atom shared between A and B. Construct an evaluation $[\![\circ]\!]$, such that $[\![p]\!]$ is either b or $\sim b$ for any atom p in A, and it is either c or $\sim c$ for any atom p in B.

By induction, we can verify that the value $[\![A]\!]$ is one of b and $\sim b$, and similarly, the value $[\![B]\!]$ is one of c or $\sim c$. Therefore, $[\![A]\!] \not\leqslant [\![B]\!]$, and since this is a model of R, we have $A \nvdash B$ in R, and hence, $A \nvdash B$ in any sublogic of R.

EXAMPLE 19 (Sugihara Models). One can modify BN4 in a number of ways. You can leave out the value b, and get Łukasiewicz's three-valued logic. Extensions to Łukasiewicz's n-valued, and infinitary logics are straightforward too. These systems all invalidate contraction, but validate weakening and the other common contraction-free structural

---

[72]Hint: set $[\![A]\!]$ = n and $[\![B]\!]$ = f. Check for yourself that this is a counterexample.

rules. Another way of modifying BN4 is to leave out the value n. This gives us the structure known as RM3, a three-valued algebra useful in the study of relevant logics, because this is a model of R. This simple three-valued model can be generalised to RM_{2n+1} for any n as follows by setting the domain of propositions to be the numbers

$$\{-n, -(n-1), \ldots, -1, 0, 1, \ldots, n-1, n\}$$

where we set $\sim a$ to be $-a$, and $\rightarrow$ and fusion are defined as follows:

$$a \rightarrow b = \left\{ \begin{array}{ll} -a \vee b & \text{if } a \leqslant b \\ -a \wedge b & \text{if } a > b \end{array} \right. \qquad a \cdot b = \left\{ \begin{array}{ll} a \wedge b & \text{if } a \leqslant -b \\ a \vee b & \text{if } a > -b. \end{array} \right.$$

Fusion is commutative (verify by eye) and associative (verify by checking case by case), with identity 0. Note that $a \cdot a = a$, so the logic satisfies both W and M — this is a model for the logic RM discussed earlier.

This model can also be extended by not stopping at $-n$ or n but by including all of $\mathbb{Z}$, the positive and negative integers. This infinite model captures exactly the logic RM in the language $\wedge, \vee, \rightarrow, \circ, \sim, t$. The infinite model has no members fit for either $\top$ or $\bot$, but they can be added as ∞ and $-\infty$ without disturbing the logic of the model.

EXAMPLE 20 (The Integers). The integers feature in the RM algebra above. The choice of the interpretation of implication in that model is only one of many different ways you could go in this structure. Another is to consider *addition* as a model for fusion. The residual for addition is obvious: it is subtraction. $X \rightarrow Y$ is $Y - X$. This structure is unlike the RM algebra in a number of ways. First, W fails, as $a \not\leqslant a + a$ whenever a is negative. Second, M fails (and so, K and K′ do too) as $a + a \not\leqslant a$ whenever a is positive. However, C and B are satisfied in this structure, so we have a structure fit for linear logic. In particular, since the structure is totally ordered, we have the distribution of conjunction over disjunction, so we have a model for *distributive* linear logic.

More interestingly, $(x \rightarrow y) \rightarrow y = x$ for each x and y. This does not hold in any boolean algebra or in any other non-trivial structure with $\top$. If $\top$ were present, then $(a \rightarrow \top) \rightarrow \top = a$ but $\top \leqslant b \rightarrow \top$ for each b (including $a \rightarrow \top$) so $\top \leqslant a$.

It is possible to define the negation $\sim a$ as $a \rightarrow 0 = -a$. However, other choices are possible. Taking b an arbitrary proposition, we can define $\sim_b a$ as $a \rightarrow b$, and the condition $(x \rightarrow y) \rightarrow y = x$ states, in effect, that $\sim_b \sim_b a = a$. Double negation introduction and elimination holds for *any* negation $\sim_b$ we choose.

This model is a way to invalidate simple consecutions in distributive linear logic. For example, can we prove $A \rightarrow (B \rightarrow C) \vdash (A \rightarrow B) \rightarrow (A \rightarrow C)$? If this holds in our structure we must have $(z - y) - x \leqslant (z - x) - (y - x)$, but this simplifies to $z \leqslant (z - x) + 2x = z + x$ (add $x + y$ to both sides). And when $x \geqslant 0$ we have $z \leqslant z + x$, but if $x < 0$ this fails. Similar manipulations can be used to invalidate other consecutions. However, some consecutions invalid in distributive linear logic do hold in the integers. We have seen that $(A \rightarrow B) \rightarrow B \vdash A$ already. Another case is $t \vdash (A \rightarrow B) \vee (B \rightarrow A)$. So the integers do not give an exact fit for distributive linear logic.

(Others have been aware that simple "counting" mechanisms can provide a useful filter for issues of validity in substructural logics [van Benthem, 1991; Kurtonina, 1995; Roorda, 1991; Pentus, 1995].)

The logic here is known as *abelian logic*: It was introduced by Meyer and Slaney, who show that it is the logic of ordered abelian groups [1989].

EXAMPLE 21 (ω Under Division). Using number systems as structures gives us rich mathematical tradition upon which we can build. However, the structures we have seen so far are all *totally ordered*: for any x and y either $x \leqslant y$ or $y \leqslant x$. This is not always desirable — it leads to the truth of $(A \rightarrow B) \vee (B \rightarrow A)$. Now some "natural" orderings of numbers are total orders, but others are not. For example, take the positive integers, ordered by *divisibility*. This is a partial ordering — indeed, a lattice ordering — in which join is the lowest common multiple (lcm) and meet is the greatest common divisor (gcd). Fusion has a natural model in multiplication.

The lattice is distributive, as $\gcd(a, \text{lcm}(b, c)) \mid \text{lcm}(\gcd(a, b), \gcd(b, c))$. With fusion modelled as multiplication, 1 is the identity of fusion and we have a distributive lattice-ordered commutative monoid with a unit. Furthermore, the monoid is square-increasing (as $a \mid a^2$), so it models the behaviour of the $\wedge, \vee, \circ, t$ part of the logic R.

How can you model a conditional residuating fusion? We want

$$xy \mid z \text{ if and only if } x \mid y \rightarrow z.$$

If y divides z, then we can set $y \rightarrow z$ to be z/y. For any x you choose, $xy \mid z$ if and only if $x \mid z/y$. However, if y does not divide z, we do not have anything to choose, as 1 is the bottom of the order we have thus far. To get a residual in every instance, we need to add another element to the ordering. It will be the lowest element in the ordering, so we will call it 0 (for reasons which will become more obvious later). We can by fiat determine that $0 \mid x$ for every x in the structure, and that $x \mid 0$ only when $x = 0$. Conjunction and disjunction are as before, with the addition that $0 \wedge x = 0$ and $0 \vee x = x$ for each x. The rule for implication is then:

$$x \rightarrow y = \begin{cases} y/x & \text{if } x \mid y, \\ 0 & \text{otherwise.} \end{cases}$$

Given 0 we need to extend the interpretation of fusion. But this is simple

$$0x = x0 = 0$$

for every x. So defined, the operation is still order-preserving, commutative, square-increasing and with 1 as the identity. This structure is a model for the *positive* part of R. To model the whole of R we need to model a De Morgan negation. That requires an order-inverting involution on the structure. To do this, we need to introduce many more elements in the structure, as no order-inverting involution can be found on what we have here before us: consider the infinite ascending chain

$$0 \mid 1 \mid 2 \mid 4 \mid \cdots \mid 2^n \mid \cdots .$$

To negate each element in the series you must get an infinite descending chain. Why? Because we need an involution: $x \mid y$ if and only if $-y \mid -x$, and in particular, if $-x = -y$, then we must have $x = y$. Each element in the inverted chain must be distinct. Alas, there

are no such chains in our structure, as every number has only finitely many divisors. So, we need to add more elements to do the job. As our notation has suggested, we will add the negative integers and ∞. The order is given by setting

$$0 \mid x \mid -y \mid \infty$$

for every positive x and y, and in particular, $-x \mid -y$ if and only if $y \mid x$. So you can read '$\mid$' as divides only when it holds between positive integers. Otherwise, it is defined by these clauses. The infinite ascending chain is then mapped onto the infinite descending chain *above it* as follows:

$$0 \mid 1 \mid 2 \mid 4 \mid \cdots \mid 2^n \mid \cdots \mid -2^n \mid \cdots \mid -4 \mid -2 \mid -1 \mid \infty.$$

The result is still a distributive lattice order, and conjunction and disjunction are obviously definable as greatest lower bound and least upper bound, respectively. Implication between all pairs of elements is defined as follows:

○ If x is negative and y is positive, $x \to y = 0$.

○ If x is positive and y is negative, then $x \to y = (-x)y$

○ If x and y are both negative, then $x \to y = -y \to -x$

Fusion is then defined by setting $xy = -(x \to -y)$, and you can show that this is commutative, square-increasing and with 1 as the identity.

The lattice is not *complete*, in that not every subset has a least upper or a greatest lower bound: The chain $0 \mid 1 \mid 2 \mid \cdots \mid 2^n \mid \cdots$ has an upper bound (any negative number will do) but no least upper bound.

This structure was first constructed by Meyer [1970b] in 1970 who used it to establish some formal properties of R. The technique of expanding an algebra to model negation is one we shall see again as an important technique in the metatheory of these logics.

EXAMPLE 22 (Algebras of Relations). A generalisation of Boolean algebras due to De Morgan [1964] and Peirce [1970] and later developed, for example, by Schröder [1995], was to consider algebras of *binary relations*. A *concrete relation algebra* is the set of all subsets of some set $D \times D$ of pairs of elements from a set D under not only the Boolean operations of intersection, union and complementation but also under new operations which exploit the relational structure.

For any two relations R and S their composition is also a relation: $R \cdot S$ is defined by setting $x(R \cdot S)y$ if and only if $(\exists z \in D)(xRz \wedge zRy)$. This is a model for fusion. Fusion has a left and right identity, 1, the identity relation on D. Furthermore, for any relation R we have its converse, given by setting $x\check{R}y$ if and only if yRx. Note that $\widetilde{(R \cdot S)} = \check{S} \cdot \check{R}$.

We can define left and right residuals for composition directly by the residuation conditions, or we can note that they are definable in terms of the Boolean connectives, fusion and converse. $R \to S = -(-S \cdot \check{R})$ and $S \leftarrow R = -(\check{R} \cdot -S)$.

It is possible to modify the behaviour of these algebras by considering restricted classes of relations. For example, we could look at algebras of reflexive relations. These are odd,

in that $1 \leqslant R$ for each R, so the bottom element of the algebra is also the identity for fusion. These algebras are closed under some of the operations at issue, but not all. The Boolean complement of a reflexive relation is not reflexive, but the conjunction or disjunction of two reflexive relations is.

Another possibility is to consider, for example, equivalence relations [Finberg *et al.*, 1996]. When is the composition of two equivalence relations an equivalence relation? It turns out that $R \cdot S$ is also commutative when $R \cdot S = S \cdot R$. And if this obtains, then their composition is the least upper bound of the two relations (in the set of equivalence relations). Therefore, a class of commuting equivalence relations forms a lattice, in which fusion is least upper bound. And it is not too hard to show that this lattice generally fails to be distributive, but it is *modular*. It satisfies the modular law

$$a \wedge (b \vee (a \wedge c)) \leqslant (a \wedge b) \vee (a \wedge c)$$

but not the more general distributive law.

Tarski [1941] helped bring relation algebras back to prominence in modern logic, and there is much contemporary research in the area, particularly in Hungary [Andéka *et al.*, 1988]. Vaughan Pratt has also considered them (and *dynamic algebras*, a tractable fragment of relation algebras) as a useful model of computation [1990].

General Structures

The algebraic study of models of substructural logics was first explicitly and comprehensively tackled by J. Michael Dunn in his doctoral dissertation from the middle 1960s [Dunn, 1966]. The techniques he used are mostly standard ones, adapted to the new context of relevant logics. There had been a long tradition of using finite algebras (also called 'matrices' for obvious reasons) to prove syntactic results about logics, such as the relevance property for R, as we have seen. Section 22 of *Entailment* Volume 1 [Anderson and Belnap, 1975] contains a good discussion of results of this sort. However, it was Dunn's work that first took such structures as a fit object of study in their own right.

For a helpful guide to the state of the art in the 1970s, Helena Rasiowa's *An Algebraic Approach to Non-classical Logics* [1974] is a compendium of results in the field. Meyer and Routley's groundbreaking paper "An Algebraic Analysis of Entailment" [1972] did a great deal of work showing how a whole host of logics fit together, all with the theme of *residuation* or the connection of fusion with implication. They showed that not only in R but also in other relevant logics, fusion is connected together with implication by the residuation postulate

$$a \circ b \leqslant c \text{ iff } a \leqslant b \to c$$

and that the natural way to ring the changes in the logic is to vary the postulates governing fusion. Here is a summary of Dunn's and Meyer and Routley's work on the general theory of algebras for substructural logics.

DEFINITION 23 (Posets). A *poset* (a partially ordered set) is a set equipped with a binary relation $\leqslant$ which is reflexive, transitive, and asymmetric. That is, $a \leqslant a$ for each a, if $a \leqslant b$ and $b \leqslant c$ then $a \leqslant c$, and if both $a \leqslant b$ and $b \leqslant a$, then $a = b$.

Posets are the basic structure of an algebra for a logic. The order is *entailment* between propositions in structure. Entailment is asymmetric as we assume that co-entailing propositions are *identical*. This is what makes propositions in this kind of structure differ from sentences in a formal language.

Extensional conjunction and disjunction enrich the poset into a familiar algebraic structure:

DEFINITION 24 (Lattices). A lattice is a partially ordered set equipped with least upper bound $\vee$ and a greatest lower bound $\wedge$.

A lattice is *distributive* if and only if $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ holds for each a, b and c.

A lattice is *bounded* if it has greatest and least elements, $\top$ and $\bot$ respectively.

For traditional substructural logics, two more additions are required to this kind of structure. First, negation, and second, fusion and implication. Let's tackle negation first.

DEFINITION 25 (Negations). A *negation* on a poset is an order inverting operation $\sim$: that is, if $a \leqslant b$ then $\sim b \leqslant \sim a$.

A negation on a lattice is *De Morgan* if $\sim\sim a = a$ and it satisfies the De Morgan identities $\sim(a \wedge b) = \sim a \vee \sim b$ and $\sim(a \vee b) = \sim a \wedge \sim b$.

A De Morgan negation in a bounded lattice is an *ortho-negation* if $a \wedge \sim a = \bot$ and $a \vee \sim a = \top$.

Note that a De Morgan negation need not be an ortho-negation. (The negations in each of the structures in the previous section are De Morgan but not ortho-negation.) An ortho-negation operation in a distributive lattice is the *Boolean* negation in that structure.[73]

Some very recent work of Dunn has charted even more possibilities for the behaviour of negation. In particular, he has shown that a basic structure in a substructural logic is a *split negation* satisfying the following residuation-like clauses

$$a \leqslant \neg b \text{ iff } b \leqslant -a.$$

Given this equivalence, both $\neg$ and $-$ are negations, and both satisfy *some* of the De Morgan inequalities but not others [Dunn, 1994].[74]

The most interesting operations in algebras for substructural logics are fusion and implication. The simplest way to define them is by residuation.

DEFINITION 26 (Residuated Pairs and Triples). $\langle \circ, \rightarrow \rangle$ is a *residuated pair* in a poset if and only if $a \circ b \leqslant c$ if and only if $a \leqslant b \rightarrow c$.

---

[73]There may be more than one ortho-negation in a lattice, but there is only one ortho-negation in a distributive lattice.

[74]In particular, $\neg a \wedge \neg b = \neg(a \vee b)$ and $\neg a \vee \neg b \leqslant \neg(a \wedge b)$ (this latter inequality is satisfied by *any* negation) but the converse can fail: $\neg(a \wedge b) \not\leqslant \neg a \vee \neg b$. The negation $\neg$ differs from intuitionistic or minimal negation, however, by not necessarily satisfying $a \leqslant \neg\neg a$. We do have, however, $a \leqslant \neg - a$ and $a \leqslant -\neg a$. For an example of a split negation, let $-A$ be $A \rightarrow f$ and let $\sim A$ be $f \leftarrow A$ in the Lambek calculus extended with a false constant f.

$\langle \circ, \rightarrow, \leftarrow \rangle$ is a *residuated triple* in a poset if and only if $a \circ b \leqslant c$ if and only if $a \leqslant b \rightarrow c$ if and only if $b \leqslant c \leftarrow a$.

If $\langle \circ, \rightarrow \rangle$ is a residuated pair then it immediately follows that $\circ$ is *isotonic* in both places with respect to the entailment ordering. That is

$$\text{if } a \leqslant a' \text{ and } b \leqslant b' \text{ then } a \circ b \leqslant a' \circ b'.$$

Implication, on the other hand, is not isotonic in both places. It is isotonic in the *consequent* place and *antitonic* in the antecedent place. That is we have

$$\text{if } a' \leqslant a \text{ and } b \leqslant b' \text{ then } a \rightarrow b \leqslant a' \rightarrow b'.$$

All of this was noticed by Meyer and Routley in the 1970s and made rigorous (and generalised to arbitrary n-place operations and residuated families) by Dunn in the 1980s and 1990s in his work on *gaggle theory* (from "ggl" for "Generalised Galois Logic": a *Galois connection* is the general phenomenon of which a residuated pair or triple is a special case) [1991; 1993].[75]

Tonicity is not the only behaviour of fusion and implication present in these models. If the poset is a *lattice* ordering, then tonicity generalises to *distribution*. It is also an elementary consequence of the residuation clause that fusion distributes over disjunction in both places

$$(a \vee a') \circ b = (a \circ b) \vee (a' \circ b) \text{ and } a \circ (b \vee b') = (a \circ b) \vee (a \circ b')$$

and implication distributes over the extensional connectives in a more complicated fashion.

$$(a \vee a') \rightarrow b = (a \rightarrow b) \wedge (a' \rightarrow b) \text{ and } a \rightarrow (b \wedge b') = (a \rightarrow b) \wedge (a \rightarrow b').$$

These sorts of structures are well known, and they appear independently in different disciplines. *Quantales* [Mulvey, 1986] are but one example. These are lattice-ordered semi-groups (so, $\circ$ is *associative*) with arbitrary disjunctions but only finite conjunctions. They appear in both pure mathematics and theoretical computer science. They are discussed a little in Vickers' *Topology via Logic* [1990], which is a useful source book of other algebraic constructions and their use in modelling processes and observation. The existence of arbitrary disjunctions means that in a quantale, implication is *definable* from fusion. If you set $a \rightarrow b$ as follows

$$a \rightarrow b = \bigvee \{x : x \circ a \leqslant b\}$$

then $\rightarrow$ satisfies the residuating condition for fusion.[76] The same definition is possible in the other direction too. If you have a lattice with arbitrary *conjunctions* (and implication

---

[75]And I have begun to sketch the obvious parallels between gaggle theory and display logic. Residuation is displaying, and isotonicity and antitonicity have connections to antecedent and consequent positions in the proof rules for a connective. When you process a fusion, the subformulas remain on the same side of the turnstile as the original formula. On the other hand, when you process an implication, the antecedent swaps sides and the consequent stays put.

[76]The distribution of $\circ$ over the infinitary disjunction is essential here.

distributes over conjunction in the right way) then you can define fusion from implication

$$a \circ b = \bigwedge \{x : a \leqslant b \to x\}.$$

This definition is key to one of the important techniques in understanding the behaviour of fusion and the connections between fusion and implication. For fusion plays no part in the Hilbert systems introducing some substructural logics. Yet it is present in the Gentzen systems (at least in the guise of the comma, if not explicitly) and in these algebras. Does the addition of fusion add anything new to the system in the language of implication? Or is the addition of fusion *conservative*? In the next section I will sketch Meyer's techniques for proving conservative extensions for many substructural logics, by way of algebraic models.

Before that, I must say a little about *truth* in these algebras. In Boolean algebras (for classical logic) and Heyting lattices (for intuitionistic logic) the *truths* in a structure are the formulas which are interpreted as the top element. There is no need for this to be the case in our structures. In the absence of K', we might have $b \not\leqslant a \to a$. That means that a true conditional (as every identity $a \to a$ is true) need not be the top element of the ordering. So, instead of picking out true propositions as those at the top of the ordering, substructural logics need to be more subtle.

DEFINITION 27 (A Truth Set). Given an algebra with $\to$, the *truth set T* is the set of all x where $a \to b \leqslant x$ for some a, b where $a \leqslant b$.[77]

The truth set is the set of all conditionals true on the basis of logic alone, and anything entailed by those conditionals. A truth set has some nice properties.

FACT 28 (A Truth Set in a Lattice is a Filter). Any truth set T in a lattice is a *filter*. (A filter is a set which is closed under $\leqslant$, and closed under conjunction.[78]) If $x, y \in T$ then $x \wedge y \in T$. If $x \leqslant x'$ then $x' \in T$ too.

Proof. That T is closed upwards is immediate. That T is closed under conjunction, note that if $x, y \in T$ then $a \to b \leqslant x$ and $a' \to b' \leqslant y$ where $a \leqslant b$ and $a' \leqslant b'$. Then $a \wedge a' \leqslant b \vee b'$, and $(a \to b) \wedge (a' \to b') = a \wedge a' \to b \vee b'$, so since $(a \to b) \wedge (a' \to b') \leqslant x \wedge y$, we have $x \wedge y \in T$ too. ∎

If the logic contains t, then the truth set is the filter *generated* by t: $T = \{x : t \leqslant x\}$. The presence of truth sets in models shows that a logic without t can be conservatively extended by it. Both conservative extension constructions — due to Meyer in the 1970s — are the topic of the next section.

Conservative Extension Theorems

Meyer's conservative extension results follow the one technique [1973a]. Suppose we have consecution invalid in a logic with a restricted language $A \nvdash B$. Then (by the soundness and completeness results for propositional structures) there is an algebra $\mathcal{A}$ and an

---

[77]Equivalently, it is the set generated by all identities $a \to a$, since $a \to a \leqslant a \to b$ if $a \leqslant b$.

[78]It is the algebraic analogue of a *theory*, which we have already seen.

interpretation $[\![\cdot]\!]$ into $\mathcal{A}$ where $[\![A]\!] \not\leqslant [\![B]\!]$. Then, we *manipulate* $\mathcal{A}$ into a new structure $\mathcal{A}'$, appropriate for the larger language, and in which we have a new interpretation which is still a counterexample to $[\![A]\!] \not\leqslant [\![B]\!]$. There are two separate techniques Meyer pioneered. One, injecting a structure $\mathcal{A}$ into it's *completion* $\mathcal{A}^*$ (giving us a way to interpret t, $\circ$, and conjunction and disjunction if those are not present), and then, taking a structure $\mathcal{A}$ and pasting on an inverted duplicate $\mathcal{A}^{op}$, in order to model negation.

EXAMPLE 29 (Mapping $\mathcal{A}$ into $\mathcal{A}^*$). The map from a propositional structure into its *completion* is given in the following way. $\mathcal{A}^*$ is defined in a number of alternative ways.

○ If $\mathcal{A}$ is only a poset or a semilattice (with $\wedge$ but not $\vee$), then $\mathcal{A}^*$ is the set of all downwardly closed sets of $\mathcal{A}$. That is, I is an element of $\mathcal{A}^*$ if and only if whenever $a \in I$ and $b \leqslant a$ then $b \in I$.

○ If $\mathcal{A}$ contains disjunction, then $\mathcal{A}^*$ is the set of all *ideals* in $\mathcal{A}$. (Ideals are dual to filters. They are closed downward, and closed under disjunction: if $a, b \in I$ then $a \vee b \in I$.)

○ If $\mathcal{A}$ contains disjunction and $\bot$, then $\mathcal{A}^*$ is the set of nonempty ideals in $\mathcal{A}$ — every ideal must contain $\bot$, the least element of $\mathcal{A}$.

$\mathcal{A}^*$ is a complete lattice — order is subsethood, the conjunction of a class of elements is their intersection, and the disjunction of a class of elements is the intersection of all elements above each element in that class. It is not difficult to show that it is completely distributive if the original lattice contains no counterexample to distributivity. If fusion is present in $\mathcal{A}$ then it is present in $\mathcal{A}^*$ too.

$$I \circ J = \{z : \exists x \in I, y \in J(z \leqslant x \circ y)\}$$

and other connectives lift in a similar way. The structural rules of $\mathcal{A}$ are preserved in $\mathcal{A}^*$.[79] This shows that $\mathcal{A}^*$ has the nice logical properties of $\mathcal{A}$.

However, since $\mathcal{A}^*$ is a complete lattice, we can do interesting things with it. If $\mathcal{A}$ doesn't contain a truth element t as a left identity for fusion, $\mathcal{A}^*$ still does. Since $\mathcal{A}^*$ is complete, you can set t to be $\bigwedge\{I \to J : I \leqslant J\}$. Then $t \circ I \leqslant J$ if and only if $t \leqslant I \to J$ if and only if $I \leqslant J$, and so, t is a left identity for fusion.

Furthermore, the map from $\mathcal{A}$ to $\mathcal{A}^*$ which sends a to $\downarrow a = \{x : x \leqslant a\}$ injects one structure into the other, preserving all of the operations in $\mathcal{A}$. Any consecution with a counterexample in $\mathcal{A}$ will have a counterexample in $\mathcal{A}^*$ too. It shows that linear logic without the additives is conservatively extended by additives which distribute, for example.

EXAMPLE 30 (Pasting $\mathcal{A}$ and $\mathcal{A}^{op}$ together). Modifying a structure in such a way as to add negation is more difficult. To add a De Morgan negation to a structure, we need an *upside down copy* $\mathcal{A}^{op}$ of $\mathcal{A}$ so that negation can be an order inverting map of period two. Following the details of this construction will be a great deal easier if we take $\mathcal{A}$ to include top and bottom elements, so from now I will do so.

---

[79]The proof is tedious but straightforward [Restall, 2000a, ch. 9].

Conjunction and disjunction in $\mathcal{A}^{op}$ can be defined as the De Morgan dual of that in $\mathcal{A}$. So, if $a, b \in \mathcal{A}^{op}$ then $a \wedge b = -(-a \vee -b)$, where $-$ is the natural map from $\mathcal{A}$ to $\mathcal{A}^{op}$ and back, sending an $\mathcal{A}$ object to its shadow in the copy $\mathcal{A}^{op}$ and *vice versa*. Defining conjunctions and disjunctions of elements between $\mathcal{A}$ and $\mathcal{A}^{op}$ depends on another decision we need take. If $a \in \mathcal{A}$ and $b \in \mathcal{A}^{op}$, then we need decide on what we take $a \wedge b$ and $a \vee b$ to be. There are three options for this, each depending on the relevant *positioning* of $\mathcal{A}$ and $\mathcal{A}^{op}$. Meyer's original choice [1973a] was to put $\mathcal{A}^{op}$ *above* $\mathcal{A}$. Then the disjunction of an element form $\mathcal{A}$ with an element from $\mathcal{A}^{op}$ will be the element from $\mathcal{A}^{op}$ and the conjunction will be the element from $\mathcal{A}$. This choice (rather than putting $\mathcal{A}^{op}$ *under* $\mathcal{A}$) is the one to take if you wish to end up with a model for the relevant logic R, for we wish to end up with $t \vdash A \vee -A$. The element t is in $\mathcal{A}$, and we wish it to be under each $a \vee -a$. But $a \vee -a$ can be *any* element in the top half of the model, so t must be under each element in the top half, so it is either the *bottom element* of the top half of the model (not likely, if any conditional is untrue at all in the original model) or it is in the bottom half.

The *other* choice for ordering the two components — putting $\mathcal{A}^{op}$ below $\mathcal{A}$ — is required if you wish the original model to satisfy K. Then, t must be $\top$, and since t is in the original model, it must be at the top of the new model, so $\mathcal{A}^{op}$ can go underneath.

There is one other natural choice for the ordering of $\mathcal{A}$ and $\mathcal{A}^{op}$, and that is to take them in *parallel*. You can paste together the top elements of both models and the bottom elements of both models (or add new top and bottom elements if you prefer) and then take the disjunction of a pair, one from $\mathcal{A}$ and $\mathcal{A}^{op}$, to be the $\top$ element of the whole structure, and the conjunction of that pair to be the $\bot$ element. This is another natural option, but the resulting lattice is not distributive if $\mathcal{A}$ is not trivial.

To make the resulting structure a model for a logic, you must define $\circ$ and $\rightarrow$ in the whole structure. Most choices are fixed in advance, if fusion is commutative in $\mathcal{A}$. Since we want $a \circ b$ to be $-(a \rightarrow -b)$, and $a \rightarrow b$ to be $-b \rightarrow -a$, we take $a \circ b$ when $a \in \mathcal{A}$ and $b \in \mathcal{A}^{op}$ to be $-(a \rightarrow -b)$ (and the dual choice when $a \in \mathcal{A}^{op}$ and $b \in \mathcal{A}$). The remaining choice is for $a \circ b$ where $a, b \in \mathcal{A}^{op}$. Here, it depends on the relative position of $\mathcal{A}$ and $\mathcal{A}^{op}$. If we add the new structure on top, take $a \circ b$ to be $\top$. If we add the new structure below, or alongside, take $a \circ b$ to be $\bot$. The new structure satisfies many of the structural rules of the old structure, and as a result, a conservative extension result for logics in the vicinity of R follows [Meyer, 1970b; 1973a; Restall, 2000a].

There is one substructural logic for which a conservative extension by negation fails: RM. RM is given by extending R with the mingle rule $A \circ A \rightarrow A$ (or equivalently, $A \vdash A \rightarrow A$). If you add mingle to positive R then the result is still a sublogic of intuitionistic logic, and as a result, total ordering $t \vdash (A \rightarrow B) \vee (B \rightarrow A)$ is not provable. This logic is called RM0. In the presence of negation, however, the addition of mingle brings along with itself the total ordering principle. (This result is due to Meyer and Parks, from 1972 [1972].)

3.2   *Categories*

In propositional structures, we abstract away from the particulars of the languages in which our propositions are expressed to focus on the propositions themselves, ordered under entailment. In propositional structures, propositions are first-class citizens, and *proofs* between propositions fade into the background. If there is a proof from A to B, then $[\![A]\!] \leqslant [\![B]\!]$. The differences between proofs from A to B are not registered in this algebraic semantics.

Models do not have to be like this. We can consider not only propositions as objects but also proofs as "arrows" between objects. If we have one proof from A to B, we might indicate this as '$f : A \longrightarrow B$,' where f is the proof. We might have another proof $g : B \longrightarrow C$, and then we could *compose* them to construct another proof $gf : A \longrightarrow C$, which runs though f and then g.

Logicians did not have to go to the trouble of inventing structures like this. It turns out that mathematical objects with just these properties have been widely studied for many decades. Categories are important mathematical structures. Category theory is a helpful language for describing constructions which appear in disparate parts of mathematics. This means that category theory is, by its nature, very abstract. This also means that category theory is rich in examples, interesting categories are models of substructural logics. In particular, I will look at one example categorical model of a logic, Girard's model of *coherence spaces*, for linear logic.[80]

To understand the role of categories as models of logic, you need to focus on one particular part of categorical technology: the *adjoint pair*. An adjoint pair is a relationship between two *functors*, and functors are structure preserving maps between categories. Thinking of a category as a model of a logic generalising an algebra, the operators such as fusion, implication and so on are all *functors* from the category to itself (or perhaps, from the category to its *opposite*, which is found by swapping arrows from a to b to go from b to a instead). Operators like fusion, which are isotonic, are really two-place maps from a category to itself, not only sending a pair of category objects to another object (their fusion) but *also* sending arrows $f : a \longrightarrow a'$ and $g : b \longrightarrow b'$ to an arrow $f \circ g : a \circ b \longrightarrow a \circ b'$.

EXAMPLE 31 (Adjunction between Fusion and Implication). In *cartesian closed categories*,[81] product: $- \times B$ is a functor $C \longrightarrow C$. Similarly $[B \Rightarrow -]$ is a functor $C \longrightarrow C$. These functors form an adjunction. If $f : A \times B \longrightarrow C$, then $\lambda f : A \longrightarrow [B \Rightarrow C]$. Conversely, if $g : A \longrightarrow [B \to C]$, then $\mathrm{ev}(g \times \mathrm{id}_B) : A \times B \longrightarrow C$. This is a bijection $\mathrm{Hom}(A \times B, C) \cong \mathrm{Hom}(A, [B \Rightarrow C])$.

---

[80]In a history like this I can only assume some category theory, and not introduce it myself. Here are some standard references: Mac Lane's *Categories for the Working Mathematician* is a very good introduction to the area [1971], readable even by those who are not working mathematicians. Barr and Wells' *Category Theory for Computing Science* is also clear, from a perspective of the theory of computation [1990]. Chapter 10 of *An Introduction to Substructural Logics* [Restall, 2000a] contains *just* the category theory you need to go through the detail of this model. Došen's paper "Deductive Completeness" is a clear introduction focussing on the use of categories in logic [1996].

[81]I can't tell you what these are, for lack of space.

This is the categorical equivalent of the residuation between extensional conjunction and intuitionistic implication. Cartesian closed categories are models of intuitionistic logic [Lambek and Scott, 1986].

Coherence Spaces

Coherence spaces arise as a model of the λ-calculus, and intuitionistic logic. They provided the first model which gave Girard an insight into the decomposition of intuitionistic implication in terms of *linear* implication and the exponential ! [1987a; 1989].

DEFINITION 32 (Coherence Spaces). A *coherence space* is a set $\mathcal{A}$ of sets, satisfying the following two conditions.

○ If $a \in \mathcal{A}$ and $b \subseteq a$ then $b \in \mathcal{A}$, and
○ If for each $x, y \in a$, $\{x, y\} \in \mathcal{A}$, then $a \in \mathcal{A}$.

But coherence spaces are much better thought of as undirected graphs. We say *a coheres with b* (in $\mathcal{A}$) if $\{x, y\} \in \mathcal{A}$. We write this: '$x \frown y (\bmod \mathcal{A})$.' The coherence relation determines the coherence space completely. Coherent sets ($a \in \mathcal{A}$) are cliques in the graph. The coherence relation is reflexive and symmetric, but not, in general, transitive.[82]

Given a coherence space $\mathcal{A}$, we define *coherence relations* as follows:

○ $x \frown y (\bmod \mathcal{A})$ iff $x \frown y (\bmod \mathcal{A})$ and $x \neq y$.
○ $x \smile y (\bmod \mathcal{A})$ iff $\{x, y\} \notin \mathcal{A}$.
○ $x \asymp y (\bmod \mathcal{A})$ iff it is not the case that $x \frown y (\bmod \mathcal{A})$.

DEFINITION 33 (Product, Sum and Negation Spaces). Given spaces $\mathcal{A}$ and $\mathcal{B}$, the coherence spaces $\mathcal{A} \wedge \mathcal{B}$ and $\mathcal{A} \vee \mathcal{B}$ are defined on the disjoint union of the points x of the graph of $\mathcal{A}$ and y of the graph of $\mathcal{B}$, as follows:

$$
\begin{array}{ll}
(0, x) \frown (0, x') (\bmod \mathcal{A} \wedge \mathcal{B}) & \text{iff } x \frown x' (\bmod \mathcal{A}) \\
(1, y) \frown (1, y') (\bmod \mathcal{A} \wedge \mathcal{B}) & \text{iff } y \frown y' (\bmod \mathcal{B}) \\
(0, x) \frown (1, y) (\bmod \mathcal{A} \wedge \mathcal{B}) & \text{always}
\end{array}
$$

$$
\begin{array}{ll}
(0, x) \frown (0, x') (\bmod \mathcal{A} \vee \mathcal{B}) & \text{iff } x \frown x' (\bmod \mathcal{A}) \\
(1, y) \frown (1, y') (\bmod \mathcal{A} \vee \mathcal{B}) & \text{iff } y \frown y' (\bmod \mathcal{B}) \\
(0, x) \frown (1, y) (\bmod \mathcal{A} \vee \mathcal{B}) & \text{never}
\end{array}
$$

Given a coherence space $\mathcal{A}$, the coherence space $\sim A$ is defined by setting $x \frown y (\bmod \sim \mathcal{A})$ if and only if $x \asymp y (\bmod \mathcal{A})$. Note that $\sim \sim \mathcal{A} = \mathcal{A}$. $\mathsf{Sgl} = \{\emptyset, \{*\}\}$, an arbitrary one-point coherence space. $\mathsf{Emp} = \{\emptyset\}$, the empty coherence space. Note that $\sim\mathsf{Sgl} = \mathsf{Sgl}$ and $\sim\mathsf{Emp} = \mathsf{Emp}$.

---

[82]Erhard's *hypercoherences* are a generalisation of coherence spaces which are richer than a graph represents [1993]. In hypercoherences, a might be a coherent set without $a' \subseteq a$ also being coherent. The category of hypercoherences is also a model of linear logic.

Note that here $\sim(\mathcal{A} \wedge \mathcal{B}) = \sim\mathcal{A} \vee \sim\mathcal{B}$, and $\sim(\mathcal{A} \vee \mathcal{B}) = \sim\mathcal{A} \wedge \sim\mathcal{B}$. Furthermore, $\mathcal{A} \wedge \mathsf{Emp} = \mathsf{Emp} \wedge \mathcal{A} = \mathcal{A} = \mathcal{A} \vee \mathsf{Emp} = \mathsf{Emp} \vee \mathcal{A}$. Emp does the job of both $\top$ and $\bot$ in the category of coherence spaces.[83]

The class of all coherence spaces can be made into a cartesian closed category, if we take the arrows to be *continuous functions*.

DEFINITION 34 (Continuous Functions). $F : \mathcal{A}{\longrightarrow}\mathcal{B}$ is *continuous* if and only if

○ If $a \subseteq b$ then $F(a) \subseteq F(b)$.

○ If $S \subseteq \mathcal{A}$ is *directed* (that is, if $a, b \in S$, then $a \cup b \in S$ too) then $F(\bigcup S) = \bigcup\{F(a) : a \in S\}$.

FACT 35 (Minimal Representatives). If $F : \mathcal{A}{\longrightarrow}\mathcal{B}$ is continuous, and if $a \in \mathcal{A}$ and $y \in F(a)$, then there is a *minimal* finite $a' \in \mathcal{A}$ where $y \in F(a')$.

Proof. If $y \in F(a)$ then $y \in F(a^*)$ for some finite a^*. Pick some *smallest* subset a' of a^* with this property. (This is possible, as a^* is finite.) ■

We want to construct a coherence space representing $F : \mathcal{A}{\longrightarrow}\mathcal{B}$. We start by defining the trace of a function.

$$\mathrm{Trace}(F) = \{(a, y) \in \mathcal{A}_{\mathsf{fin}} \times |\mathcal{B}| : y \in F(a) \text{ and } a \text{ is minimal}\}$$

Note that $\mathrm{Trace}(F) \subseteq \mathcal{A}_{\mathsf{fin}} \times |\mathcal{B}|$ has the following properties.

○ If $(a, y), (a, y') \in \mathrm{Trace}(F)$ then $y \smile y' (\mathrm{mod}\ \mathcal{B})$.

○ If $a' \subseteq a$, $(a, y), (a', y) \in \mathrm{Trace}(F)$, then $a = a'$.

Conversely, if $\mathfrak{F}$ is any set with these two properties, then define $F_{\mathfrak{F}}$ by setting

$$F_{\mathfrak{F}}(a) = \{y \in |\mathcal{B}| : \exists a' \subseteq a \text{ where } (a', y) \in \mathfrak{F}\}.$$

We can represent continuous functions by their traces. In fact, if F is continuous, then $F = F_{\mathrm{Trace}(F)}$. Can we define a coherence relation on traces? Consider the special case where there are *two* minimal representatives, that is, $(a, y), (a', y) \in \mathrm{Trace}(F)$. Under what circumstances are they coherent? Unfortunately, we need more information in order to define a coherence relation — we need a relationship between a and a'. We can show that in a particular class of continuous functions, there is always a *unique* minimal a.

DEFINITION 36 (Stable Functions). $F : \mathcal{A}{\longrightarrow}\mathcal{B}$ is *stable* if it is continuous, and in addition, whenever $a, a', a \cup a' \in \mathcal{A}$, then $F(a \cap a') = F(a) \cap F(a')$.

With stable functions, we can choose a unique minimal representative a.

FACT 37 (Unique Minimal Representatives). $F : \mathcal{A}{\longrightarrow}\mathcal{B}$ is stable if and only if for each $a \in \mathcal{A}$, where $y \in F(a)$, there is a unique minimal $a' \in \mathcal{A}_{\mathsf{fin}}$ such that $y \in F(a')$.

---

[83]This shows how categories have a kind of flexibility unavailable to posets. In a poset, $\top = \bot$ only if the poset is trivial. In a category, $\top$ and $\bot$ might be identical or isomorphic, without the category structure being trivial. Yes, there will be arrows from every object to every other object, but it is not the case that all objects are isomorphic.

Proof. For left to right, it is straightforward to check that $a' = \bigcap\{a^* \in \mathcal{A} : a^* \subseteq a,$ where $y \in F(a^*)\}$ is the required a'. For right to left, monotonicity tells us that $F(a \cap a') \subseteq F(a)$ and $F(a \cap a') \subseteq F(a')$, so $F(a \cap a') \subseteq F(a) \cap F(a')$. Conversely, if $a, a', a \cup a' \in \mathcal{A}$, then if $y \in F(a)$ and $y \in F(a')$, then y is in $F(a'')$ for a unique minimal a''. Therefore $a'' \subseteq a$ and $a'' \subseteq a'$, so $a'' \subseteq a \cap a'$, and hence $y \in F(a'') \subseteq F(a \cap a')$, as desired. ∎

The next result is simple to verify.

FACT 38 (Characterising Stable Functions). If F is stable, then whenever $(a, y), (a', y') \in \text{Trace}(F)$

- If $a \cup a' \in \mathcal{A}$ then $y \bigcirc y' (\text{mod } \mathcal{B})$.
- If $a \cup a' \notin \mathcal{A}$ then $y = y' (\text{mod } \mathcal{B})$.

Conversely, if the set $\mathfrak{F}$ satisfies these conditions, then $F_{\mathfrak{F}}$ is stable. □

Given this, we can define $\mathcal{A} \supset \mathcal{B}$. $|\mathcal{A} \supset \mathcal{B}| = \mathcal{A}_{\text{fin}} \times |\mathcal{B}|$ as follows: $(a, y) \bigcirc (a', y')$ mod $\mathcal{A} \supset \mathcal{B}$ if and only if

- If $a \cup a' \in \mathcal{A}$ then $y \bigcirc y' (\text{mod } \mathcal{B})$.
- If $a \cup a' \in \mathcal{A}$ then $y = y' (\text{mod } \mathcal{B})$.

That is, $\mathcal{A} \supset \mathcal{B} = \{\text{Trace}(F) \mid F : \mathcal{A} \longrightarrow \mathcal{B} \text{ is stable}\}$.

The category of coherent spaces and stable functions between them is cartesian closed. This construction is *obviously* a two-stage process. It *begs* to be decomposed. We should define a coherence space $!A$ on the set of finite coherent sets of $\mathcal{A}_{\text{fin}}$ as follows:

$$a \bigcirc a' (\text{mod } !A) \text{ iff } a \cup a' \in \mathcal{A}$$

and define *linear* implication $\mathcal{A} \to \mathcal{B}$ by setting $(x, y) \bigcirc (x', y') (\text{mod } \mathcal{A} \to \mathcal{B})$ if and only if

- If $x \bigcirc x' (\text{mod } \mathcal{A})$ then $y \bigcirc y' (\text{mod } \mathcal{B})$.
- If $x \bigcirc x' (\text{mod } \mathcal{A})$ and $y = y'$ then $x = x'$.

Note that $\mathcal{A} \to \mathcal{B}$ is (isomorphic to) $\sim\mathcal{B} \to \sim\mathcal{A}$. Furthermore, $\mathsf{Sgl} \to \mathcal{A}$ is (isomorphic to) $\mathcal{A}$ and $\mathcal{A} \to \mathsf{Sgl}$ is (isomorphic to) $\sim\mathcal{A}$. The operation $\supset$ stands to stable functions as $\to$ stands to a new kind of function: the *linear* functions.

DEFINITION 39 (Linear Maps). F is a *linear* map if and only if whenever $A \subseteq \mathcal{A}$ is linked (that is, if $a, b \in A$ then $a \cup b \in A$) then $F(\bigcup A) = \bigcup\{F(a) : a \in A\}$.

If F is linear then F is stable (this is straightforward) and in addition, if $x \in F(a)$ then the minimal b where $x \in F(b)$ is a *singleton*. It follows that the trace of F can be simplified. The *linear trace* of a linear map F is defined as follows:

$$\text{Trlin}(F) = \{(x, y) : y \in F(\{x\})\}.$$

Therefore, $\mathcal{A} \to \mathcal{B} = \{\text{Trlin}(F) \mid F : \mathcal{A} \longrightarrow \mathcal{B} \text{ is linear}\}$.

Given $\to$, we can see that it is connected by an adjunction to a natural fusion operation. We can define $\mathcal{A} \circ \mathcal{B}$ as follows: $|\mathcal{A} \circ \mathcal{B}| = |\mathcal{A}| \times |\mathcal{B}|$, and $(x, y) \frown (x', y')$ if and only if $x \frown x' \,(\mathrm{mod}\ \mathcal{A})$ and $y \frown y' \,(\mathrm{mod}\ \mathcal{B})$.

FACT 40 (The Adjunction between Fusion and Implication). In the category of coherence spaces with linear maps

$$\mathrm{Hom}(\mathcal{A} \circ \mathcal{B}, C) \cong \mathrm{Hom}(\mathcal{A}, \mathcal{B} \to C)$$

is an adjunction for all $\mathcal{A}$, $\mathcal{B}$ and C.

This, with the associativity and commutativity of $\circ$, together with the behaviour of !, shows that the category of coherence spaces and linear maps is a model of linear logic.

Some very recent work of Schalk and de Paiva's on *poset-valued sets* [2004] generalises coherence spaces in an interesting direction. They show that coherence spaces and hypercoherences can be seen as maps from $\mathsf{Set} \times \mathsf{Set}$ to the algebra $\mathsf{RM3}$.[84] If $x \frown y$ then $f(x, y) = \mathsf{t}$, if $x \smile y$ then $f(x, y) = \mathsf{f}$, if $x = y$ then $f(x, y) = \mathsf{b}$. The logical operators of negation, fusion and implication then lift from the algebra to the coherence spaces. (In other words, if $\mathcal{A} : \mathsf{Set} \times \mathsf{Set} \to \mathsf{RM3}$ is a coherence space, then $\sim\!\mathcal{A}$ is the map composing $\mathcal{A}$ with $\sim : \mathsf{RM3} \to \mathsf{RM3}$. The same goes for the other operations.) Different categorical models in the style of coherence spaces can then be given by varying the target algebra. I suspect that using some of the algebras known in the substructural literature will lead to interesting categorical models of linear logic and related systems.

Girard has shown that *Banach spaces* can be used in place of coherent spaces to model linear logic [1996]. The norm in a Banach space takes the place of the coherence relation. As we shall see later, it is not the only point at which geometric intuitions have come to play a role in substructural logic.

3.3   *Frames*

The study of modal logic found new depth and vigour with the advent of possible worlds semantics. As we have seen, algebras are useful models of substructural logics. However, they are so *close* to the proof theory of these logics that they do not provide a great deal of new information, either about the intrinsic properties of the logic in question, or about how it is to be applied. Models in terms of *frames* are one way to extract more information. Perhaps this is because frames are a further step removed from the logic in an important sense. In algebras, each formula in the language is interpreted as an element in the algebra. In frames, each formula is not interpreted as an element in the frame — the elements in the frame lie *underneath* the interpretation of formulas. Formulas are interpreted as *collections* of frame elements. Therefore the interpretations of connectives on a frame are themselves *decomposed*. They are no longer simply *functions* on algebras satisfying specified conditions. Their action on sets of frame elements is factored through their action on individual frame elements. As a result, frame interpretations of logics can be thought to carry more information than algebras. In addition, frame semantics is

---

[84]They do not recognise that the algebra is already quite studied in the relevant logic literature.

suggestive of applications of logics. Just as the idea of the interpretation of a proposition as a set of possible worlds, or a set of times or a set of locations has driven the application of different models of modal or temporal logics, so the interpretation of frame semantics for substructural logics has led to their use in diverse applications. But enough of scene-setting. Let's start with the first attempts to give precise frame semantics for substructural logics. As before, our story starts with the relevant logic R.

Operational Frames

The idea of frame semantics for relevant logics occurred independently to Routley and to Urquhart in the late 1960s and early 1970s. Routley's techniques are more general than Urquhart's, but Urquhart's were published first, and are the simplest to introduce, so we will start with them.

Consider the constraints for developing a frame semantics for a relevant logic. The bare bones of any frame semantics are as follows. A *frame* is a set of objects (call them *points*, though "worlds", "situations", "set-ups" and other names have all been used), and a *model* on that frame is a relation ⊩ which indicates what formulas are true at what points. We read "$x \Vdash A$" as "A is true at x." Typically, the relation ⊩ is constrained by inductive clauses that indicate how the truth (or otherwise) of a complex formula at each point is determined by the truth (or otherwise) of its subformulas. Given a particular model, then, we say that *A entails B* on that model if and only if for every point x, if $x \Vdash A$ then $x \Vdash B$. Entailment is preservation of truth at all points in a model.

This is the bare bones of a frame semantics for a logic. Consider how this determines what we can do to interpret relevant implication. It is axiomatic for a relevant logic that the entailment from A to $B \rightarrow B$ can fail. In frame terms this means that we must have points in our models in which $B \rightarrow B$ can fail. This means that the interpretation of implication must differ from any kind of frame interpretation of conditionals seen before. For a *strict* conditional $A \Rightarrow B$ to be true at a world, we need to check all accessible worlds, to see if B is true whenever A is true. As a result, $B \Rightarrow B$ is true at every world. Similarly, for *counterfactual* conditionals $A \;\Box\!\!\rightarrow\; B$, we check the *nearby* worlds where A is true, to see if B is true there too. Again, $B \;\Box\!\!\rightarrow\; B$ is true, because we check the consequent at the very same points in the model where we have taken the antecedent to be true. Something different must be done for a relevant conditional. At the very least we need to check the value of the consequent somewhere at places other than simply where we have checked the antecedent.

Urquhart's innovation was a natural way to do just this [1972a; 1972b; 1972c; 1972d]. Consider again what an implication $A \rightarrow B$ says. To be committed to $A \rightarrow B$ is to be committed to B whenever we gain the information that A. To put it another way, a body of information warrants $A \rightarrow B$ if and only if whenever you *update* that information with new information which warrants A, the resulting (perhaps new) body of information warrants B. Putting this idea in technical garb, we get a familiar-looking inductive clause from a frame semantics:

　○ $x \Vdash A \rightarrow B$ if and only if for each y, if $y \Vdash A$ then $x \sqcup y \Vdash B$.

But this inductive clause has a new twist. Unlike the clauses for strict or counterfactual conditionals, in this clause we check the consequent and antecedent at different points in the model structure. The way is open for $B \rightarrow B$ to fail.

Let's take some time to examine the detail of this clause. We have a class of points (over which "x" and "y" vary), and a function $\sqcup$ which gives us new points from old. The point $x \sqcup y$ is supposed, on Urquhart's interpretation, to be the body of information given by *combining* x with y. The properties we take combination to have will influence the properties of the conditional. First up, let's consider our old enemy, $A \vdash B \rightarrow B$. For this to fail, we need to have a point x where $x \nVdash B \rightarrow B$, and for this, we need just some y where $y \Vdash B$ and $x \sqcup y \nVdash B$. This means that combination of bodies of information cannot satisfy this *hereditary* condition:

> ○   If $x \Vdash A$ then $x \sqcup y \Vdash A$ left hereditary condition.

Similarly, if we are to have $A \vdash B \rightarrow A$ to fail, then combination cannot satisfy the *dual* hereditary condition.

> ○   If $x \Vdash A$ then $y \sqcup x \Vdash A$ right hereditary condition.

This means that combination is sometimes *nonmonotonic* in a natural sense. Sometimes when a body of information is combined with another body of information, some of the original body of information might be lost. This is simplest to see in the case motivating the failure of $A \vdash B \rightarrow A$. A body of information might tell us that A. However, when we combine it with something which tells us B, the resulting body of information might no longer warrant A (as A might conflict with B). Combination might not simply result in the *addition* of information. It may well warrant its *revision*.

To model the logic R, combination must satisfy a number of properties:

> ○ $x \sqcup y = y \sqcup x$ commutativity
> ○ $(x \sqcup y) \sqcup z = x \sqcup (y \sqcup z)$ associativity
> ○ $x \sqcup x = x$ *idempotence*.

Commutativity gives us *assertion*, associativity gives us *prefixing* and *suffixing*, and *idempotence* gives us *contraction*, as is easily verified. For example, consider assertion: to verify that $A \vdash (A \rightarrow B) \rightarrow B$, suppose that $x \Vdash A$. To show that $x \Vdash (A \rightarrow B) \rightarrow B$, take a y where $y \Vdash A \rightarrow B$. We wish to show that $x \sqcup y \Vdash B$. By commutativity, $x \sqcup y = y \sqcup x$, and since $y \Vdash A \rightarrow B$ and $x \Vdash A$, we can apply the conditional clause at y to give $y \sqcup x \Vdash B$. So, $x \sqcup y \Vdash B$ as desired.

These frame properties, governing the behaviour of $\sqcup$ are very similar in scope to the structural rules governing fusion and intensional combination in different proof theories. This is no surprise, as $\sqcup$ is the frame analogue of fusion. It comes as no surprise, then, that as you vary conditions on $\sqcup$ you can model different substructural logics.

Keeping the analogy afloat, then, we can see how these models might interpret *theoremhood* in our logics. In analogy with the proof theory and algebraic models of our logics, we can see that there are two different grades of truth. It is one thing for a formula

to be true everywhere in a model — this corresponds to being entailed by the Church true constant $\top$. It is another thing for it to be a tautology, for it to be entailed by the Ackermann true constant t. Identities are entailed by t. What corresponds to being a tautology in this sense in our models? Clearly being true at every point is ruled out, as identities can fail at different points in a model. Continuing the interpretation of points as bodies of information, if we can have bodies of information which do *not* warrant all of the tautologies of logic, then we need some way of talking about which bodies of information do. The simplest approach (and the one which Urquhart took) is to take a special body of information 0 to stand for "logic." A natural condition to take on 0 is that it is a left identity for composition

$$\circ \quad 0 \sqcup x = x \quad \text{left identity}$$

In this way, $0 \Vdash A \to B$ if and only if for each x, if $x \Vdash A$ then $x \Vdash B$ — so the conditionals warranted by *logic* correspond to exactly the entailments valid in the frame. The identity point 0 does a good job of modelling logic.

The interpretation of points as bodies of information warrants a simple interpretation of conjunction as well. The usual clause

$\circ$ $x \Vdash A \wedge B$ if and only if $x \Vdash A$ and $x \Vdash B$

is uncontroversial. If a body of information warrants $A \wedge B$, it warrants A and it warrants B, and conversely. Adding this clause to the semantics gives us the conjunction and implication fragment of R (and its neighbours, varying the behaviour of of $\sqcup$).

Intensional conjunction is also straightforward. We can add fusion as the object-language witness of composition:

$\circ$ $x \Vdash A \circ B$ if and only if for some y, z where $x = y \sqcup z$, $y \Vdash A$ and $z \Vdash B$.

It is instructive to verify that in these models, that $A \circ B \vdash C$ if and only if $A \vdash B \to C$. Residuation between $\to$ and $\circ$ corresponds to the *universal* clause modelling $\to$ interacting with the *existential* clause modelling $\circ$.

Let's now turn to soundness and completeness with respect to these models. To prove soundness of a proof theory with respect to these models, it is required only to show that everything provable in the proof theory holds in the model (either holds at 0 for a Hilbert system, or holds over the entire frame for a proof theory which delivers consecutions). As usual, verifying soundness is a straightforward matter of checking axioms and rules.

There are two different ways to prove the completeness of a proof theory with Urquhart's operational models. Again, as usual, the common technique is to provide a *counterexample* for an unprovable formula (or consecution). Both techniques use a *canonical model* construction, familiar from the worlds semantics for modal logics. Where these constructions differ is in the stuff out of which the points in the model are made. The first, and most general kind of canonical model we can provide for an operational semantics is the *theory* model, in which the points are all of the *theories* of the logic in question.

DEFINITION 41 (The *Theory* Canonical Model). The set of points is the set $\mathcal{T}$ of theories. The identity theory is the set L of all of the tautologies of the logic. The composition

relation ⊔ is defined as follows:

$$S \sqcup T = \{B : (\exists A)(A \to B \in S \text{ and } A \in T)\}$$

and $S \Vdash A$ if and only if $A \in S$.

To verify that the *theory* canonical model is indeed a canonical model we must show that ⊔ so defined satisfies all of the conditions of a composition relation, and that ⊩ satisfies the recursive conditions of an evaluation relation.

To show that ⊔ satisfies the conditions of composition, you need first show that ⊔ is indeed a *function* on the class of theories: that if S and T are theories, so is $S \sqcup T$. The verification of this fact is elementary. The frame conditions on ⊔ correspond quite neatly to axioms or structural rules.

To show that ⊩ satisfies the recursive conditions, you need show that $A \wedge B \in T$ if and only if $A, B \in T$ (which is an immediate consequence of the definition of a theory) and that $A \to B \in S$ if and only if for each T where $A \in T$, $B \in S \sqcup T$. The verification from left to right is an immediate consequence of the definition of ⊔. The verification from right to left is simplest to prove in the contrapositive: that if $A \to B \notin S$ then there is a T where $A \in T$ and $B \notin S \sqcup T$. Finding such a T is easy here: let $T = \{C : A \vdash C\}$. If $B \in S \sqcup T$ then there is some $C \in T$ where $C \to B \in S$. Since $A \vdash C$, it follows that $C \to B \vdash A \to B$ (by monotonicity of $\to$) and $A \to B \in S$ contrary to what we have assumed.

It is possible to extend this kind of completeness proof to show that the condition for fusion models this connective correctly too.

DEFINITION 42 (The *Finite Set* Canonical Model). The points are the finite *sets* of formulas. Composition ⊔ is set union. $\{A_1, \ldots, A_n\} \Vdash B$ if and only if

$$\vdash A_1 \to (\cdots \to (A_n \to B)).$$

(The permutation axiom shows that the order of presentation in the set is irrelevant in this definition.)

It is not difficult to show that this is indeed a model — that the recursive clause defining $\to$ is satisfied.

This is a simple model which gives a straightforward counterexample to any invalid argument. If $A \nvdash B$ then $\{A\}$ is the point in the model invalidating the argument: $\{A\} \Vdash A$ and $\{A\} \nVdash B$.

Operational frames are important models of other substructural logics too.

EXAMPLE 43 (Language frames). A *language* frame on alphabet $\mathcal{A}$ is the collection of all strings on that alphabet, with ⊔ defined as concatenation.

Language frames are a model of the Lambek calculus. The composition operation ⊔ is associative but not commutative (except in the case where $\mathcal{A}$ is a singleton). It was an open question for many years whether or not the Lambek calculus is complete for Language frames. Mati Pentus showed that it is, using an ingenious (and difficult!) model construction argument pasting chains of partial models together to form a string

model [1995; 1998]. Different frames for the Lambek calculus feature prominently in some recent work on the system and its linguistic applications [Moortgat, 1988; Morrill, 1994].

EXAMPLE 44 (Domain Spaces). Models of the λ-calculus [Abramsky and Jung, 1994; Gunter and Scott, 1990; Scott, 1973; 1980] are models for substructural logics too. Scott's famous model construction involves a topological space D such that D is isomorphic to the space $[D \rightarrow D]$ of continuous functions from D to itself. Each element of D is paired with a function in $[D \rightarrow D]$, so can think of the objects equally well as functions. Therefore, there is a two-place operation of application on the domain. Consider $x(y)$ — the application of x to y. We can assign types to functions in this model by "reading" the model as a frame for a logic. If we set $x \sqcup y$ to be $x(y)$, then this is an operational frame:

$$x \Vdash A \rightarrow B \text{ if and only if for each } y, \text{ where } y \Vdash A, x(y) \Vdash B.$$

In other words, x is of type $A \rightarrow B$ if and only if whenever given an input of type x, the output is of type B. This gives us a plausible notion of function *typing*. For example, $\lambda x.(x + 1)$ will have type $Even \rightarrow Odd$ and $Odd \rightarrow Even$. The function $\lambda x.\lambda y.(2x + y)$ has type $N \rightarrow (Odd \rightarrow Odd)$ (whatever number x is, if y is odd, so is $2x + y$) but it does not have type $Odd \rightarrow (N \rightarrow Odd)$ (if y is even, the output will be even, not odd). This is an example demonstrating the failure of the permutation-related rule: $A \rightarrow (B \rightarrow C) \vdash B \rightarrow (A \rightarrow B)$.[85]

This is an important model because it motivates the failure of not only commutativity of $\sqcup$ but also *associativity*. There is no sense in which $x(y(z))$ ought be equal to $(x(y))(z)$. Function typing models a very weak substructural logic.

Urquhart considered adding *disjunction* to operational frames, with the natural clause:

∘ $x \Vdash A \vee B$ if and only if $x \Vdash A$ or $x \Vdash B$.

However, this is not as satisfactory as its cousin for conjunction. For one thing, R models extended with this clause validate the following formula

(1) $(A \rightarrow B \vee C) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$

which is not valid in R.[86] Secondly, and more importantly, the interpretation in terms of *pieces of information* simply doesn't motivate the straightforward clause for disjunction. Pieces of information may well warrant disjunctions without warranting either disjunct. To interpret disjunction in operational models (and to get a logic in the vicinity of R or any of the other logics we are interested in) you can do one of two things. One approach, taken by Ono [1985; 1992], Došen [1988; 1989] and Wansing [1993], is to admit some kind of *closure operator* on the frame: $A \vee B$ is true not only at points where A is true

---

[85]We can use other connectives to expand the type analysis of terms. Conjunction clearly makes sense in this interpretation: $x \Vdash A \wedge B$ if and only if $x \Vdash A$ and $x \Vdash B$. In this way, we have models not only for typing functions with $\rightarrow$ but also with intersection. These are models for the Torino type system $\lambda \cap$ [Barendregt *et al.*, 1983; Coppo *et al.*, 1981; Hindley, 1983; Venneri, 1994].

[86]This is not to say that the operational semantics with this disjunction clause hasn't been investigated. See some interesting papers of Charlewood: [1978; 1981], following on from a result of Fine [1976].

and where *B* is true, but also at some more points, found by *closing* the original set under some operation. Doing this will almost invariably invalidate distribution, and we will look at one example of this kind of semantics in a couple of section's time, when we come to phase space models for linear logic.

A related method, and one which validates distribution, was discovered by Kit Fine in the mid 1970s [1974; 1988]. He showed that if you have a *two tiered* collection of points, the whole class *S* with a special subset *P* of *prime* points (in analogy with prime theories) which respect disjunction. For points in *P*, a disjunction is true if and only if at least one disjunct is. For arbitrary points in *S* this may fail. For an arbitrary point in *S*, however, you have a guarantee that it can be *covered* by a point in *P*. For each $s \in S$ there is at least one $s' \in P$ where $s \sqsubseteq s'$. This means that disjunctions are at least promissory notes: although a disjunct may not be true given *this* body of information, it is possible for the information to be filled out so that you get one or other disjunct. Then, a disjunction, in a Fine model, is evaluated like this:

○ $x \Vdash A \vee B$ if and only if for each $y \in T$ where $x \sqsubseteq y$, $y \Vdash A$ or $y \Vdash B$.

Fine's models will satisfy distribution, and model the positive fragment of R nicely. They do so at the cost of requiring *special* bodies of information, those which are prime. They also have the cost of requiring a new notion $\sqsubseteq$, of informational inclusion. This requires a new condition on frames, the hereditary condition, familiar from models for intuitionistic logic:

DEFINITION 45 (Hereditary Condition). If $x \Vdash p$ and $x \sqsubseteq y$ then $y \Vdash p$ too.

To show that the hereditary condition extends from atomic propositions to all propositions, a further model condition is required to validate the inductive step for the conditional. You need to assume that

$$\text{If } x \sqsubseteq x' \text{ and } y \sqsubseteq y' \text{ then } x \sqcup y \sqsubseteq x' \sqcup y'.$$

Given this clause, we indeed have a model for positive R. The cost has been a complication of the clause for disjunction, the requirement that we have a two-tiered universe of points, and a hereditary condition on points. This is not the only way to model the whole of R. Routley and Meyer, independently of Fine, came to an equally powerful semantics, with a slightly smaller set of primitive notions. Before looking at the Routley–Meyer semantics in the next section, I must say a little about negation in the operational semantics.

How one interprets negation depends to a great extent on the intended interpretation. The *Boolean* clause for negation

○ $x \Vdash -A$ if and only if $x \nVdash A$

is marvellously appropriate in string models of the Lambek calculus (a string is of type "not a noun" just when it is not of type "noun") and in function typing (a function like $\lambda x.x^2$ has type $-\text{Even} \rightarrow -\text{Even}$: it sends inputs which are not even to outputs which are also not even) but it is terrible when it comes to taking points as bodies of information. There is little reason to think that a body of information *x* warrants the negation of *A* just when it fails to warrant *A*. Bodies of information can be incomplete (warranting neither

a claim nor its negation) and they can be inconsistent (warranting — you might think misleadingly — a claim and its negation). Something else too must be done to model negation. Fine had a treatment of negation in his models, but it too appeals essentially to the two-tiered nature of a model, and it is simpler in the Routley–Meyer incarnation.

Routley–Meyer Frames

Routley and Meyer [1972a; 1972b; 1973; 1982] chose to keep the interpretation for disjunction simple, and to generalise the interpretation for implication. The central feature of a Routley–Meyer frame is the *ternary relation R*. The clause for implication then is:

○ $x \Vdash A \rightarrow B$ if and only if for each y, z where $Rxyz$, if $y \Vdash A$ then $z \Vdash B$.

This is a generalisation of the operational semantics. An operational frame is a Routley–Meyer frame where $Rxyz$ holds if and only if $x \sqcup y = z$. The interpretation of R is similarly a generalisation of that for $\sqcup$. Reading the implication clause "in reverse" (as assigning meaning to R and not to $\rightarrow$)[87] we have that $Rxyz$ if and only if the *laws* (or conditionals) in x, applied to the *facts* (antecedents) in y give outcomes (consequents) true in z. Or more shortly, applying x to y gives an outcome included in z. That this is a genuine *relation* means that applying x to y might give no outcome at all. On the other hand, we might have $Rxyz$ and $Rxyz'$ for different z and z'. The result of applying x and y is no doubt a body of information, but it might not be a prime body of information. For example we might have $x \Vdash A \rightarrow B \vee C$ and $y \Vdash A$. Applying the information in x to that in y will give $B \vee C$, without giving us any guidance on which of B or C it is to be. And this is possible even if x is prime — for in R we don't have the counterintuitive entailment $A \rightarrow B \vee C \vdash (A \rightarrow B) \vee (A \rightarrow C)$, so we have no reason to think that x might contain either $A \rightarrow B$ or $A \rightarrow C$. So, in this case, we'd have two points z and z' where $Rxyz$ and $Rxyz'$. At z we can have B and at z' we can have C. In this way, we verify $A \rightarrow B \vee C$ at x, all the time using prime points.

We only have a semantics for the positive part of the logic R when endow the ternary relation R with some more properties. Routley and Meyer's original properties are best stated with the use of some shorthand.

○ "R^2abcd" is shorthand for $(\exists x)(Rabx \wedge Rxcd)$.
○ Given the distinguished point 0, we let "$a \sqsubseteq b$" be shorthand for $R0ab$.

These abbreviations make sense, given the interpretations of the concepts at hand. R^2abcd conjoins application. You apply a to b and get a result in x (for some x) which we then apply to c to get a result in d. One way of thinking of this is applying a to b and applying all of this to c. The inclusion relation is defined by looking at what happens when you apply *logic* to a state. Applying *logic* to a ought to result in nothing more than a. So, if $R0ab$ if and only if a is included in b.

---

[87]Which, frankly, is exactly what is done in cases of interpreting the accessibility relation in a modal logic as "relative possibility".

Fine has suggested writing *Rabc* as "$b \sqsubseteq_a c$", and reading it as: according to *a*, *b* is contained in *c* [1974].[88] In this case, $\sqsubseteq$ is $\sqsubseteq_0$, containment from the point of view of *logic*.

Here are the postulates Routley and Meyer gave to make their semantics model the logic R.

○ (Identity) *R0aa* for each *a*.

○ (Commutativity) If *Rabc* then *Rbac*.

○ (Pasch's Postulate) If R^2abcd then R^2acbd.

○ (Idempotence) *Raaa* for each *a*.

○ (Heredity) If *Rabc* and $a' \sqsubseteq a$ then *Ra'bc*.

These postulates parallel the postulates for $\sqcup$. Identity and heredity govern the behaviour of 0, making it fit to do the job of *t*, and to be a place to witness logical truths. Commutativity corresponds to the commutativity of fusion, Pasch's postulate corresponds to B′: an equivalent postulate, given commutativity, would be

○ (Associativity) If R^2abcd then $R^2a(bc)d$.

Where "$R^2a(bc)d$" is read as $(\exists x)(Rbcx \wedge Raxd)$.

Idempotence does the job of WI. So, we have a match with the postulates for an operational frame. And as with operational frames, ringing the changes with regard to the behaviour of *R* will result in different logical systems.

Soundness of Routley–Meyer models is a straightforward matter of showing that each provable consecution is valid on each model. (A valid consecution is, as usual, one which is preserved at every point.) To interpret consecutions, you must have an interpretation of fusion, but that is as you would expect.

○ $x \Vdash A \circ B$ iff there are y, z where *Ryzx*, $y \Vdash A$ and $z \Vdash B$.

A fusion is true at a point in a model when it is the composition of two points, at which the "fusejuncts" are, respectively, true. Logically true formulas are then always true at 0 in a Routley–Meyer model.

Demonstrating completeness, as always for a semantics like this, is much more involved. As usual, it is a canonical model construction. To construct a canonical model for a logic like R, instead of dealing with all *theories*, as we could with operational models, we must deal in *prime theories*.[89] But here, not just any prime theories will do. In these models, each point is closed under consequence as defined at the point 0. This is a fundamental fact about Routley–Meyer models:

FACT 46 (Semantic Entailment). *A* entails *B* in a Routley–Meyer model (that is, for all x, if $x \Vdash A$ then $x \Vdash B$) if and only if $0 \Vdash A \to B$.

This means that these points are not only prime theories, they are prime 0-theories. They are closed under the "logic of 0." And here, what is going on at 0 may be more

---

[88]This is a plausible proposal, provided that you are aware that in models for R, $\sqsubseteq_a$ may fail to be reflexive, as is needed to form a counterexample to $A \vdash B \to B$.

[89]Defined at Definition 5 on page 312.

than what amounts to "logic" according to the logic in question. In particular, this is the case at R, at least if negation is around. (Bear with the fact that I haven't told you how to interpret negation yet.) For R proves $A \vee \sim A$, and so, by the primeness of the point 0, we will have either $0 \Vdash A$ or $0 \Vdash \sim A$. In any particular model, 0 will validate more than "logic alone". So, to construct a model, we will first construct a prime theory T for the base point 0, and then the other points in the model will be prime T-theories: theories closed under the inferences licensed by T.

DEFINITION 47 (The *Prime Theory* Canonical Model on T). Given a prime, regular theory T, *prime theory* canonical model on T is populated by prime T-theories. The identity point is T itself. The ternary relation R is defined as follows:

$$RUVW \text{ if and only if } \{B : (\exists A)(A \rightarrow B \in U \text{ and } A \in V)\} \subseteq W$$

and $U \Vdash A$ if and only if $A \in U$.

Note the similarity of the definition of R here to the definition of $\sqcup$ on the canonical theory model on page 364. Here, R is defined by composition of theories, but the composition of two prime T-theories may not itself be a prime theory, so we resort to the ternary relation.

Proving that this is indeed a *model* is a matter of checking all of the clauses. The difficult conditions are the *existential* ones, according to which there is a point in the model with certain properties. An example is one half of the implication clause: if $A \rightarrow B \notin U$, we need to find V, W where $RUVW$, $A \in V$ and $B \notin W$. This is a matter of using Belnap's Pair Extension Theorem[90] twice. First, to construct V you use the pair $\langle \{A\}, \{C : C \rightarrow B \in U\} \rangle$, and extend it to get a prime V. Then, for W you use the pair $\langle U \sqcup V, \{B\} \rangle$. The result will be the two prime theories you wish. The same techniques work for the other difficult clauses. The canonical prime theory model is indeed a model.

This shows the ubiquity of the pair extension theorem in the metatheory of distributive substructural logics. Prime theories play the part here of consistent and complete theories in the metatheory of classical intensional logics.

I have said nothing about the treatment of negation. The Routleys' innovation is to understand that negation can be modelled by another operator which takes us away from the current point of evaluation. The Boolean clause will not do. The alternative is this:

○ $x \Vdash \sim A$ if and only if $x^* \nVdash A$

where $*$ is a map of period two on the set of points in a model. That is, $x^{**} = x$. This indeed suffices to make $\sim$ a De Morgan negation on the model. Adding the following condition

○ (Contraposition) If $Rxyz$ then Rxz^*y^*

results in the model validating the contraposition axiom.

There has been a great deal of debate centred around the interpretation of the $*$ operator [Copeland, 1979; 1986; Meyer and Martin, 1986].[91] There is no doubt that the $*$

---

[90]Fact 10 on p. 313.

[91]Not to be confused with the $*$ of display logic, which simply means "not" in the metatheory of structures.

operation is not particularly perspicuous in and of itself. (Being told that it turns set-ups "inside out" is not particularly enlightening.) Instead of pursuing that debate here (which largely burned out), I will merely quote an insight from Belnap and Dunn:

> ... we are convinced of the high probability that a mathematical apparatus of such proven formal power will eventually find its concrete applications and its resting place in intuition (think of tensors) [1992, p. 164].

This, I think, has been borne out in the later development of the Routley–Meyer semantics, and its applications. But to find a plausible interpretation of *, and to understand the semantics more fully, we need to work with it some more. As it stands so far, the Routley–Meyer construction might seem *ad hoc* and fit simply for R and its neighbours. For although you can ring the changes with *some* of the rules (commutativity, associativity, contraction) others seem hopelessly fixed. The models, as they stand, do not appear *natural* in the way that Kripke models for modal logic do.

This is merely an appearance. Recent work (dating from the 1990s, and chiefly due to Dunn, on gaggle theory) has shown that ternary frames are completely natural models for substructural logics in just the same way as Kripke models interpret normal modal logics [Dunn, 1991; 1993; 1994; 1995; Restall, 2000a].

DEFINITION 48 (Ternary Frames). A *ternary frame* is a set with a ternary relation R on that set. The connectives $\wedge$, $\vee$, $\top$, $\bot$, $\rightarrow$, $\circ$, $\leftarrow$ can be defined on a ternary frame as follows:

○ $x \Vdash A \wedge B$ if and only if $x \Vdash A$ and $x \Vdash B$

○ $x \Vdash A \vee B$ if and only if $x \Vdash A$ or $x \Vdash B$

○ $x \Vdash \bot$ never

○ $x \Vdash \top$ always

○ $x \Vdash A \rightarrow B$ if and only if for each y, z where $Rxyz$, if $y \Vdash A$ then $z \Vdash B$.

○ $x \Vdash A \leftarrow B$ if and only if for each y, z where $Ryxz$, if $y \Vdash A$ then $z \Vdash B$.

○ $x \Vdash A \circ B$ iff there are y, z where $Ryzx$, $y \Vdash A$ and $z \Vdash B$.

Many structural rules come with a corresponding conditions on R.[92]

There are no restrictions on R in such an interpretation. It models distributive lattice operations, together with the residuated triple $\langle \circ, \rightarrow, \leftarrow \rangle$. A soundness and completeness result, using standard techniques, works for this frame semantics.

Interpreting R is a tricky business, as we have seen. Probably no non-circular definition (one which doesn't appeal to conditionality) will be possible. However, some interesting *explications* of R have been tried in the applied semantics of relevant logics. One answer which has some cachet at present explains R in terms of *situation theory*. If the points in a model are *circumstances* or *situations* of some kind, then R indicates the degree to which situations can carry information about other situations. In particular, *Rabc* holds just when circumstance *a* acts as a *information channel* from *b* to *c*. There is a significant

---

[92]Some, however, require an inclusion relation $\sqsubseteq$, to be defined below.

growing recent literature on the connections between traditional situation theory and the semantics of relevant logics [Barwise and Seligman, 1993; Barwise *et al.*, 1996; Mares, 1997; Restall, 1994; 1995].

Some conditions (such as the condition for double negation elimination, which is too complex to discuss here [Restall, 2000b]) require talk of a relation of inclusion between points in models. This is not surprising, if in the intended interpretation, points are possibly *incomplete* (think of models for intuitionistic logic) then sometimes the relation of *extension* or *inclusion* might play a role.

DEFINITION 49 (Inclusion). $\sqsubseteq$ is an *inclusion* relation on a ternary frame if and only if it is reflexive, transitive and asymmetric, and in addition

  ∘ For all x, y, z if $Rxyz$ and $x' \sqsubseteq x$, $y' \sqsubseteq y$ and $z \sqsubseteq z'$ then $Rx'y'z'$.[93]

A model on a ternary frame with an inclusion relation must satisfy the hereditary condition on atomic formulas

  ∘ If $x \Vdash p$ and $x \sqsubseteq x'$ then $x' \Vdash p$

The clause linking $\sqsubseteq$ and R suffices to prove the hereditary lemma: complex formulas involving $\rightarrow, \circ, \leftarrow$ satisfy the hereditary lemma if their constituents do.

Inclusion, as a relation between points in a model, is simple to explain given an interpretation of these points. If points are situations, then $a \sqsubseteq b$ just when a is a "subsituation" of b. The situation of my bedroom is a subsituation of the situation of my house. An inconsistent circumstance described by the first chapter of some fiction may well be a substitution of an inconsistent circumstance described by the whole book. Given these explications of inclusion, the connection between it and R is plausible. As x shrinks to x', it connects *more* pairs of circumstances, as for a given antecedent circumstance there are *more* possible consequent circumstances. Given x, as y shrinks to y', again, there are more possible consequent circumstances, as y' gives us less information to constrain possible consequents. These explications are probably not *reductions* of the notion, but they go some way to explain their appeal and their use.

Another significant role in models is played by the distinguished point 0 in Routley–Meyer models. This point plays the part of modelling t,

DEFINITION 50 (Truth Set). T is a *truth set* in a ternary model with inclusion if and only if.

  ∘ $RTab$ if and only if $a \sqsubseteq b$.

where "$RTab$" stands for $(\exists x)(x \in T \wedge Rxab)$. A truth set is *reduced* if it has the form $\{x : 0 \sqsubseteq x\}$ for some point 0.

A truth set does the job of recording frame consequence. The $\rightarrow$ formulas true at every point in T are exactly the entailments witnessed by the entire model. For some applications (in particular, using frames to prove the admissibility of disjunctive syllogism [Routley and Meyer, 1972a]) reduced truth sets are desirable. A great deal of work

---

[93]This is quite a defensible condition as it stands, but it's more general than it needs to be to prove the hereditary lemma for all formulas [Restall, 2000a, ch. 11].

has gone into showing the circumstances in which a logic has a semantics with a reduced truth set [Slaney, 1987; Priest and Sylvan, 1992; Restall, 1993; 1995]. On the other hand, in our intended interpretation, it is by no means obvious.

Most contentious is the interpretation of negation. Some of Dunn's recent work, however, has served to take the sting out of * [Dunn, 1994; Restall, 1999]. Dunn notes that * is a particular case of a more understandable clause for negation:[94]

DEFINITION 51 (Compatibility). A compatibility relation C on a frame is an arbitrary two-place relation. Negation is interpreted using C as follows:

○ $x \Vdash \sim A$ if and only if for all y, if xCy then $y \nVdash A$.

If the frame uses an inclusion relation, compatibility is related to inclusion as follows:

○ If xCy, $x' \sqsubseteq x$ and $y' \sqsubseteq y$ then $x'Cy'$.

If you think of x and y as compatible just when there are no clashes between them, then these clauses are defensible. A circumstance warrants $\sim A$ just when there's no compatible circumstance in which A holds. So, $\sim A$'s holding in a circumstance just when A is *ruled out* by that circumstance: there is no compatible circumstance in which A. If circumstances are possibly incomplete (they might be compatible with more than just themselves) and if they are possibly inconsistent (not compatible with themselves: containing an internal contradiction) then we have counterexamples to the paradoxes of implication from before. We may have $x \Vdash A \wedge \sim A$ (when it is not the case that xCx) or we may have $x \nVdash A \vee \sim A$ (when $x \nVdash A$, but xCy where $y \Vdash A$).[95]

But what of the dreaded *? It can be seen to be a special case of C. The behaviour of C is wrapped up by * just if a^* is the unique $\sqsubseteq$-maximum of the set $\{x : aCx\}$ of all points compatible with a. If this set has a unique $\sqsubseteq$-maximum, a^*, then indeed $a \Vdash \sim A$ just when $a^* \nVdash A$ [Restall, 1999].

So, the ternary frame semantics can be "deconstructed" into individual components, each of which may be explained and applied in different circumstances. Here are some other examples of ternary frames which have been useful in the study of substructural logics.

EXAMPLE 52 (Two-Dimensional Frames). Given a set D, we can define a frame on the set $D \times D$ of pairs of D elements, by defining the ternary relation R on $D \times D$, setting $R\langle a, b \rangle \langle c, d \rangle \langle e, f \rangle$ if and only if $b = c$, $e = a$ and $b = f$. In other words, $\langle a, b \rangle$ composes with $\langle b, c \rangle$ to result in $\langle a, c \rangle$, and no other relations hold between pairs. So, the evaluation conditions on these two-dimensional frames reduce as follows:

○ $\langle a, b \rangle \Vdash A \circ B$ if and only if for some $c \in D$, $\langle a, c \rangle \Vdash A$ and $\langle c, b \rangle \Vdash B$.

○ $\langle a, b \rangle \Vdash A \rightarrow B$ if and only if for each $c \in D$ if $\langle b, c \rangle \Vdash A$ then $\langle a, c \rangle \Vdash B$.

---

[94]The expression in terms of *compatibility* is mine. Dunn uses a relation of incompatibility, expressed by the symbol "⊥", which is already overloaded here.

[95]If we impose no constraints on compatibility at all, negation still satisfies a range of logical properties: it is order inverting (if $A \vdash B$ then $\sim B \vdash \sim A$) and so it automatically satisfies two of the De Morgan inferences ($\sim(A \vee B) \vdash \sim A \wedge \sim B$ and $\sim A \vee \sim B \vdash \sim(A \wedge B)$). In addition, since it is a *universal* operator, we have one more De Morgan inference ($\sim A \wedge \sim B \vdash \sim(A \vee B)$), and its degenerate case $\top \vdash \sim\bot$.

∘ $\langle a, b \rangle \Vdash B \leftarrow A$ if and only if for each $c \in D$ if $\langle c, a \rangle \Vdash A$ then $\langle c, b \rangle \Vdash B$.

In this frame, the ternary R reduces to a *partial* function on pairs. This function is associative but not symmetric, where defined. The point set is flat — there is no natural notion of inclusion to be imposed. This frame has a truth set, but in this case it is not reduced: it is the set $\{\langle a, a \rangle : a \in D\}$.

These models are studied by van Benthem, Došen and Orłowska [van Benthem, 1991; Došen, 1992a; Orłowska, 1988] in the context of substructural logics, and they have blossomed into their own industry, under the suggestive name 'arrow logics' [Marx and Mausch, 1996]. In these logics, we think of the points $\langle a, b \rangle$ as transitions, or arrows, from a to b.

These are important frames for they are closely related to language frames in a number of respects — the relation R is functional: if $Rxyz$ and $Rxyz'$ then $z = z'$. However, in this case the relation is *partial*. For some x, y there is no z such that $Rxyz$.

EXAMPLE 53 (Mitchell's *IE models*). Mitchell's IE models, are models of linear logic with distribution of $\wedge$ over $\vee$ [1997]. In these models, points are *pairs* $\langle m, n \rangle$ whose elements are taken from a commutative monoid R of *resources*. As R is a commutative monoid, there is an operation $+$ on R, such that $m + n = n + m$, with an identity 0, such that $m + 0 = m = 0 + m$. We evaluate propositions at points as follows: $\langle m, n \rangle \Vdash A \circ B$ if and only if for some n_1, n_2 where $n = n_1 + n_2$, $\langle m + n_1, n_2 \rangle \Vdash A$ and $\langle m + n_2, n_1 \rangle \Vdash B$. $\langle m, n \rangle \Vdash {\sim}A$ if and only if $\langle n, m \rangle \nVdash A$. $\langle m, n \rangle \Vdash A \rightarrow B$ if and only if for each m_1, m_2 where $m = m_1 + m_2$, if $\langle n + m_2, m_1 \rangle \Vdash A$ then $\langle m_2, n + m_1 \rangle \Vdash B$. Conjunction, disjunction, $\top$ and $\bot$ are defined in the usual way.

Early antecedents of the frame semantics for substructural logics can be found in Jónsson and Tarski's work on the representation of Boolean algebras with operators [1951; 1952]. This work presents what amounts to a soundness and completeness result for frames of substructural logics (with Boolean negation), though it takes a certain amount of hindsight to see it as such. The papers are written very much from the perspective of algebra and representation theory.

An extensive study of the properties of ternary frames is given in Routley, Meyer, Brady and Plumwood's *Relevant Logics and their Rivals* [1982]. Gabbay [1972] also gave a ternary relational semantics for implication, independently of the tradition of Routley and Meyer. Frames can be viewed from an algebraic point of view. The class of propositions of the frame is a completely distributive lattice under intersection and union, and it is equipped with the appropriate operators, defined by the clauses in the evaluation conditions. For example, the implication clause gives us

$$\alpha \rightarrow \beta = \big\{ x : (\forall y, z \in \mathcal{F})(Rxyz \rightarrow (y \in \alpha \rightarrow z \in \beta)) \big\}$$

Similarly, ${\sim}\alpha = \{x : (\forall y \in \mathcal{F})(xCy \rightarrow y \notin \alpha)\}$, and so on. We will call the resulting propositional structure 'Alg$(\mathcal{F})$' the *algebra* of the frame $\mathcal{F}$. Furthermore, any interpretation $\Vdash$ on a frame gives you an evaluation $v_\Vdash$ given by setting $v_\Vdash(A)$ to be $[\![A]\!]$. The connections with algebra run even deeper. Our canonical model is constructed out of

prime theories in a language. A similar construction can work with the *prime filters* of a propositional structure. Dunn's work on gaggles [1991; 1993; 1994; 1995] generalises the results here to operators with arbitrary arity. An *n*-ary operator is modelled with an $n + 1$-place relation.

Duality theory is the study of the relationship between algebras and their representations in terms of frames. There is an important strand of recent work in the semantics of substructural logic exploring duality theory in this context [Hartonas, 1996; 1997a; Hartonas and Dunn, 1997; Sambin and Vaccaro, 1988; Urquhart, 1996].

Meyer and Mares have done important work on the particular case of adding an S4-type necessity for R [1993], and they have shown that disjunctive syllogism is admissible in this case, using the frame semantics to prove it. Meyer and Mares have also studied the extensions of these logics with Boolean negation [1992; 1993].

Study of the frame conditions corresponding to rules brings forward questions of canonicity and correspondence. When is the canonical frame for a logic itself a model of the logic? This is not always the case in modal logics, and also, not always the case in substructural logics. There has been some work in attempting to pin down the class of substructural logics for which canonicity holds [Ghilardi and Meloni, 1997; Kurtonina, 1995; Restall, 2000a].

Not all logics have connectives which are amenable to the treatment of accessibility relations. We will see this when we consider ! from linear logic. Another case is the counterfactual conditional. These are more aptly modelled by *neighbourhood frames*. There has been, as yet, only a little work considering how neighbourhood frames can be used in a substructural setting [Akama, 1997; Fuhrmann and Edwin, 1994; Mares and Fuhrmann, 1995].

Projective Frames and Undecidability

R is undecidable. Alasdair Urquhart proved this in his ground-breaking papers [1983; 1984]. The general idea is a straightforward one: encode a known undecidable problem into the language of R. Meyer showed how to do this in the 1960s, by constructing a simple substructural logic, such that deciding what was a theorem in that logic would enable you to solve the word problem for free semigroups [Meyer, 1968; Meyer and Routley, 1973c]. That logic was not particularly natural. (It was the Lambek calculus together with just enough contraction to enable you to represent the deducibility problem as a conditional.) The logic was not particularly like R. The insights that helped decide the issue for R came from an unexpected quarter — projective geometry. To see why projective geometry gave the necessary insights, we will first consider a simple case, the undecidability of the system KR. KR is given by adding $A \wedge \sim A \rightarrow B$ to R. A KR frame is one satisfying the following conditions (given by adding the clause that $a = a^*$ to the conditions for an R frame).[96]

---

[96]My presentation of these results is indebted to many discussions with Pragati Jain [1997].

> R0*ab* iff $a = b$ *Rabc* iff *Rbac* iff *Racb* (total permutation)
> *Raaa* for each a R^2abcd only if R^2acbd

The clauses for the connectives are standard, with the proviso that $a \Vdash \sim A$ iff $a \nVdash A$, since $a = a^*$.

Urquhart's first important insight was that KR frames are surprisingly similar to projective spaces. A *projective space* $\mathcal{P}$ is a set P of points, and a collection L of subsets of P called *lines*, such that any two distinct points are on exactly one line, and any two distinct lines intersect in exactly one point. But we can define projective spaces instead through the ternary relation of collinearity. Given a projective space $\mathcal{P}$, its collinearity relation C is a ternary relation satisfying the condition:

> *Cabc* iff $a = b = c$, or a, b and c are distinct and they lie on a common line.

If $\mathcal{P}$ is a projective space, then its collinearity relation C satisfies the following conditions,

> *Caaa* for each a. *Cabc* iff *Cbac* iff *Cacb*. C^2abcd only if C^2acbd.

provided that every line has at least four points (this last requirement is necessary to verify the last condition). Conversely, if we have a set with a ternary relation C satisfying these conditions, then the space defined with the original set as points and the sets $l_{ab} = \{c : Cabc\} \cup \{a, b\}$ where $a \neq b$ as lines is a projective space.

Now the similarity with KR frames becomes obvious. If $\mathcal{P}$ is a projective space, the frame $\mathcal{F}(\mathcal{P})$ generated by P is given by adjoining a new point 0, adding the conditions *C0aa*, *Ca0a*, and *Caa0*, and by taking the extended relation C to be the accessibility relation of the frame.

Projective spaces have a naturally associated undecidable problem. The problem arises when considering the *linear subspaces* of projective spaces. A subspace of a projective space is a subset which is also a projective space under its inherited collinearity relation. Given any two linear subspaces $\mathcal{X}$ and $\mathcal{Y}$, the subspace $\mathcal{X} + \mathcal{Y}$ is the set of all points on lines through points in $\mathcal{X}$ and points in $\mathcal{Y}$.

In KR frames there are propositions which play the role of linear subspaces in projective spaces. We need a convention to deal with the extra point 0, and we simply decree that 0 should be *in* every "subspace." Then linear subspaces are equivalent to the *positive idempotents* in a frame. That is, they are the propositions X which are *positive* (so $0 \in X$) and *idempotent* (so $X = X \circ X$). Clearly, for any formula A and any KR model $\mathcal{M}$, the extension of A, $\|A\|$ in $\mathcal{M}$ is a positive idempotent iff $0 \Vdash A \wedge (A \leftrightarrow A \circ A)$. It is then not too difficult to show that if A and B are positive idempotents, so are $A \circ B$ and $A \wedge B$, and that t and $\top$ are positive idempotents.

Given a projective space $\mathcal{P}$, the lattice algebra $\langle \mathcal{L}, \cap, + \rangle$ of all linear subspaces of the projective space, under intersection and $+$ is a modular geometric lattice. That is, it is a complete lattice, satisfying these conditions:

Modularity $a \geqslant c \Rightarrow (\forall b)\big(a \cap (b + c) \leqslant (a \cap b) + c\big)$

Geometricity Every lattice element is a join of atoms, and if a is an atom and X is a set where $a \leqslant \Sigma X$ then there's some finite $Y \subseteq X$, where $a \leqslant \Sigma Y$.

The lattice of linear subspaces of a projective space satisfies these conditions, and in fact, any modular geometric lattice is isomorphic to the lattice of linear subspaces of some projective space. Furthermore the lattice of positive idempotents of any KR frame is also a modular geometric lattice.

The undecidable problem which Urquhart uses to prove the undecidability of KR is now simple to state. Hutchinson [1973] and Lipshitz [1974] proved this result:

FACT 54 (Modular Lattice Word Problem). The word problem for a class of modular lattices which includes the subspace lattice of an infinite dimensional projective space is undecidable.

That is, given any class of modular lattices, the *word* problem is the problem for deciding for any *problem* in the language of lattices (of the form "if $v_1 = w_1 \cdots v_n = w_n$ then $v_{n+1} = w_{n+1}$" where each v_i and w_i are terms in the language of lattices, on variables $x_1, \ldots, x_m$) whether or not it holds in this class of lattices.

Now, given an infinite dimensional projective space in which every line includes at least four points $\mathcal{P}$, the logic of the frame $(\mathcal{P})$ is said to be a *strong* logic. Our undecidability theorem then goes like this:

FACT 55 (Undecidability for KR). Any logic between KR and a strong logic is undecidable.

Proof. Consider a modular lattice problem

$$\text{If } v_1 = w_1 \cdots v_n = w_n \text{ then } v = w$$

stated in a language with variables x_i ($i = 1, 2, \ldots$) constants 1 and 0, and the lattice connectives $\cap$ and $+$. Fix a map into the language of KR by setting $x_i^r = p_i$ for variables, $0^r = t$, $1^r = \top$, $(v \cap w)^r = v^r \wedge w^r$ and $(v + w)^r = v^r \circ w^r$. The translation of our modular lattice problem is then the KR formula

$$\left(B \wedge (v_1^r \leftrightarrow w_1^r) \wedge \cdots \wedge (v_n^r \leftrightarrow w_n^r) \wedge t \right) \to (v^r \leftrightarrow w^r)$$

where the formula B is the conjunction of all formulas $p_i \wedge (p_i \leftrightarrow p_i \circ p_i)$ for each p_i appearing in the formulae v_j^r or w_j^r.

We will show that given a particular infinite dimensional projective space (with every line containing at least four points) $\mathcal{P}$, then the word problem is valid in the lattice of linear subspaces of $\mathcal{P}$ if and only if its translation is provable in L, for any logic L intermediate between KR and the logic of the frame $\mathcal{F}(\mathcal{P})$.

If the translation of the word problem is valid in L, then it holds in the frame $\mathcal{F}(\mathcal{P})$. Consider the word problem. If it were invalid, then there would be linear subspaces $x_1, x_2, \ldots$ in the space $\mathcal{P}$ such that each $v_i = w_i$ would be true while $v \neq w$. Construct a model on the frame $\mathcal{F}(\mathcal{P})$ as follows. Let the extension of p_i be the space x_i together with the point 0. It is then simple to show that $0 \Vdash B$, as each p_i is a positive idempotent. In addition, $0 \Vdash t$, and $0 \Vdash v_i^r \leftrightarrow w_i^r$, for the extension of each v_i^r and w_i^r will be the spaces picked out by v_i and w_i (both with the obligatory 0 added). However, we would have $0 \nVdash v^r \leftrightarrow w^r$, since the extensions of v^r and w^r were picked out to differ. This would

amount to a counterexample to the translation of the word problem, which we said was valid. As a result, the word problem is valid in the space $\mathcal{P}$. The converse reasoning is straightforward. Deciding the logic would give us a decision for the word problem.   ∎

Unfortunately, these techniques do not work for systems weaker than KR. The proof that positive idempotents are modular uses essentially the special properties of KR. Not every positive idempotent in R is modular. Nonetheless, the techniques of the proof can be extended to apply to a much wider range of systems. You do not need to restrict your attention to modular lattices to construct an undecidable word problem. But to do that, you need to examine Lipshitz and Hutchinson's proof more carefully. In the rest of this section, I will hint at the structure of Urquhart's undecidability proof for R and other logics. For detail, the reader is urged to consult Urquhart's original paper [1984] or my retelling of the proof [Restall, 2000a, ch. 15]

Lipshitz and Hutchinson proved that the word problem for modular lattices was undecidable by embedding into that problem the already known undecidable word problem for semigroups. It is enough to show that a structure can define a free associative binary operation, for then you will have the tools for representing arbitrary semigroup problems. Urquhart showed that this could be done without resorting to the full power of a modular lattice.

It suffices to have an 0-*structure*, and a *modular* 4-*frame* defined within that 0-structure. An 0-structure is a set equipped with the following structure

- It has a semilattice join operator $\sqcap$, defining an order $\leqslant$;
- It has a commutative and associative binary operator $+$;
- $x \leqslant y \ \Rightarrow \ x + z \leqslant y + z$;
- $0 + x = x$;
- $y \geqslant 0 \ \Rightarrow \ x \sqcap (x + y) = x$;

A 4-frame in a 0-structure is a set $\{a_1, a_2, a_3, a_4\} \cup \{c_{ij} : i \neq j, i, j = 1, \ldots, 4\}$ such that

- The a_is are *independent*. If $G, H \subseteq \{a_1, \ldots, a_4\}$ then $(\Sigma G) \sqcap (\Sigma H) = \Sigma(G \cap H)$ (where $\Sigma \varnothing = 0$)
- If $G \subseteq \{a_1, \ldots, a_4\}$ then ΣG is modular
- $a_i + a_i = a_i$
- $c_{ij} = c_{ji}$
- $a_i + a_j = a_i + c_{ik}$; $c_{ij} \sqcap a_j = 0$, if $i \neq j$
- $(a_i + a_k) \sqcap (c_{ij} + c_{jk}) = c_{ik}$ for distinct i, j, k

Given the 4-frame, we can define a semigroup structure. For each distinct i, j, we define the set L_{ij} to be $\{x : x + a_j = a_i + a_j \text{ and } x \sqcap a_j = 0\}$. Then if $b \in L_{ij}$ and $d \in L_{jk}$ where i, j, k are distinct, we set $b \otimes d = (b + d) \sqcap (a_i + a_k)$. It follows (through some manipulation) that $b \otimes d \in L_{ik}$. Then, we can define a semigroup operation '.' on L_{12} by:

$$x.y = (x \otimes c_{23}) \otimes (c_{31} \otimes y)$$

It is quite an involved operation to show that this is associative. Furthermore, in certain circumstances, the operation is *freely* associative. Given a countably infinite-dimensional vector space $\mathcal{V}$, its lattice of subspaces is a 0-structure, and it is possible to define a modular 4-frame in this lattice of subspaces, such that any countable semigroup is isomorphic to a subsemigroup of L_{12} under the defined associative operation.

The rest of the work of the undecidability proof involves showing that this construction can be modelled in a logic. Perhaps surprisingly, it can all be done in a weak logic like $\mathrm{TW}[\wedge, \vee, \rightarrow, \top, \bot]$. We can do without negation by defining it implicationally as usual: Pick a distinguished propositional atom f, and by defining $-A$ to be $A \rightarrow f$, t to be $-f$, and $A:B$ to be $-(A \rightarrow -B)$. A is a *regular* proposition iff $--A \leftrightarrow A$ is provable. The *regular* propositions form an 0-structure, under the assumption of the formula $\Theta = \{R(t, f, \top, \bot), N(t, f, \top, \bot), -\top \leftrightarrow \bot\}$. where $R(A)$ is $--A \leftrightarrow A$, $N(A)$ is $(t \rightarrow A) \rightarrow A$, and $R(A, B, \ldots)$ is $R(A) \wedge R(B) \wedge \cdots$ and similarly for N. So, we can show that the conditions for an 0-structure hold in the regular propositions, assuming Θ as a premise. To interpret the 0-structure conditions we model $\sqcap$ by $\wedge$, $+$ by $:$ and 0 by t. To model a 4-frame in the 0-structure, Define $K(A)$ to be $R(A) \wedge (A \wedge -A \leftrightarrow \bot) \wedge (A \vee -A \leftrightarrow \top) \wedge (A : -A \leftrightarrow -A) \wedge (A \leftrightarrow A:A)$. Then we can show the following

$$K(A), R(B, C), C \rightarrow A \;\Rightarrow\; A \wedge (B:C) \leftrightarrow (A \wedge B):C.$$

Then the conditions for a 4-frame go as follows: Choose distinct atomic formulas $A_1, \ldots, A_4$ and $C_{12}, \ldots, C_{34}$ to match $a_1, \ldots, a_4$ and $c_{12}, \ldots, c_{34}$. One independence axiom is then

$$(A_1 : A_2 : A_3) \wedge (A_2 : A_3 : A_4) \leftrightarrow (A_2 : A_3)$$

and one modularity condition is

$$K(A_1 : A_3 : A_4).$$

Let Π be the conjunction of the statements expressing that the propositions A_i and C_{ij} form a 4-frame in the 0-structure of regular propositions. In any algebra in which $\Theta \cup \Pi$ is true, the lattice of regular propositions is a 0-structure, and the denotations of the propositions A_i and C_{ij} form a 4-frame. Finally, when coding up a semigroup problem with variables $x_1, x_2, \ldots, x_m$, we will need formulas doing duty for these variables: We need a condition to pick out the fact that p_i (standing for x_i) is in L_{12}. We define $L(p)$ to be $(p : A_2 \leftrightarrow A_1 : A_2) \wedge (p \wedge A_2 \leftrightarrow t)$. Then the semigroup operation on elements of L_{12} can be defined in terms of $\wedge$ and $:$ and the formulas A_i and C_{ij}. We assume that done, and we will take it that there is an operation $\cdot$ on formulas which picks out the operation on L_{12}. Then we have the following:

FACT 56 (Deducibility from TW to KR is undecidable). For any logic between $\mathrm{TW}[\wedge, \vee, \rightarrow, \top, \bot]$ and KR, the Hilbert deducibility problem is undecidable.

Proof. Take a semigroup problem which is known to be undecidable. It may be presented in the following way

$$\text{If } v_1 = w_1 \ldots v_n = w_n \text{ then } v = w$$

where each term v_i, w_i is a term in the language of semigroups, constructed out of the variables $x_1, x_2, \ldots, x_m$ for some m. The translation of that problem into the language of $\mathrm{TW}[\wedge, \vee, \rightarrow, \top, \bot]$ is the deducibility problem

$$\Theta, \Pi, L(p_1, \ldots, p_m), v_1^r \leftrightarrow w_1^r, \ldots, v_n^r \leftrightarrow w_n^r \Rightarrow v^r \leftrightarrow w^r$$

where each translation u^r of each term u is defined recursively by setting x_i^r to be p_i, and $(u_1.u_2)^r$ to be $u_1^r \cdot u_2^r$.

For any logic between TW and KR the word problem in semigroups is valid if and only if its translation is valid in that logic. If the word problem is valid in the theory of semigroups, its translation must be valid, for given the truth of Θ and Π and $L(p_1, \ldots, p_m)$, the operator $\cdot$ is provably a semigroup operation on the propositions in L_{12} in the algebra of the logic, and the terms v_i and w_i satisfy the semigroup conditions. As a result, v^r and w^r pick out the same propositions, and we have a proof of $v^r \leftrightarrow w^r$.

Conversely, if the word problem is invalid, then it has an interpretation in the semigroup $\mathcal{S}$ defined on L_{12} in the lattice of subspaces of an infinite dimensional vector space. The lattice of subspaces of this vector space is the 0-structure in our countermodel. Consider the argument for KR. There, the subspaces were the positive idempotents in the frame. The other propositions in the frame were *arbitrary* subsets of points. Something similar can work here. On the vector space, consider the subsets of points which are closed under multiplication (that is, if $x \in \alpha$, so is kx, where k is taken from the field of the vector space). This is a De Morgan algebra, defining conjunction and disjunction by means of intersection and union as is usual. Negation is modelled by set difference. The fusion $\alpha \circ \beta$ of two sets of points is the set $\{x + y : x \in \alpha \text{ and } y \in \beta\}$. It is not too difficult to show that this is commutative and associative, and square increasing, when the vector space is in a field of characteristic other than 2, since if $x \in \alpha$ then $x = \frac{1}{2}x + \frac{1}{2}x \in \alpha \circ \alpha$. Then $\alpha \rightarrow \beta$ is simply $-(\alpha \circ -\beta)$. This is an algebraic model for KR, and the regular propositions in this model are exactly the subspaces of the vector space. It follows that our counterexample in the 0-structure is a counterexample in a model of KR to the translation of the word problem. As a result, the translation is not provable in KR or in any weaker logic.   ∎

This result applies to systems between TW and KR, and it shows that the deducibility problem is undecidable for any of these systems. In the presence of the *modus ponens* axiom $A \wedge (A \rightarrow B) \wedge t \rightarrow B$, this immediately yields the undecidability of the *theoremhood* problem, as the deducibility problem can be rewritten as a single formula.

$$\big(\Theta \wedge \Pi \wedge L(p_1, \ldots, p_m) \wedge (v_1^r \leftrightarrow w_1^r) \wedge \cdots \wedge (v_n^r \leftrightarrow w_n^r) \wedge t\big) \rightarrow (v^r \leftrightarrow w^r)$$

As a result, the theoremhood problem for logics between T and KR is undecidable. In particular, R, E and T are all undecidable.

The restriction to TW is necessary in the theorem. Without the prefixing and suffixing axioms, you cannot show that the lattice of regular propositions is closed under the 'fusion-like' connective ':'.

Before moving on to our next section, let us mention that these geometric methods have been useful not only in proving the undecidability of logics, but also in showing that interpolation fails in R and related logics [Urquhart, 1993].

Phase Spaces

Not all substructural logics are distributive, and not all point models validate distribution. In this section, we will look at phase spaces for linear logic as an example of a frame invalidating distribution. Before launching into the definition (due to Girard [1987a]) I will set the scene with some historical precedents.

An important idea germane to the representation of non-distributive lattices is that of a Dedekind–MacNeille closure [Davey and Priestley, 1990; Grätzer, 1978; Hartonas, 1997a; MacNeille, 1937; Troelstra, 1992].

EXAMPLE 57 (Dedekind–MacNeille Frames). Consider a poset with order $\sqsubseteq$. Define '$y \sqsubseteq \alpha$' for a set α to mean $y \sqsubseteq x$ for each $x \in \alpha$. Then the *closure* $\Gamma\alpha$ of a set α of points can be given as follows:

$$\Gamma\alpha = \{z : \forall y(y \sqsubseteq \alpha \rightarrow y \sqsubseteq z)\}.$$

Consider the closure operation on the class of all theories of some logic. If α is a set of theories, then suppose that $y \sqsubseteq \alpha$. This is equivalent to saying that $y \sqsubseteq \bigcap \alpha$: y is no bigger than the intersection of the set of the theories in α, which is itself a theory. So, if $y \sqsubseteq z$, then we must have $\bigcap \alpha \sqsubseteq z$ too. If $x \in \Gamma\alpha$, then anything true in all of α must also be true in x. So in these frames, to model disjunction we require $x \Vdash A \vee B$ if and only if $x \in \Gamma(\llbracket A \rrbracket \cup \llbracket B \rrbracket)$.

Sambin and others have used the notion of a "pretopology" (in our language, a set with a closure operator) not only as a model of substructural logics but also as a constructive generalisation of a topological space [Hartonas, 1997b; Sambin, 1989; 1993; 1995]. Došen [1988; 1989], Ono and Komori [1985], and Ono [1992] have also given semantics involving a closure operation

This is not the only way to avoid distribution. In a model without a notion of inclusion, we can get by with a *negation* to define a closure operator:

EXAMPLE 58 (Goldblatt Frames). Consider *orthologic*: an ortho-negation combined lattice logic. Here a frame will most likely appeal to a two-place compatibility relation C to deal with negation. The compatibility relation is reflexive (so $A \wedge \sim A \vdash \bot$) and symmetric (so $A \vdash \sim\sim A$). Robert Goldblatt showed (in [1974]) how to deal with disjunction by considering a simple compatibility frame $\langle P, C \rangle$, where P is a set of points (unordered by any inclusion relation) and C is a symmetric, reflexive compatibility relation on P.[97] Conjunction and negation are modelled in the standard way:

○ $x \Vdash A \wedge B$ iff $x \Vdash A$ and $x \Vdash B$

○ $x \Vdash \sim A$ iff for each y where xCy, $y \nVdash A$.

However, as it stands, this semantics does not validate $\sim\sim A \vdash A$. To add an extra condition on C to validate double negation elimination would result in C being the identity relation and the logic would collapse into classical propositional logic. Goldblatt's insight

---

[97] J. L. Bell presents an interesting philosophical analysis of Goldblatt's semantics, in which C is understood as *proximity* [Bell, 1986].

was to instead restrict the evaluation of propositions on the frame to those propositions for which $\sim\sim A \vdash A$ is valid. In the process, you reject distribution.

Given a set $\alpha \subseteq P$, let $\alpha^\sim = \{y : \forall x(x \in \alpha \rightarrow \sim yCx)\}$, or equivalently, $\{y : \forall x(yCx \rightarrow x \notin \alpha)\}$. Therefore, for any evaluation, $[\![A]\!]^\sim = [\![\sim A]\!]$. A set $\alpha \subseteq P$ is said to be *C-closed* if and only if $\alpha = \alpha^{\sim\sim}$. The *C*-closed sets will model our propositions. Since *C* is symmetric, $\alpha \subseteq \alpha^{\sim\sim}$.

A disjunction is true not only at the points at which either disjunct is true but also at the closure of that set of points. Here, however, it is *C*-closure at work.

○ $x \Vdash A \vee B$ iff $x \in ([\![A]\!] \cup [\![B]\!])^{\sim\sim}$

Girard's *phase spaces* (1987) are a generalisation of Goldblatt's compatibility frames (discovered independently of Goldblatt's work, despite being 10 years later).

EXAMPLE 59 (Phase Spaces). A *phase space* is a quadruple $\langle P, \cdot, 1, 0\rangle$ in which $\langle P, \cdot, 1\rangle$ is a commutative monoid, and in which 0 is a distinguished subset of *P*. The elements of *P* are *phases*, and 0 is the set of *orthogonal phases* of *P*.[98] In a phase space, the binary operator $\cdot$ is used for the ternary relation for implication. Here, $Rxyz$ if and only if $x \cdot y = z$. For any subset $G \subseteq P$, the *dual* $G^\sim$ of *G* is defined as follows:

$$G^\sim = \{x \in P : \text{ for all } y \in G(x \cdot y \in 0)\}.$$

In other words, $G^\sim$ is the set of all objects which send each element of *G* (by the monoid operation) to 0. For any set *G* of phases, $G^{\sim\sim}$ is the *closure* of *G*. It is not too hard to verify that this is indeed a closure operation, by showing the following:

○ $G \subseteq G^{\sim\sim}$.

○ $G^{\sim\sim\sim} = G^\sim$.

○ If $G \subseteq H$ then $H^\sim \subseteq G^\sim$.

○ $G = G^{\sim\sim}$ iff $G = H^\sim$ for some $H \subseteq P$.

The closed sets are called *facts*. The set of facts can be equipped with a natural monoid operation, $(G \cdot H)^{\sim\sim}$, where $G \cdot H$ is defined in the obvious way as $\{x \cdot y : x \in G \text{ and } y \in H\}$. This operation is residuated by the operation $\rightarrow$ defined by setting $G \rightarrow H = \{x : \forall y \in G(x \cdot y \in H)\}$, which can be shown to equal $(GH^\sim)^\sim$.

For negation, we define xCy to hold if and only if $x \cdot y \notin 0$. *C* is symmetric, given the commutativity composition, and the negation of a fact *G* is $G^\sim$. The negation of a fact is itself a fact.

It follows that this is a model for linear logic without exponentials. *R* satisfies the conditions for C and B, as composition is associative and commutative. The set $1 = \{1\}^{\sim\sim}$ is the identity (both left and right) for fusion.

Phase spaces are a particular kind of closure frame. They are special in a number of ways. Not only is the closure operation defined by negation, and not only are the structural rules B and C satisfied, but the accessibility relation underlying the frame is

---

[98]In the linear logic literature, '⊥' is used instead of '0' for the set of orthogonal phases. We use ⊥ for the bottom element of a lattice, so we will use 0 for the set of orthogonal phases.

functional. Nevertheless, phase spaces are still a faithful model for linear logic. We have the following theorem.

FACT 60 (Soundness and Completeness in Phase Spaces). $X \vdash A$ is provable in linear logic if and only if $X \vdash A$ holds in all phase spaces.

Proof. The soundness result is straightforward as usual. For completeness, we construct the canonical phase space out of formulas. The operator $\cdot$ on this frame is fusion. If you wish to think of a ternary relation, think of $RABC$ iff $A = B \circ C$. Then for 0, we have $\{A : 0 \vdash \sim A\}$. The *false* elements are the set of all formulae whose negations are provable, as you would expect. This is the correct choice, as $G^\sim = \{A : \forall B \in G(B \vdash \sim A)\}$, and so, $G^{\sim\sim} = \{A : \forall B \in G^\sim(B \vdash \sim A)\} = \{A : \forall B(\forall C \in G \text{ where } C \vdash \sim B)B \vdash \sim A\}$. Verifying the details is no more difficult in this case than in Urquhart's operational models for the conjunction/implication fragment of R. ∎

The definition of a phase space gives us a nice result. It motivates an embedding of the whole of multiplicative additive linear logic into its $(\rightarrow, \wedge, t)$ fragment. You choose f to be some arbitrary proposition, a translation as follows (where we set $\sim A = A \rightarrow f$).

$$
\begin{aligned}
p^t &= \sim\sim p \\
(A \wedge B)^t &= \sim\sim(A^t \wedge B^t) \\
(A \vee B)^t &= \sim(\sim A^t \wedge \sim B^t) \\
(A \circ B)^t &= \sim(A^t \rightarrow \sim B^t) \\
(A \rightarrow B)^t &= A^t \rightarrow B^t \\
t^t &= \sim\sim t
\end{aligned}
$$

FACT 61 (Embedding using $\wedge$, $\rightarrow$ and t). $A \vdash B$ holds in multiplicative, additive linear logic if and only if $A^t \vdash B^t$ in the $(\rightarrow, \wedge, t)$ fragment.

Proof. First, if $A^t \vdash B^t$ is provable then $A^t \vdash B^t$ is provable in linear logic, and in particular, it is provable when we choose $\sim t$ for f. In this case, A^t is equivalent to A in linear logic, and therefore, $A \vdash B$ is provable.

Conversely, if $A^t \vdash B^t$ does not hold in the $(\rightarrow, \wedge, t)$ fragment then in the canonical model (constructed simply out of *theories*) we have a counterexample to $A^t \vdash B^t$. Construct a phase space out of this model. The phases are the theories in the canonical model. The monoid operation is theory fusion, and the set 0 is $\{x : f \in x\}$. It is straightforward to check that any set of the form $[\![\sim\sim A]\!]$ in the original canonical model is a fact in the phase space we are constructing. Construct an interpretation of the language of linear logic by setting $[\![A]\!]$ in the phase space to be $[\![A^t]\!]$ in the canonical model. As the definition of the translation ^{t} mimics the evaluation clauses in a phase space, this is an acceptable phase space evaluation, and it is one which invalidates $A \vdash B$, so this consecution fails in linear logic. ∎

Note that this construction works in logics other than linear logic. For example, it will work to embed the whole of R without distribution into $R[\rightarrow, \wedge, t]$, for if the original model satisfies W, so will the phase model for R without distribution.

We will end this section by sketching how to cope with non-normal modal operators, such as ! and ? of linear logic. The difficulty with operators like these is the way the distribution properties of normal operators fail. We do not have $!A \wedge !B \vdash !(A \wedge B)$. So, we cannot use standard accessibility relations. However, something is possible.

DEFINITION 62 (Topolinear Spaces). A phase space with a set $\mathfrak{F}$ of facts satisfying the following conditions:

- If $X \subseteq \mathfrak{F}$ then $\bigcap X \in \mathfrak{F}$.
- If $F, G \in \mathfrak{F}$ then $F + G \in \mathfrak{F}$.
- If $F \in \mathfrak{F}$ then $F + F = F$.
- $\bigcap \mathfrak{F} = 0$.

is called a *topolinear space*. G is a *closed fact* iff $G \in \mathfrak{F}$, and G is a *open fact* iff $G^{\sim} \in \mathfrak{F}$.

Now, given any fact G, the *consideration* of G, $?G$, is

$$?G = \bigcap \{F : G \subseteq F \text{ and } F \in \mathfrak{F}\}$$

It is simply the smallest element of $\mathfrak{F}$ containing G. Its dual, the *affirmation* of G, $!G$ is

$$!G = (\bigsqcup \{H : H \subseteq G \text{ and } H^{\sim} \in \mathfrak{F}\})^{\sim \sim}$$

These are duals, as you can readily check.

LEMMA 63 (Duality of ! and ?). *For any fact G, $!(G^{\sim}) = (?G)^{\sim}$, and dually, $?(G^{\sim}) = (!G)^{\sim}$.*   □

This definition gives us a semantics for the exponentials. The semantics does as we would expect: by construction $G \subseteq ?G$, for any fact G, so by duality, $!G \subseteq G$. Furthermore, $?G$ is itself a closed fact, so $?G = ??G$, and dually, $!G = !!G$. Similarly, all of the closed facts are fixed points for fission, $?G + ?G = ?G$, and by duality, $!G \circ !G = !G$. Finally, $0 \subseteq ?G$ by construction, so by duality $!G \subseteq t$, and by the behaviour of t, $G \circ t = t$ gives $F \circ !G \subseteq F$.

Each of these simple verifications shows that the construction of ! and ? satisfies the rules for the exponentials in linear logic. This gives us the first part of the following fact.

FACT 64 (Soundness and Completeness in Topolinear Spaces). $X \vdash A$ is provable in LL if and only if $X \vdash A$ holds in all topolinear spaces.

Proof. As we have seen, the rules for the exponentials hold in topolinear spaces. For the converse, we must verify that the canonical topolinear space satisfies the conditions required for a topolinear space. So how should we construct the canonical topolinear space?

We will use the canonical phase space we have seen to construct a set of closed facts. Obviously, each $\{A : ?B \vdash A\}$ ought to be a closed fact for any choice of B. This cannot be the whole thing, as the intersection of a class of closed facts is not necessarily a set of the form $\{A : ?B \vdash A\}$. So we add these intersections. For any class of formulae B_i, we

will let $\bigcap_i \{A : ?B_i \vdash A\}$ be a closed fact. Once we do this, it is straightforward to check that $?[\![A]\!] = \bigcap \{F : [\![A]\!] \subseteq F$ and $F \in \mathfrak{F}\}$ for any formula A in the canonical model structure. The duality of ? and !, together with the duality of their defining conditions, ensures that the result for ! holds too. ∎

This kind of closure operation works well to model the exponentials in phase spaces.

4 LOOSE ENDS

Let me end this whirlwind tour through the history of substructural logic by indicating what I take to be some interesting directions for further research.

4.1 Paradox

Untutored intuitions about collections might lead you to believe that for any property, there is a collection of all and only those things which have that property. Formally, you might try this:

$$a \in \{x : \phi(x)\} \dashv\vdash \phi(a)$$

An object a is in the collection $\{x : \phi(x)\}$ of all of the ϕs if and only if a has property ϕ. This is the *naïve membership scheme*. Russell has shown that from naïve membership scheme, paradox follows. Consider the *Russell set* $\{x : x \notin x\}$. As an instance of the general scheme of membership, we have *Russell's paradox*:

$$\{x : x \notin x\} \in \{x : x \notin x\} \leftrightarrow \{x : x \notin x\} \notin \{x : x \notin x\}$$

The Russell set is a member of itself if and only if it is not a member of itself. In many traditional logics (classical or intuitionistic propositional logic, for example) from $p \leftrightarrow \sim p \vdash p \wedge \sim p$, and from this, anything at all.

The mainstream response to Russell's paradox is to calm our enthusiasm for the naïve membership scheme and to hunt around for weaker theories of set membership which are not so extravagant.[99]

However, this is not the only possibility. There is a motivation to consider logics in which we can retain the naïve membership scheme. Clearly, something must be done with the logic of negation, as we wish to retain propositions p such that $p \leftrightarrow \sim p$, without everything following from this. There are generally two options, logics with "gaps" or "gluts," corresponding to the point in the inference from $p \leftrightarrow \sim p$ to $p \wedge \sim p$ to $\bot$ which is taken to fail. A logic allows "gaps" if it the first inference fails, for p could then be "neither true nor false." A logic allows "gluts" if the second inference fails. Plenty of work has been done on both options for a number of years [Gilmore, 1974; 1986; Priest, 1987; Restall, 1992]

---

[99]There is some interesting work in this area, attempting to admit sets which are self-membered, without paradox [Aczel, 1988; Barwise and Moss, 1997].

However, it is not just the logic of negation which must be non-classical in order to retain the naïve membership scheme. Curry's paradox [Geach, 1955; Meyer and Routley, 1979]. Curry's paradox shows that more must be done, if the logic is to contain implication. Consider $\{x : x \in x \to F\}$, for some false proposition F.

$$\{x : x \in x \to F\} \in \{x : x \in x \to F\} \leftrightarrow (\{x : x \in x \to F\} \in \{x : x \in x \to F\} \to F)$$

This paradox reveals that there is a proposition p such that $p \leftrightarrow (p \to F)$, and as the following deduction shows, it is hard to avoid the inference to F:

$$
\cfrac{
  \cfrac{
    \cfrac{
      \cfrac{
        \cfrac{p \vdash p \to F \quad p \vdash p}{p; p \vdash F} (\to E)
      }{p \vdash F} [\text{WI}]
    }{0 \vdash p \to F} (\to I)
    \quad p \to F \vdash p
  }{\vdash p} [\text{Cut}]
  \quad
  \cfrac{
    \cfrac{
      \cfrac{p \vdash p \to F \quad p \vdash p}{p; p \vdash F} (\to E)
    }{p \vdash F} [\text{WI}]
  }{}
}{\vdash F} [\text{Cut.}]
$$

As the choice of F is arbitrary, we must attempt to stop this somewhere. A number of people have taken the step of *contraction* as the one to blame [Brady, 1983; Brady and Routley, 1989; Brady, 1989; Bunder, 1985; White, 1979]. However, contraction is a useful inferential move. It is *required* in mathematical induction. The step to, say, $F5$ from $F0 \wedge (\forall x)(Fx \to Fx + 1)$ requires the use of the premise no less than six times. Doing without contraction seems a little like cutting off one's nose to spite one's face. Can better be done here?

4.2 Relevant Predication

Dunn's *Relevant Predication* program is an interesting application of relevant logic to the clarification of philosophical issues [1987; 1990a; 1990b; 1996b; 1996a] and [Kremer, 1997]. A theory of relevant implication is used to attempt to mark out the distinction between *genuine properties* — say, my height, which is a genuine property of me — and *fake properties* — say, my height, as a fake property of you. I am indeed such that I am under 1.8 metres tall, and *you* are such that I am under 1.8 metres tall. But in the first case I have described how *I* am, and in the second, I haven't described any genuine property of *you*.

Classical logic is not good at marking out such a distinction, for if Hx stands for 'x is under 1.8 metres tall', and g stands for Greg, and h stands for you, then Hx is true of x iff it is under 1.8 metres tall, and $(Hg \wedge x = x) \vee Hg$ is true of something iff I am under 1.8 metres tall. Why is one a 'real' property and the other not? If we can reason using relevant implication, we can make the following distinction: It is true that if x is Greg then x is under 1.8 metres. However, it is *not* true that if x is you then Greg is under 1.8 metres. At least, it is plausible that this conditional fail, when read as a relevant conditional. This can be cashed out as follows.

DEFINITION 65 (Relevant Predication). F is a *relevant property of a* (written $(\rho x F x)a$) if and only if $(\forall x)(x = a \rightarrow Fx)$.

If F is a relevant property of a then Fa holds (quite clearly) and if F and G are relevant properties of a then so is their conjunction, and the disjunction of any relevant property with anything at all is still a relevant property.

Relevant logics excel at telling you what follows from what as a matter of *logic* — this gives us an interesting picture of the logical structure of relevant predication. However, that is only half the story. Applying the *semantics* of relevant logics ought to give us insight into what it is for a relevant implication to be *true*. That task is as yet, incomplete.

Monism and Pluralism

One debate in philosophical logic has been inspired by work in relevance and substructural logic, and we have already seen a hint of it in the discussion of disjunctive syllogism in Section 2.4. This is a debate between *pluralists* [Beall and Restall, 2000; 2001] and *monists* [Priest, 1999; Read, 1995] with respect to logical consequence. Is there *one* relation of deductive logical consequence (relative, say, to a particular choice of language, if this is a concern), or are there more than one? To make the discussion particular, given a particular instance of the inference of disjunctive syllogism

$$A \vee B, \sim A \vdash_? B$$

should the reasoner accept the inference as valid, reject it as invalid, or is there more to be said? In an interpretation which gives a counterexample to this inference, we may have a "point" x such that $x \Vdash A$, $x \Vdash \sim A$ and $x \not\Vdash B$. What are we to say about this?

Monists will say that if the choice of interpretations is correct, then this provides a counterexample to the inference. If the choice of interpretations is not a good one (if the interpretations are a model of a logic but not of the *One True Logic*) then the argument may well still be valid. For example, Priest [1999] argues that for an argument to be valid, it must be that in *every* circumstance in which the premises are true so is the conclusion, and the *One True Logic* is one which is sound and complete for the intended interpretation on the actual class of circumstances. Any logical consequence relation other than this either undergenerates by adding extra circumstances (which are alleged counterexamples to *really* valid arguments) or overgenerates by missing some out (which are counterexamples to invalid arguments missed out by the logic which is too strong).

Pluralists about logical consequence, on the other hand will say that a logic (and its attendant interpretations) may give us some information about the inference, but that this may not be the whole story about its validity or otherwise. For example, a pluralist may agree that there are indeed circumstances in which the premises of a disjunctive syllogism are true and the conclusion untrue. However, this choice of circumstances may include special circumstances not always considered: it includes *impossible* circumstances, as one would expect, if we are taking *relevance* seriously. It is natural too, to consider only possible circumstances, and if these are the only circumstances to consider, then

disjunctive syllogism ought be considered valid in this new, restricted sense. It is a lesson of relevant logic and its semantics that these are choices which can be made. For a monist, there is one definitive best answer to this choice. For a pluralist, both sides may have competing merits.

Pluralism extends beyond our interpretation of the semantics into our interpretation of proof theory too. Substructural logics have shown us that there is remarkable robustness in the interpretation of a conditional by means of the residuation clause:

$$\frac{X, A \vdash B}{X \vdash A \rightarrow B.}$$

However, the introduction and elimination rules for a conditional laid down by this clause, does *not* determine the meaning of the conditional.[100] These rules only pick out a fixed interpretation in combination with some account of the behaviour of the structural feature of the *comma*.

At the very least, relevant and substructural logics have provided so many new tools for understanding logical consequence that they have put the issue of pluralism on the agenda. Clarifying these options will deepen our understanding of logical consequence.

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

[Abramsky and Jung, 1994] Samson Abramsky and Achim Jung. Domain Theory. In Samson Abramsky, Dov M. Gabbay, and T. S. E. Maibaum (eds.), *Handbook of Logic in Computer Science*, volume 3, pages 1–168. Clarendon Press, Oxford, 1994.

[Ackermann, 1956] Wilhelm Ackermann. Begründung Einer Strengen Implikation. *Journal of Symbolic Logic*, 21:113–128, 1956.

[Aczel, 1988] Peter Aczel. *Non-Well-Founded Sets*. Number 14 in *CSLI Lecture Notes*. CSLI Publications, Stanford CA, 1988.

[Ajdukiewicz, 1935] K. Ajdukiewicz. Die Syntaktische Konnexität. *Studia Philosophica*, 1:1–27, 1935.

[Akama, 1997] Seiki Akama. Relevant Counterfactuals and Paraconsistency. In *Proceedings of the First World Conference on Paraconsistency*, Gent, Belgium, 1997.

[Andéka *et al.*, 1988] Hajnal Andéka, Steven Givant, and István Németi. Decision Problems for Equational Theories of Relation Algebras. *Memoirs of the American Mathematical Society*, 126(604), 1997.

---

[100]Despite the overwhelming literature on introduction and elimination rules determining the meanings of logical connectives [Brandom, 1994; Dummett, 1991; Harman, 1986].

[Anderson and Belnap, 1959] A. R. Anderson and N. D. Belnap Jr. Modalities in Ackermann's 'Rigorous Implication'. *Journal of Symbolic Logic*, 24:107–111, 1959.

[Anderson and Belnap, 1961] A. R. Anderson and N. D. Belnap Jr. Enthymemes. *Journal of Philosophy*, 58:713–723, 1961.

[Anderson and Belnap, 1962] A. R. Anderson and N. D. Belnap Jr. The Pure Calculus of Entailment. *Journal of Symbolic Logic*, 27:19–52, 1962.

[Anderson and Belnap, 1962] A. R. Anderson and N. D. Belnap Jr. Tautological Entailments. *Philosophical Studies*, 13:9–24, 1962.

[Anderson and Belnap, 1975] A. R. Anderson and Nuel D. Belnap. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton NJ, 1975.

[Anderson and Belnap, 1992] A. R. Anderson, Nuel D. Belnap, and J. Michael Dunn. *Entailment: The Logic of Relevance and Necessity*, volume 2. Princeton University Press, Princeton NJ, 1992.

[Avron, 1986] Arnon Avron. On Purely Relevant Logics. *Notre Dame Journal of Formal Logic*, 27:180–194, 1986.

[Avron, 1992] Arnon Avron. Whither Relevance Logic? *Journal of Philosophical Logic*, 21:243–281, 1992.

[Bar-Hillel, 1953] Y. Bar-Hillel. A Quasiarithmetical Notation for Syntactic Description. *Language*, 28:47–58, 1953.

[Barendregt *et al.*, 1983] H. P.Barendregt, M. Coppo, and M. Dezani-Ciancaglini. A Filter Lambda Model and the Completeness of Type Assignment. *Journal of Symbolic Logic*, 48:931–940, 1983.

[Barr and Wells, 1990] Michael Barr and Charles Wells. *Category Theory for Computing Science*. Prentice-Hall, 1990.

[Barwise *et al.*, 1996] J. Barwise, D. Gabbay, and C. Hartonas. Information Flow and the Lambek Calculus. In J. Seligman and D. Westerståhl, eds, *Logic, Language and Computation, Proceedings of Information-Oriented Approaches to Logic, Language and Computation*, 58:47–62. *CSLI Lecture Notes*, CSLI Publications, 1996.

[Barwise and Etchemendy, 2000] Jon Barwise and John Etchemendy. *Language, Proof and Logic*. Seven Bridges Press, 2000.

[Barwise and Moss, 1997] Jon Barwise and Larry Moss. *Vicious Circles*. CSLI Publications, 1997.

[Barwise and Seligman, 1993] Jon Barwise and Jerry Seligman. Imperfect Information Flow. *Proceedings of the 8th Annual IEEE Symposium on Logic in Computer Science*, pages 252–260, 1993.

[Beall and Restall, 2000] J. C. Beall and Greg Restall. Logical Pluralism. *Australasian Journal of Philosophy*, 78:475–493, 2000.

[Beall and Restall, 2001] J. C. Beall and Greg Restall. Defending Logical Pluralism. In Bryson Brown and John Woods (eds), *Logical Consequence: Rival Approaches. Proceedings of the 1999 Conference of the Society of Exact Philosophy*, pages 1–22. Hermes, Stanmore, 2001.

[Bell, 1986] J. L. Bell. A New Approach to Quantum Logic. *British Journal for the Philosophy of Science*, 37:83–99, 1986.

[Bell and Slomson, 1969] J. L. Bell and A. B. Slomson. *Models and Ultraproducts: An Introduction*. North-Holland, Amsterdam, 1969.

[Bellin, 1991] G. Bellin. Proof Nets for Multiplicative and Additive Linear Logic. Technical Report, Department of Computer Science, Edinburgh University, 1991. ecs-lfcs 91–161.

[Belnap, 1967] Nuel D. Belnap. Special Cases of the Decision Problem of Relevant Implication. *Journal of Symbolic Logic*, 32:431–432, 1967. Abstract

[Belnap, 1977a] Nuel D. Belnap. How a Computer Should Think. In G. Ryle (ed.), *Contemporary Aspects of Philosophy*, pages 30–56. Oriel Press, 1977.

[Belnap, 1977b] Nuel D. Belnap. A Useful Four-Valued Logic. In J. Michael Dunn and George Epstein (eds.), *Modern Uses of Multiple-Valued Logics*, pages 8–37. Reidel, Dordrecht, 1977.

[Belnap, 1982] Nuel D. Belnap. Display Logic. *Journal of Philosophical Logic*, 11:375–417, 1982.

[Belnap, 1990] Nuel D. Belnap. Linear Logic Displayed. *Notre Dame Journal of Formal Logic*, 31:15–25, 1990.

[Belnap, 1993] Nuel D. Belnap. Life in the Undistributed Middle. In Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*, pages 31–41. Oxford University Press, 1993.

[Belnap and Dunn, 1981] Nuel D. Belnap and J. Michael Dunn. Entailment and the Disjunctive Syllogism. In F. Fløistad and G. H. von Wright (eds.), *Philosophy of Language / Philosophical Logic*, pages 337–366. Martinus Nijhoff, The Hague, 1981. Reprinted as Section 80 in *Entailment* Volume 2, [Anderson and Belnap, 1992].

[Belnap *et al.*, 1980] N. D. Belnap Jr., A. Gupta, and J. Michael Dunn. A consecutive calculus for positive relevant implication with necessity. *Journal of Philosophical Logic*, 9:343–362, 1980.

[Belnap and Wallace, 1961] N. D. Belnap Jr. and J. R. Wallace. *A Decision Procedure for the System $E_{\bar{I}}$ of Entailment With Negation.* Technical Report 11, Contract No. SAR/609 (16), Office of Naval Research, New Haven, 1961. Also published as [Belnap and Wallace, 1965].

[Belnap and Wallace, 1965] N. D. Belnap Jr. and J. R. Wallace. A Decision Procedure for the System $E_{\bar{I}}$ of Entailment With Negation. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 11:261–277, 1965.

[van Benthem, 1991] Johan van Benthem. Language in Action. *Journal of Philosophical Logic*, 20:225–263, 1991.

[Benton et al., 1992] Nick Benton, G. M. Bierman, J. Martin E. Hyland, and Valeria de Paiva. Linear λ-Calculus and Categorical Models Revisited. In E. Börger (eds.), *Proceedings of the Sixth Workshop on Computer Science Logic*. Volume 702 of *Lecture Notes in Computer Science*, pages 61–84, Springer-Verlag, 1992.

[Benton et al., 1992] Nick Benton, G. M. Bierman, J. Martin E. Hyland, and Valeria de Paiva. *Term Assignment for Intuitionistic Linear Logic*. Technical Report 262, Computer Laboratory, University of Cambridge, August 1992.

[Benton et al., 1993] Nick Benton, G. M. Bierman, J. Martin E. Hyland and Valeria de Paiva. A Term Calculus for Intuitionistic Linear Logic. In M. Bezem and J. F. Groote (eds.), *Proceedings of the International Conference on Typed Lambda Calculi and Applications*. Volume 664 of *Lecture Notes in Computer Science*, pages 75–90. Springer-Verlag, 1993.

[Bimbó, 2001] Katalin Bimbó. Semantics for Structurally Free Logics $LC+$. *Logic Journal of the IGPL*, 9(4):557–571, 2001.

[Bimbó and Dunn, 1998] Katalin Bimbó and J. Michael Dunn. Two Extensions of the Structurally Free Logic LC. *Logic Journal of the IGPL*, 6(3):493–424, 1998.

[Blute et al., 1996] Richard Blute, J. R. B. Cockett, R. A. G. Seely, and T. H. Trimble. Natural Deduction and Coherence for Weakly Distributive Categories. *Journal of Pure and Applied Algebra*, 13(3): 229–296, 1996.

[Brady, 1983] Ross T. Brady. The Simple Consistency of a Set Theory Based on the Logic CSQ. *Notre Dame Journal of Formal Logic*, 24:431–449, 1983.

[Brady, 1988] Ross T. Brady. A Content Semantics for Quantified Relevant Logics I. *Studia Logica*, 47:111–127, 1988.

[Brady, 1989] Ross T. Brady. The Non-Triviality of Dialectical Set Theory. In Graham Priest, Richard Routley, and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pages 437–470. Philosophia Verlag, 1989.

[Brady, 1991] Ross T. Brady. Gentzenization and Decidability of Some Contraction-less Relevant Logics. *Journal of Philosophical Logic*, 20:97–117, 1991.

[Brady, 1996] Ross T. Brady. Relevant Implication and the Case for a Weaker Logic. *Journal of Philosophical Logic*, 25:151–183, 1996.

[Brady and Routley, 1989] Ross T. Brady and Richard Routley. The Non-Triviality of Extensional Dialectical Set Theory. In Graham Priest, Richard Routley and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pages 415–436. Philosophia Verlag, 1989.

[Brandom, 1994] Robert B. Brandom. *Making It Explicit*. Harvard University Press, 1994.

[Bunder, 1985] Martin Bunder. *BCK-Predicate Logic as a Foundation of Multiset Theory*. Technical Report, Mathematics Department, University of Wollongong, 1985.

[Charlewood, 1978] G. W. Charlewood. *Representations of Semilattice Relevance Logic*. PhD thesis, University of Toronto, 1978.

[Charlewood, 1981] G. W. Charlewood. An axiomatic version of positive semi-lattice relevance logic. *Journal of Symbolic Logic*, 4:233–239, 1981.

[Church, 1941] Alonzo Church. *The Calculi of Lambda-Conversion*. Number 6 in *Annals of Mathematical Studies*. Princeton University Press, Princeton NJ, 1941.

[Church, 1951a] Alonzo Church. The Weak Positive Implication Calculus. *Journal of Symbolic Logic*, 16:238, 1951. Abstract of The Weak Theory of Implication [Church, 1951b].

[Church, 1951b] Alonzo Church. The Weak Theory of Implication. In A. Menne, A. Wilhelmy and H. Angsil (eds.), *Kontroliertes Denken: Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, pages 22–37. Kommissions-Verlag Karl Alber, Munich, 1951. Abstracted in *The Weak Positive Implication Calculus* [Church, 1951a].

[Cockett and Seely, 1997] J. R. B. Cockett and R. A. G. Seely. Proof Theory for full intuitionistic linear logic, bilinear logic, and MIX categories. *Theory and Applications of categories*, 3(5):85–131, 1997.

[Copeland, 1979] B. J. Copeland. On When a Semantics is not a Semantics: some reasons for disliking the Routley-Meyer semantics for relevance logic. *Journal of Philosophical Logic*, 8:399–413, 1979.

[Copeland, 1986] B.J. Copeland. What is a Semantics for Classical Negation? *Mind*, 95:478–490, 1986.

[Coppo *et al.*, 1981] M. Coppo, M. Dezani-Ciancaglini and B. Venneri. Functional Characters of Solvable Terms. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 27:45–58, 1981.

[Curry, 1950] Haskell B. Curry. *A Theory of Formal Deducibility*. Volume 6 of *Notre Dame Mathematical Lectures*. Notre Dame University Press, 1950.

[Curry and Feys, 1958] Haskell B. Curry and R. Feys. *Combinatory Logic*, volume 1. North-Holland, Amsterdam, 1958.

[Curry *et al.*, 1972] Haskell B. Curry, J. Roger Hindley, and Jonathan P. Seldin. *Combinatory Logic*, volume 2. North-Holland, Amsterdam, 1972.

[Danos, 1990] Vincent Danos. *La Logique Linéaire Appliquée à l'étude de divers processus de normalisation (et principalement du λ-calcul)*. PhD thesis, Université de Paris VII, 1990.

[Danos and Regnier, 1989] Vincent Danos and Laurent Regnier. The Structure of Multiplicatives. *Archive of Mathematical Logic*, 28:181–203, 1989.

[Davey and Priestley, 1990] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.

[Došen, 1988] Kosta Došen. Sequent Systems and Groupoid Models, Part 1. *Studia Logica*, 47:353–386, 1988.

[Došen, 1989] Kosta Došen. Sequent Systems and Groupoid Models, Part 2. *Studia Logica*, 48:41–65, 1989.

[Došen, 1992a] Kosta Došen. A Brief Survey of Frames for the Lambek Calculus. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 38:179–187, 1992.

[Došen, 1992b] Kosta Došen. The First Axiomatisation of Relevant Logic. *Journal of Philosophical Logic*, 21:339–356, 1992.

[Došen, 1996] Kosta Došen. Deductive Completeness. *Bulletin of Symbolic Logic*, 2:243–283, 1996.

[Dragalin, 1987] Albert Grigorevich Dragalin. *Mathematical Intuitionism: Introduction to Proof Theory*. Volume 67 of *Translations of Mathematical Monographs*. American Mathematical Society, 1987.

[Dummett, 1991] Michael Dummett. *The Logical Basis of Metaphysics*. Harvard University Press, 1991.

[Dunn, 1966] J. Michael Dunn. *The Algebra of Intensional Logics*. PhD thesis, University of Pittsburgh, 1966.

[Dunn, 1970] J. Michael Dunn. Algebraic Completeness for *R*-mingle and its Extensions. *Journal of Symbolic Logic*, 35:1–13, 1970.

[Dunn, 1971] J. Michael Dunn. An Intuitive Semantics for First Degree Relevant Implications (abstract). *Journal of Symbolic Logic*, 36:363–363, 1971.

[Dunn, 1973] J. Michael Dunn. A 'Gentzen' System for Positive Relevant Implication. *Journal of Symbolic Logic*, 38:356–357, 1974. (Abstract).

[Dunn, 1976a] J. Michael Dunn. Intuitive Semantics for First-Degree Entailments and Coupled Trees. *Philosophical Studies*, 29:149–168, 1976.

[Dunn, 1976b] J. Michael Dunn. A Variation on the Binary Semantics for RM. *Relevance Logic Newsletter*, 1:56–67, 1976.

[Dunn, 1986] J. Michael Dunn. Relevance Logic and Entailment. In Dov M. Gabbay and Franz Günthner (eds.), *Handbook of Philosophical Logic*, volume 3, pages 117–229. Reidel, Dordrecht, 1986.

[Dunn, 1987] J. Michael Dunn. Relevant Predication 1: The Formal Theory. *Journal of Philosophical Logic*, 16:347–381, 1987.

[Dunn, 1990a] J. Michael Dunn. Relevant Predication 2: Intrinsic Properties and Internal Relations. *Philosophical Studies*, 60:177–206, 1990.

[Dunn, 1990b] J. Michael Dunn. Relevant Predication 3: Essential Properties. In J. Michael Dunn and Anil Gupta (eds.), *Truth or Consequences*, pages 77–95. Kluwer, 1990.

[Dunn, 1991] J. Michael Dunn. Gaggle Theory: An Abstraction of Galois Connections and Residuation with Applications to Negation and Various Logical Operations. In J. van Eijck (ed.), *Logics in AI, Proceedings European Workshop JELIA 1990*. Volume 478 of *Lecture Notes in Computer Science*. Springer-Verlag, 1991.

[Dunn, 1993] J. Michael Dunn. Partial-Gaggles Applied to Logics with Restricted Structural Rules. In Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*, pages 63–108. Oxford University Press, 1993.

[Dunn, 1994] J. Michael Dunn. Star and Perp: Two Treatments of Negation. In James E. Tomberlin (ed.), *Philosophical Perspectives*, volume 7, pages 331–357. Ridgeview Publishing Company, Atascadero, California CA, 1994.

[Dunn, 1995] J. Michael Dunn. Positive Modal Logic. *Studia Logica*, 55:301–317, 1995.

[Dunn, 1996a] J. Michael Dunn. Is Existence a (Relevant) Predicate? *Philosophical Topics*, 24:1–34, 1996.

[Dunn, 1996b] J. Michael Dunn. Relevant Predication: A Logical Framework for Natural Properties. In J. Earman and J. D. Norton (eds.), *The Cosmos of Science. Pittsburgh–Konstanz Series in the Philosophy and History of Science*. University of Pittsburgh Press and Universitaets Verlug Konstanz, 1996.

[Dunn and Epstein, 1977] J. Michael Dunn and G. Epstein. *Modern Uses of Multiple-Valued Logic*. Reidel, Dordrecht, 1977.

[Dunn and Restall, 2001] J. Michael Dunn and Greg Restall. Relevance Logic. In Dov M. Gabbay (ed.), *Handbook of Philosophical Logic*, volume 6, pages 1–136. Kluwer Academic Publishers, 2002. Second edition.

[Dunn and Meyer, 1997] J.Michael Dunn and Robert K. Meyer. Combinators and Structurally Free Logic. *Logic Journal of the IGPL*, 5:505–537, 1997.

[Ehrhard, 1993] Thomas Ehrhard. Hypercoherences: a strongly stable model of linear logic. *Mathematical Structures in Computer Science*, 3:365–385, 1993.

[Finberg *et al.*, 1996] David Finberg, Matteo Mainetti, and Gian-Carlo Rota. The Logic of Commuting Equivalence Relations. In P.Agliano, A.Ursini, (eds.), *Logic and Algebra (Pontignan, 1994)*, pages 69–96. Volume 180 of *Lecture Notes in Pure and Applied Mathematics*. Dekker, 1996.

[Fine, 1974] Kit Fine. Models for Entailment. *Journal of Philosophical Logic*, 3:347–372, 1974.

[Fine, 1976] Kit Fine. Completeness for the Semilattice Semantics with Disjunction and Conjunction. *Journal of Symbolic Logic*, 41:560, 1976.

[Fine, 1988] Kit Fine. Semantics for Quantified Relevance Logic. *Journal of Philosophical Logic*, 17:27–59, 1988.

[Fitch, 1952] F. B. Fitch. *Symbolic Logic*. Roland Press, New York, 1952.

[Fitting, 1989] Melvin C. Fitting. Bilattices and the Semantics of Logic Programming. *Journal of Logic Programming*, 11(2):91–116, 1989.

[Frege, 1993] Gottlob Frege. *Grundgesetze der Arithmetik, Begriffsschriftlich abgeleitet*. Verlag Hermann Pohle, Jena, 1893–1903. Parts translated in *Gottlob Frege: Logical Investigations* [Geach and Black, 1952].

[Friedman and Meyer, 1992] Harvey Friedman and Robert K. Meyer. Whither Relevant Arithmetic? *Journal of Symbolic Logic*, 57:824–831, 1992.

[Fuhrmann and Edwin, 1994] André Fuhrmann and Edwin D. Mares. On S. *Studia Logica*, 53:75–91, 1994.

[Gabbay, 1972] D. M. Gabbay. A General Theory of the Conditional in Terms of a Ternary Operator. *Theoria*, 38:39–50, 1972.

[Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Reidel, Dordrecht, 1976.

[Galmiche, 2000] D. Galmiche. Connection methods in Linear Logic and Proof Nets Construction. *Theoretical Computer Science*, 232:231–272, 2000.

[Geach, 1955] P. T. Geach. On Insolubilia. *Analysis*, 15:71–72, 1955.

[Geach, 1958] P. T. Geach. Entailment. *Aristotelian Society supplementary volume*, 32:157–172, 1958.

[Geach and Black, 1952] Peter Geach and Max Black. *Translations from the Philosophical Writings of Gottlob Frege*. Oxford University Press, 1952.

[Gentzen, 1934] Gerhard Gentzen. Untersuchungen über das logische Schliessen. *Math. Zeitschrift*, 39:176–210 and 405–431, 1934. Translated in [Gentzen, 1969].

[Gentzen, 1969] Gerhard Gentzen. M. E. Szabo (ed.). *The Collected Papers of Gerhard Gentzen*. North-Holland, Amsterdam, 1969.

[Ghilardi and Meloni, 1997] Silvio Ghilardi and Giancarlo Meloni. Constructive Canonicity in Non-Classical Logic. *Annals of Pure and Applied Logic*, 86:1–32, 1997.

[Giambrone, 1985] Steve Giambrone. TW_+ and RW_+ are Decidable. *Journal of Philosophical Logic*, 14:235–254, 1985.

[Gilmore, 1974] Paul C. Gilmore. The Consistency of partial Set Theory without Extensionality. In T. Jech (ed.), *Axiomatic Set Theory*. Volume 13 of *Proceedings of Symposia in Pure Mathematics*, pages 147–153, Providence RI, American Mathematical Society, 1974.

[Gilmore, 1986] Paul C. Gilmore. Natural Deduction Based Set Theories: A New Resolution of the Old Paradoxes. *Journal of Symbolic Logic*, 51:393–411, 1986.

[Girard, 1987a] Jean-Yves Girard. Linear Logic. *Theoretical Computer Science*, 50:1–101, 1987.

[Girard, 1987b] Jean-Yves Girard. *Proof Theory and Logical Complexity*. Bibliopolis, Naples, 1987.

[Girard, 1995] Jean-Yves Girard. Geometry of Interaction III: accomodating the additives. In J. Y. Girard, Y. Lafont and L. Regnier (eds.), *Advances in Linear Logic*, pages 1–42. Cambridge University Press, Cambridge, 1995.

[Girard, 1996] Jean-Yves Girard. Coherent Banach Spaces: A Continuous Denotational Semantics. *Electronic Notes in Theoretical Computer Science*, 3:1–24. 1996.

[Girard *et al.*, 1989] Jean-Yves Girard, Yves Lafont, and Paul Taylor. *Proofs and Types*. Volume 7 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1989.

[Goldblatt, 1974] Robert Goldblatt. Semantic Analysis of Orthologic. *Journal of Philosophical Logic*, 3:19–35, 1974. Reprinted as chapter 3 of [Goldblatt, 1993].

[Goldblatt, 1993] Robert Goldblatt. *Mathematics of Modality*. CSLI Publications, 1993.

[Goré, 1998] Rajeev Goré. Substructural Logics on Display. *Logic Journal of the IGPL*, 6(3):451–604, 1998.

[Grätzer, 1978] George Grätzer. *General Lattice Theory*. Academic Press, 1978.

[Gunter and Scott, 1990] C. A. Gunter and D. S. Scott. Semantic Domains. In J. van Leeuwen (ed.), *Handbook of Theoretical Computer Science*, pages 633–674. Elsevier Science Publishers B. V., 1990.

[Harman, 1986] Gilbert Harman. The meanings of logical constants. In Ernest Le Pore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, pages 125–134. Blackwell, Oxford, 1986.

[Harrop, 1956] R. Harrop. On Disjunctions and Existential Statements in Intuitionistic Systems of Logic. *Mathematische Annalen*, 132:347–361, 1956.

[Hartonas, 1996] C. Hartonas. Order-Duality, Negation and Lattice Representation. In H. Wansing (ed.), *Negation: Notion in Focus*, pages 27–37. De Gruyter Publication, New York–Berlin, 1996. Paper presented at the conference Analyomen 2. Leipzig, 1994. Workshop on Negation.

[Hartonas, 1997a] C. Hartonas. Duality for Lattice-Ordered Algebras and Normal Algebraizable Logics. *Studia Logica*, 58(3): 403–450, 1997.

[Hartonas, 1997b] C. Hartonas. Pretopology Semantics for Bimodal Intuitionistic Linear Logic. *Journal of the Interest Group in Pure and Applied Logics*, 5(1):65–78, 1997.

[Hartonas and Dunn, 1997] C. Hartonas and J. Michael Dunn. Stone Duality for Lattices. *Algebra Universalis*, 37:391–401, 1997.

[Hazen, 1997] Allen Hazen. Aspects of Russell's Logic in 1906. Letter, dated October 21, 1997.

[Heyting, 1956] Arend Heyting. *Intuitionism: An Introduction*. North-Holland, Amsterdam, 1956.

[Hindley, 1983] J. R. Hindley. The Simple Semantics for Coppo-Dezani-Sallé Types. In M. Dezani-Ciancaglini and H. Montanari (eds.), *International Symposium on Programming*, pages 212–226. Volume 137 of *Lecture Notes in Computer Science*. Springer-Verlag, 1983.

[Hodges, 1993] Wilfrid Hodges. *Model Theory*. Cambridge University Press, 1993.

[Howard, 1980] W. A. Howard. The Formulae-as-types Notion of Construction. In J. P. Seldin and J. R. Hindley (eds.), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 479–490. Academic Press, London, 1980.

[Hutchinson, 1973] G. Hutchinson. Recursively Unsolvable Word Problems of Modular Lattices and Diagram Chasing. *Journal of Algebra*, 385–399, 1973.

[Hyde and Priest, 2000] Dominic Hyde and Graham Priest (eds.) *Sociative Logics and their Applications: Essays by the Late Richard Sylvan*. Ashgate, 2000.

[Jain, 1997] Pragati Jain. *Undecidability of Relevant Logics*. Technical Report, tr-arp-06-97, Automated Reasoning Project, 1997.

[Jónsson and Tarski, 1951] Bjarni Jónsson and Alfred Tarski. Boolean Algebras with Operators: Part I. *American Journal of Mathematics*, 73:891–939, 1951.

[Jónsson and Tarski, 1952] Bjarni Jónsson and Alfred Tarski. Boolean Algebras with Operators: Part II. *American Journal of Mathematics*, 75:127–162, 1952.

[Kleene, 1962] S. C. Kleene. Disjunction and Existence under Implication in Elementary Intuitionistic Formalisms. *Journal of Symbolic Logic*, 27:11–18, 1962. This paper has an addendum. See [Kleene, 1963].

[Kleene, 1963] S. C. Kleene. An Addendum. *Journal of Symbolic Logic*, 28:154–156, 1963. Addendum to Disjunction and Existence under Implication in Elementary Intuitionistic Formalisms [Kleene, 1962].

[Kracht, 1996] Markus Kracht. Power and Weakness of the Modal Display Calculus. In Heinrich Wansing (ed.), *Proof Theory of Modal Logic*, pages 93–121. Kluwer Academic Publishers, Dordrecht, 1996.

[Kremer, 1997] Philip Kremer. Dunn's Relevant Predication, Real Properties and Identity. *Erkenntis*, 47:37–65, 1997.

[Kripke, 1959] Saul A. Kripke. The Problem of Entailment. *Journal of Symbolic Logic*, 24:324, 1959. Abstract.

[Kurtonina, 1995] Natasha Kurtonina. *Frames and Labels: A Modal Analysis of Categorial Inference*. PhD thesis, Institute for Logic, Language and Computation, University of Utrecht, 1995.

[Lambek, 1958] Joachim Lambek. The Mathematics of Sentence Structure. *American Mathematical Monthly*, 65:154–170, 1958.

[Lambek, 1961] Joachim Lambek. On the Calculus of Syntactic Types. In R. Jacobsen (ed.), *Structure of Language and its Mathematical Aspects*, pages 166–178. Number XII of *Proceedings of the Symposium on Applied Mathematics*. American Mathematical Society, 1961.

[Lambek, 1969] Joachim Lambek. Deductive Systems and Categories II. In Peter Hilton (ed.), *Category Theory, Homology Theory and their Applications II*, pages 76–122. Volume 86 of *Lecture Notes in Mathematics*. Springer-Verlag, 1969.

[Lambek and Scott, 1986] Joachim Lambek and Philip J. Scott. *Introduction to Higher Order Categorical Logic*. Cambridge University Press, 1986.

[Lemmon, 1965] E. J. Lemmon. *Beginning Logic*. Nelson, 1965.

[Lewis, 1986] David K. Lewis. *On the Plurality of Worlds*. Blackwell, Oxford, 1986.

[Lewy, 1958] Casimir Lewy. Entailment. *Aristotelian Society supplementary volume*, 32:123–142, 1958.

[Lincoln *et al.*, 1992] P. Lincoln, J. Mitchell, A. Scenrov, and N. Shankar. Decision Problems for Propositional Linear Logic. *Annals of Pure and Applied Logic*, 56:239–311, 1992.

[Lipshitz, 1974] L. Lipshitz. The Undecidability of the Word Problems for Projective Geometries and Modular Lattices. In *Transactions of the American Mathematical Society*, pages 171–180. American Mathematical Society, 1974.

[Mac Lane, 1971] Saunders Mac Lane. *Categories for the Working Mathematician*. Number 5 in *Graduate Texts in Mathematics*. Springer-Verlag, 1971.

[MacNeille, 1937] H. M. MacNeille. Partially Ordered Sets. *Transactions of the American Mathematical Society*, 42:416–460, 1937.

[Maraist *et al.*, 1995] John Maraist, Martin Odersky, David N. Turner and Philip Wadler. Call-by-name, call-by-value, call-by-need, and the linear lambda calculus. In *11th International Conference on the Mathematical Foundations of Programming Semantics*, New Orleans, Lousiana, March–April 1995. Originally produced as Technical Report FP-94-12, Department of Computing Science, University of Glasgow.

[Mares, 1997] Edwin Mares. Relevant Logic and the Theory of Information. *Synthese*, 109:345–360, 1997.

[Mares and Fuhrmann, 1995] Edwin D. Mares and André Fuhrmann. A Relevant Theory of Conditionals. *Journal of Philosophical Logic*, 24:645–665, 1995.

[Mares and Meyer, 1992] Edwin D. Mares and Robert K. Meyer. The Admissibility of γ in R4. *Notre Dame Journal of Formal Logic*, 33:197–206, 1992.

[Mares and Meyer, 1993] Edwin D. Mares and Robert K. Meyer. The Semantics of R4. *Journal of Philosophical Logic*, 22:95–110, 1993.

[Martin, 1984] E. P. Martin. Noncircular Logic. *Journal of Symbolic Logic*, 49:1427, 1984. Abstract.

[Martin and Meyer, 1982] E. P. Martin and R. K. Meyer. Solution to the P-W problem. *Journal of Symbolic Logic*, 47:869–886, 1982.

[Marx and Mausch, 1996] Maarten Marx, László Pólos and Michael Mausch (eds.), *Arrow Logic and Multi-modal Logic*. CSLI Publications, 1996.

[McRobbie and Belnap Jr., 1979] M. A. McRobbie and N. D. Belnap Jr. Relevant Analytic Tableaux. *Studia Logica*, 38:187–200, 1979.

[Meredith and Prior, 1963] C. A. Meredith and A. N. Prior. Notes on the Axiomatics of the Propositional Calculus. *Notre Dame Journal of Formal Logic*, 4:171–187, 1963.

[Meyer, 1976a] R. K. Meyer. Ackermann, Takeuti and Schnitt; γ for Higher-order Relevant Logics. *Bulletin of the Section of Logic*, 5:138–144, 1976.

[Meyer, 1976b] R. K. Meyer. A General Gentzen System for Implicational Calculi. *Relevance Logic Newsletter*, 1:189–201, 1976.

[Meyer, 1991] R. K. Meyer, M. Bunder and L. Powers. Implementing the 'Fool's Model' for Combinatory Logic. *Journal of Automated Reasoning*, 7:597–630, 1991.

[Meyer, 1966] Robert K. Meyer. *Topics in Modal and Many-valued Logic*. PhD thesis, University of Pittsburgh, 1966.

[Meyer, 1968] Robert K. Meyer. An undecidability result in the theory of relevant implication. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 14:255–262, 1968.

[Meyer, 1970a] Robert K. Meyer. E and S4. *Notre Dame Journal of Formal Logic*, 11:181–199, 1970.

[Meyer, 1970b] Robert K. Meyer. R_I—the Bounds of Finitude. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 16:385–387, 1970.

[Meyer, 1971] Robert K. Meyer. On Coherence in Modal Logics. *Logique et Analyse*, 14:658–668, 1971.

[Meyer, 1973a] Robert K. Meyer. Conserving Positive Logics. *Notre Dame Journal of Formal Logic*, 14:224–236, 1973.

[Meyer, 1973b] Robert K. Meyer. Intuitionism, Entailment, Negation. In H. Leblanc (ed.), *Truth, Syntax and Modality*, pages 168–198. North-Holland, Amsterdam, 1973.

[Meyer, 1976] Robert K. Meyer. Metacompleteness. *Notre Dame Journal of Formal Logic*, 17:501–517, 1976.

[Meyer, 1978] Robert K. Meyer. *Why I am not a Relevantist*. Technical Report 1, Logic Group, RSSS, Australian National University, 1978.

[Meyer, 1998] Robert K. Meyer. ⊃E is Admissible in 'True' Relevant Arithmetic. *Journal of Philosophical Logic*, 27:327–351, 1998.

[Meyer and Dunn, 1969] Robert K. Meyer and J. Michael Dunn. *E*, *R* and γ. *Journal of Symbolic Logic*, 34:460–474, 1969.

[Meyer and Mares, 1993] Robert K. Meyer and Edwin Mares. Semantics of Entailment 0. In Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*. Oxford University Press, 1993.

[Meyer and Martin, 1986] Robert K. Meyer and Errol P. Martin. Logic on the Australian Plan. *Journal of Philosophical Logic*, 15:305–332, 1986.

[Meyer and Parks, 1972] Robert K. Meyer and Zane Parks. Independent Axioms of Sobociński's three-valued logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 18:291–295, 1972.

[Meyer and Restall, 1996] Robert K. Meyer and Greg Restall. 'Strenge' Arithmetic. *Automated Reasoning Project*, Australian National University, 1996. Submitted for publication.

[Meyer and Routley, 1972] Robert K. Meyer and Richard Routley. Algebraic Analysis of Entailment. *Logique et Analyse*, 15:407–428, 1972.

[Meyer and Routley, 1973a] Robert K. Meyer and Richard Routley. Classical Relevant Logics I. *Studia Logica*, 32:51–66, 1973.

[Meyer and Routley, 1973b] Robert K. Meyer and Richard Routley. Classical Relevant Logics II. *Studia Logica*, 33:183–194, 1973.

[Meyer and Routley, 1973c] Robert K. Meyer and Richard Routley. An Undecidable Relevant Logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 19:289–397, 1973.

[Meyer and Routley, 1979] Robert K. Meyer, Richard Routley, and J. Michael Dunn. Curry's Paradox. *Analysis*, 39:124–128, 1979.

[Meyer and Slaney, 1989] R. K. Meyer and J. K. Slaney. Abelian Logic from A to Z. In Graham Priest, Richard Sylvan and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pages 245–288, Philosophia Verlag, 1989.

[Meyer *et al.*, 1995] Robert K. Meyer, Michael A. McRobbie, and Nuel D. Belnap. Linear Analytic Tableaux. In *Proceedings of the Fourth Workshop on Theorem Proving with Analytic Tableaux and Related Methods*, pages 278–293. Volume 918 of *Lecture Notes in Computer Science*. 1995.

[Minc, 1972] G. Minc. Cut-Elimination Theorem in Relevant Logics. In J. V. Matijasevic and O. A. Silenko (eds.), *Isslédovaniá po konstructivnoj mathematiké i matematičeskoj logike V*, pages 90–97. Izdatél'stvo Nauka, 1972. English translation in [Minc, 1976].

[Minc, 1976] G. Minc. Cut-Elimination Theorem in Relevant Logics. *The Journal of Soviet Mathematics*, 6:422–428, 1976. English translation of the original article [Minc, 1972].

[Minc, 1977] G. Minc. Closed Categories and the Theory of Proofs. *Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematischeskogo Instituta im, V. A. Steklova AN SSSR*, 68:83–114, 1977. In Russian with an English summary.

[Mitchell, 1996] John C. Mitchell. *Foundations for Programming Languages*. MIT Press, 1996.

[Mitchell, 1997] William P. R. Mitchell. The Carcinogenic Example. *Logic Journal of the IGPL*, 5(6):795–810, 1997.

[Moh, 1950] Shaw-Kwei Moh. The Deduction Theorems and Two New Logical Systems. *Methodos*, 2:56–75, 1950.

[Moortgat, 1988] Michiel Moortgat. *Categorial Investigations: Logical Aspects of the Lambek Calculus*. Foris, Dordrecht, 1988.

[De Morgan, 1964] A. de Morgan. On the Syllogism: IV, and on the Logic of Relations. *Transactions of the Cambridge Philosophical Society*, 10:331–358, 1864. Read before the Cambridge Philosophical Society on April 23, 1860.

[Morrill, 1994] Glyn Morrill. *Type Logical Grammar: Categorial Logic of Signs*. Kluwer, Dordrecht, 1994.

[Mulvey, 1986] C. J. Mulvey. *Second Topology Conference*, Rendiconti del Circolo Matematico di Palermo, series 2, supplement number 12, pages 99–104. 1986.

[O'Hearn and Pym, 1999] P. O'Hearn and D. Pym. The Logic of Bunched Implications. *Bulletin of Symbolic Logic*, 5:215–244, 1999.

[Ono and Komori, 1985] H. Ono and Y. Komori. Logic without the Contraction Rule. *Journal of Symbolic Logic*, 50:169–201, 1985.

[Ono, 1992] Hiroakira Ono. Semantics for Substructural Logics. In Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*, pages 259–291. Oxford University Press, 1993.

[Orlov, 1928] I. E. Orlov. The Calculus of Compatibility of Propositions (in Russian). *Matematicheskiĭ Sbornik*, 35:263–286, 1928.

[Orłowska, 1988] E. Orłowska. Relational Interpretation of Modal Logics. *Bulletin of the Section of Logic*, 17:2–14, 1988.

[Peirce, 1970] Charles S. Peirce. Description of a Notation for the Logic of Relatives, Resulting from an Amplification of the Conceptions of Boole's Calculus of Logic. *Memoirs of the American Academy of Science*, 9:317–378, 1870.

[Pentus, 1995] Mati Pentus. Models for the Lambek Calculus. *Annals of Pure and Applied Logic*, 75(1–2):179–213, 1995.

[Pentus, 1998] Mati Pentus. Free monoid completeness of the Lambek calculus allowing empty premises. In J. M. Larrazabal, Lascar D., and Mints G. (eds.), *Logic Colloquium '96: proceedings of the colloquium held in San Sebastian, Spain, July 9–15, 1996*, pages 171–209. Volume 12 of *Lecture Notes in Logic*. Springer, 1998.

[Pottinger, 1979] G. Pottinger. On analysing relevance constructively. *Studia Logica*, 38:171–185, 1979.

[Pratt, 1990] V. R. Pratt. Dynamic Algebras as a Well-Behaved Fragment of Relation Algebras. In *Algebraic Logic and Universal Algebra in Computer Science*. Volume 425 in *Lecture Notes in Computer Science*. Springer-Verlag, 1990.

[Prawitz, 1965] Dag Prawitz. *Natural Deduction: A Proof Theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.

[Priest, 1980] Graham Priest. Sense, Entailment and *Modus Ponens*. *Journal of Philosophical Logic*, 9:415–435, 1980.

[Priest, 1987] Graham Priest. *In Contradiction: A Study of the Transconsistent*. Martinus Nijhoff, The Hague, 1987.

[Priest, 1999] Graham Priest. *Logic: One or Many?* Typescript, 1999.

[Priest, 2000] Graham Priest. Motivations for Paraconsistency: The Slippery Slope from Classical Logic to Dialetheism. In Graham Priest, Diderik Batens, Chris Mortensen and Jean-Paul van Bendegem (eds.), *Frontiers of Paraconsistency*, pages 223–232. Kluwer Academic Publishers, 2000.

[Priest and Sylvan, 1989] Graham Priest and Richard Sylvan. Reductio Ad Absurdum et Modus Tollendo Ponens. In Graham Priest, Richard Sylvan, and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pages 613–626. Philosophia Verlag, 1989.

[Priest and Sylvan, 1992] Graham Priest and Richard Sylvan. Simplified Semantics for Basic Relevant Logics. *Journal of Philosophical Logic*, 21:217–232, 1992.

[Priest *et al.*, 1989] Graham Priest, Richard Sylvan, and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Philosophia Verlag, 1989.

[Rasiowa, 1974] H. Rasiowa. *An Algebraic Approach to Non-classical Logics*. North-Holland, Amsterdam, 1974.

[Read, 1981] Stephen Read. What is Wrong with Disjunctive Syllogism? *Analysis*, 41:66–70, 1981.

[Read, 1988] Stephen Read. *Relevant Logic*. Basil Blackwell, Oxford, 1988.

[Read, 1995] Stephen Read. *Thinking about Logic*. Oxford University Press, 1995.

[Restall, 1992] Greg Restall. A Note on Naïve Set Theory in *LP*. *Notre Dame Journal of Formal Logic*, 33:422–432, 1992.

[Restall, 1993] Greg Restall. Simplified Semantics for Relevant Logics (and some of their rivals). *Journal of Philosophical Logic*, 22:481–511, 1993.

[Restall, 1994] Greg Restall. A Useful Substructural Logic. *Bulletin of the Interest Group in Pure and Applied Logic*, 2(2):135–146, 1994.

[Restall, 1995] Greg Restall. Display Logic and Gaggle Theory. *Reports in Mathematical Logic*, 29:133–146, 1995.

[Restall, 1995] Greg Restall. Four Valued Semantics for Relevant Logics (and some of their rivals). *Journal of Philosophical Logic*, 24:139–160, 1995.

[Restall, 1995] Greg Restall. Information Flow and Relevant Logics. In Jerry Seligman and Dag Westerståhl (eds.), *Logic, Language and Computation: The 1994 Moraga Proceedings*, pages 463–477. CSLI Publications, 1995.

[Restall, 1998] Greg Restall. Displaying and Deciding Substructural Logics 1: Logics with contraposition. *Journal of Philosophical Logic*, 27:179–216, 1998.

[Restall, 1999] Greg Restall. Negation in Relevant Logics: How I Stopped Worrying and Learned to Love the Routley Star. In Dov Gabbay and Heinrich Wansing (eds.), *What is Negation?*, pages 53–76. Volume 13 of *Applied Logic Series*. Kluwer Academic Publishers, 1999.

[Restall, 2000a] Greg Restall. *An Introduction to Substructural Logics*. Routledge, 2000.

[Restall, 2000b] Greg Restall. Defining Double Negation Elimination. *Logic Journal of the IGPL*, 8(6):853–860, 2000.

[Restall, 2000c] Greg Restall. Laws of Non-Contradiction, Laws of the Excluded Middle and Logics. In Graham Priest and JC Beall (eds.), *The Law of Non-Contradiction*. To appear.

[Roorda, 1991] D. Roorda. *Resource Logics: Proof-theoretical Investigations*. PhD thesis, Amsterdam, 1991.

[Routley, 1984] Richard Routley. Relevantism, Material Detachment, and the Disjunctive Syllogism Argument. *Canadian Journal of Philosophy*, 14:167–188, 1984.

[Routley and Meyer, 1972a] Richard Routley and Robert K. Meyer. Semantics of Entailment — II. *Journal of Philosophical Logic*, 1:53–73, 1972.

[Routley and Meyer, 1972b] Richard Routley and Robert K. Meyer. Semantics of Entailment — III. *Journal of Philosophical Logic*, 1:192–208, 1972.

[Routley and Meyer, 1973] Richard Routley and Robert K. Meyer. Semantics of Entailment. In Hugues Leblanc (ed.), *Truth Syntax and Modality, Proceedings of the Temple University Conference on Alternative Semantics*, pages 194–243. North-Holland, Amsterdam, 1973.

[Routley *et al.*, 1982] Richard Routley, Val Plumwood, Robert K. Meyer, and Ross T. Brady. *Relevant Logics and their Rivals*. Ridgeview, 1982.

[Russel, 1906] Bertrand Russell. The Theory of Implication. *American Journal of Mathematics*, 28:159–202, 1906.

[Sambin, 1989] Giovanni Sambin. Intuitionistic Formal Spaces and their Neighbourhood. In C. Bonotto, R. Ferro, S. Valentini, and A. Zanardo (eds.), *Logic Colloquium '88*, pages 261–286. North-Holland, Amsterdam, 1989.

[Sambin, 1993] Giovanni Sambin. The Semantics of Pretopologies. In Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*, pages 293–307. Oxford University Press, 1993.

[Sambin, 1995] Giovanni Sambin. Pretopologies and Completeness Proofs. *Journal of Symbolic Logic*, 60(3):861–878, 1995.

[Sambin and Vaccaro, 1988] Giovanni Sambin and V. Vaccaro. Topology and duality in modal logic. *Annals of Pure and Applied Logic*, 37:249–296, 1988.

[Schalk and de Paiva, 2004] Andrea Schalk and Valeria de Paiva. Poset-valued sets or how to build models for linear logics. *Theoretical Computer Science*, 315(1):83–107, 2004.

[Schellinx *et al.*, 1995] Harold Schellinx, Vincent Danos and Jean-Baptiste Joinet. On the Linear Decoration of Intuitionistic Derivations. *Archive of Mathematical Logic*, 33:387–412, 1995.

[Schröder, 1995] E. Schröder. *Vorlesungen über die Algebra der Logik (exacte Logik), Volume 3, Algebra und Logik der Relative, Part I.*. Leipzig, 1895.

[Schröder-Heister and Došen, 1993] Peter Schröder-Heister and Kosta Došen (eds.), *Substructural Logics*. Oxford University Press, 1993.

[Scott, 1973] Dana Scott. Models for Various Type-Free Calculi. In Patrick Suppes, Leon Henkin, Athanase Joja and Gr. C. Moisil (eds.), *Logic, Methodology and Philosophy of Science IV*, pages 157–187. North-Holland, Amsterdam, 1973.

[Scott, 1980] Dana Scott. Lambda Calculus: Some Models, Some Philosophy. In J. Barwise, H. J. Keisler and K. Kunen (eds.), *The Kleene Symposium*, pages 223–265. North-Holland, Amsterdam, 1980.

[Slaney, 1989] J. K. Slaney. RWX is not Curry-paraconsistent. In Graham Priest, Richard Sylvan and Jean Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pages 472–480. Philosophia Verlag, 1989.

[Slaney, 1987] John K. Slaney. Reduced Models for Relevant Logics without WI. *Notre Dame Journal of Formal Logic*, 28:395–407, 1987.

[Slaney, 1990] John K. Slaney. A General Logic. *Australasian Journal of Philosophy*, 68:74–88, 1990.

[Smiley, 1959] T. J. Smiley. Entailment and Deducibility. *Proceedings of the Aristotelian Society (new series)*, 59:233–254, 1959.

[Stalnaker, 1984] Robert Stalnaker. *Inquiry*. Bradford Books–MIT Press, 1984.

[Tarski, 1941] Alfred Tarski. On The Calculus of Relations. *Journal of Symbolic Logic*, 6:73–89, 1941.

[Tarski, 1956] Alfred Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Clarendon Press, Oxford, 1956. Translated by J. H. Woodger.

[Tennant, 1992] Neil Tennant. *Autologic*. Edinburgh University Press, 1992.

[Tennant, 1994] Neil Tennant. The Transmission of Truth and the Transitivity of Deduction. In Dov Gabbay (ed.), *What is a Logical System?*, pages 161–177. Volume 4 of *Studies in Logic and Computation*. Oxford University Press, Oxford, 1994.

[Thistlewaite *et al.*, 1988] Paul Thistlewaite, Michael McRobbie, and Robert K. Meyer. *Automated Theorem Proving in Non-Classical Logics*. Wiley, New York, 1988.

[Troelstra, 1992] A. S. Troelstra. *Lectures on Linear Logic*. CSLI Publications, 1992.

[Urquhart, 1972a] Alasdair Urquhart. The Completeness of Weak Implication. *Theoria*, 37:274–282, 1972.

[Urquhart, 1972b] Alasdair Urquhart. A General Theory of Implication. *Journal of Symbolic Logic*, 37:443, 1972.

[Urquhart, 1972c] Alasdair Urquhart. Semantics for Relevant Logics. *Journal of Symbolic Logic*, 37:159–169, 1972.

[Urquhart, 1972d] Alasdair Urquhart. *The Semantics of Entailment*. PhD thesis, University of Pittsburgh, 1972.

[Urquhart, 1983] Alasdair Urquhart. Relevant Implication and Projective Geometry. *Logique et Analyse*, 26:345–357, 1983.

[Urquhart, 1984] Alasdair Urquhart. The Undecidability of Entailment and Relevant Implication. *Journal of Symbolic Logic*, 49:1059–1073, 1984.

[Urquhart, 1986] Alasdair Urquhart. Many-Valued Logics. In Dov M. Gabbay and Franz Günthner (eds.), *Handbook of Philosophical Logic, Volume 3*, pages 71–116. Reidel, Dordrecht, 1986.

[Urquhart, 1990] Alasdair Urquhart. The Complexity of Decision Procedures in Relevance Logic. In J. Michael Dunn and Anil Gupta (eds.), *Truth or Consequences*, pages 77–95. Kluwer, 1990.

[Urquhart, 1993] Alasdair Urquhart. Failure of Interpolation in Relevant Logics. *Journal of Philosophical Logic*, 22:449–479, 1993.

[Urquhart, 1996] Alasdair Urquhart. Duality for Algebras of Relevant Logics. *Studia Logica*, 56:263–276, 1996.

[Urquhart, 1997] Alasdair Urquhart. *The Complexity of Decision Procedures in Relevance Logic II*. Available from the author, University of Toronto, 1997.

[Venneri, 1994] Betti Venneri. Intersection Types as Logical Formulae. *Journal of Logic and Computation*, 3:1–17, 1994.

[Vickers, 1990] Steven Vickers. *Topology via Logic*. Volume 5 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1990. Revised edition.

[Wright, 1957] G. H. von Wright. *Logical Studies*. Routledge & Kegan Paul, 1957.

[Wadler, 1990] Philip Wadler. Linear Types can Change the World!. In M. Broy and C. Jones (eds.), *Programming Concepts and Methods*, Sea of Galilee, Israel, April 1990. North-Holland, Amsterdam, 1990. IFIP TC 2 Working Conference.

[Wadler, 1991] Philip Wadler. Is there a Use for Linear Logic? In *ACM Conference on Partial Evaluation and Semantics-Based Program Manipulation*, New Haven, Connecticut, June 1991, pages 255–273. Volume 26, number 9. ACM Press, 1991.

[Wadler, 1992a] Philip Wadler. Comprehending monads. *Mathematical Structures in Computer Science*, 2:461–493, 1992. Special issue of selected papers from the 6th Conference on Lisp and Functional Programming.

[Wadler, 1992b] Philip Wadler. There's No Substitute for Linear Logic. In *Workshop on Mathematical Foundations of Programming Semantics*, Oxford, UK, April 1992.

[Wadler, 1993a] Philip Wadler. A Syntax for Linear Logic. In *9th International Conference on the Mathematical Foundations of Programming Semantics*, New Orleans, Lousiana, April, 1993, pages 513–529. Springer-Verlag, 1993.

[Wadler, 1993b] Philip Wadler. A Taste of Linear Logic. In *Mathematical Foundations of Computer Science*, pages 185–210. Volume 711 of *Lecture Notes in Computer Science*. Springer-Verlag, 1993.

[Wansing, 1993] Heinrich Wansing. *The Logic of Information Structures*. Volume 681 in *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1993.

[Wansing, 1994] Heinrich Wansing. Sequent Calculi for Normal Propositional Modal Logics. *Journal of Logic and Computation*, 4:125–142, 1994.

[Wansing, 1998] Heinrich Wansing. *Displaying Modal Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[White, 1979] Richard White. The Consistency of the Axiom of Comprehension in the Infinite Valued Predicate Logic of Łukasiewicz. *Journal of Philosophical Logic*, 8:503–534, 1979.

A.N. PRIOR'S LOGIC

Peter Øhrstrøm and Per Hasle

INTRODUCTION

The greatest achievement of A.N. Prior (1914–1969) was without doubt his development of modern temporal logic. From the mid-1950s and onwards, he almost singlehandedly laid out the foundation for this important discipline of modern logic. In the first decades after his death, Prior was remembered almost exclusively for this undeniably great achievement. His work was regarded as a milestone, but also as superseded by later developments, and his works were not much referred to in the 1970s and 1980s. However, since the early 1990s we have witnessed a strong resurgence of interest in Prior's work. Thus, for instance, he is the direct source of important recent developments such as 'hybrid logic'. Moreover, it has become clear that his work in logic had a scope much broader than "merely" temporal logic: Prior made important contributions to deontic logic, modal logic, the theory of quantification, theories of truth, and the history of logic. In his work he also discussed questions of ethics, free will, and theological problems. Clearly, any important researcher will to some degree be led from subject to subject, and will receive inspiration from the reaction of colleagues. Certainly this was the case with Prior too — as forcefully witnessed by the huge correspondence now kept in the Bodleian Library — but the breadth of Prior's interests was not just a case of "one thing leading to another", as observed already in Anthony Kenny's memorial paper on Prior: "[Prior] constantly returned to the same central and unchanging themes. Throughout his life, for instance, he worked away at the knot of problems surrounding determinism: first as a predestinarian theologian, then as a moral philosopher, finally as a metaphysician and logician" [Kenny, 1970, p. 348]. In this article we shall explore not only Prior's major contributions to logic, but also the central interests underlying this work in general. Even though the focus of interest will be the systematic character of Prior's logic, the central themes within his work can only be discussed in a satisfactory manner while also considering Prior's background and motivations. We hence begin with a sketch of the main stages in Prior's life and their relation to his work.

Arthur Norman Prior was born on December 4th, 1914 in Masterton in the North Island of New Zealand. His mother died a few weeks after his birth. His father was a doctor and a medical officer during the First World War, and Arthur was brought up by his aunts and grandparents. Both of his grandfathers were Methodist ministers. It is obvious that Prior's upbringing in a Christian family

formed an important background for his later works in philosophy and logic. In his works he often referred to theological problems and concepts, and he frequently quoted the church fathers and other religious writers.

Prior went to Otago University at Dunedin in 1932. He set out to study medicine, but after a short time he instead went into philosophy and psychology. In 1934 he attended Findlay's courses on ethics and logic. Through Findlay, Prior became interested in the history of logic and was introduced to Prantl's textbooks. His M.A. thesis was devoted to this subject.

In the introduction to *Logic and the Basis of Ethics*, Prior wrote about Findlay: "I owe to his teaching, directly or indirectly, all that I know of either Logic or Ethics" [1949, p. xi]. Of course, this statement would not have been true if uttered somewhat later in Prior's career, but it reflects Findlay's great importance for its early stages. It also reflects on Prior's great willingness to credit others, a characteristic which permeates his entire work.

Prior was brought up as a Methodist. However, during his first year as a Philosophy student at Otago University, 18–19 of age, he instead joined the Presbyterian denomination. The reason for this shift was dissatisfaction with the lack of systematicity in Methodist theology and especially its emphasis on the importance of having a personal conversion experience. Prior had not had, and never was to have, any such experience himself. During his B.A. studies in Philosophy, he attended courses at the Presbyterian Knox Hall with a view to entering the Presbyterian ministry. This intention was never realised, but he was for many years to come a practising member of the Presbyterian community. In particular, he became a very active member of the Student Christian Movement (SCM), and wrote a considerable number of papers for the movements magazine *The Student*. His latest contribution to this magazine was written as late as 1955 and it was entitled 'Speaking about God' [1955c].

As a young man Prior was also very much influenced by socialism as well as pacifism. Prior was to remain a socialist for his entire life, whereas he gave up pacifism (around 1942) as well as Presbyterianism (in the 1950s, see later discussion). To Prior, there was no contradiction between socialism and Christianity, a position not unknown yet unusual in his day. Major theological thinkers who influenced him included Karl Barth, Emil Brunner, and to some extent Søren Kierkegaard [1940a; 1940b; 1940c].

In the years about 1940 Prior, however, found himself in a crisis of belief. Around this time, and for a few years to follow, Prior had become interested in Freudian psychoanalysis. While he saw no difficulty in reconciling socialism and Christianity, he saw the latter as shattered by the insights gained from psychoanalytical thought. During these years he wrote the article *Can religion be discussed?* [1942], in which he advocated an almost atheistic position. This view, however, does not seem to have lasted very long. Prior was still an active Presbyterian, and later in the 1940s he again wrote papers in defence of predestinarian theology. In the long run, the decisive challenge to Prior's Christian beliefs proved to come, neither from Freudianism (in which he entirely lost interest), nor from socialism,

but from the very centre of Presbyterian theology, namely its teaching of predestination and its rejection of free will — an important theme, to which we shall return.

Although 'conservative' in his theological outlook as a Christian, he was never a 'fundamentalist'. This is made quite clear already in the unpublished *A Modernist Stocktaking* [Prior, Unpublished e], which warns against taking for granted the gains of 'Modernism', especially the right to free and critical inquiry. The paper deals with the position of Christianity in the face of Modernism. It rejects fundamentalism, but otherwise embraces Christianity — warning, however, against a 'bringing-up-to-date' of Christianity such as the one taking place in Nazi Germany at the time.

In 1943 Arthur Prior married Mary Wilkinson. From 1943 until the end of World War Two, he served in the Royal New Zealand Air Force. In view of Nazism and the World War, Prior had given up his pacifist leanings.

Prior's first employment at Canterbury University College was in 1946, where he continued his writings on philosophical and religious questions (see for instance [1946; 1947; 1948a; 1948b]. A vacancy had been made when Karl Popper left, and Prior — at least technically speaking — took over Popper's position. In 1949 Prior was appointed a Senior Lecturer. At this time, Prior was still strongly committed to theological studies, and he was working on a book on the history and thought of Scottish (Presbyterian) Theology. Unfortunately, the Priors' house burned down in March 1949. After the fire, in which some of his drafts perished, he gave up the project on Scottish Theology. His main intellectual interests from then on veered toward philosophy, ethics, and logic.

Prior became an elder of the Presbyterian Church in 1951. Clearly, by then he must have been revising his former attempts to defend the doctrine of predestination, but apparently this did not at the time shake his fundamental Christian belief.

In 1949, Prior's first book, *Logic and the Basis of Ethics*, was published by Oxford University Press. The book was very well received. It was also a turning point in the sense that after this publication, logical approaches to philosophical problems — as well as logic in its own right — came to dominate Prior's work. In 1953 he became a professor of philosophy.

During 1950 and 1951 Prior wrote a manuscript for a book with the working title *The Craft of Logic*. This book was, however, never published as a whole, but P.T. Geach and A.J.P. Kenny edited parts of it, which were published as "The Doctrine of Propositions and Terms" [1976a]. In the first chapter of the book, *Propositions and Sentences* (originally written ca. 1950) the author among other things analysed Aristotle's view on some of the problems concerning time and tense. Prior found that, according to the ancient as well as the medieval view, a proposition may be true at one time and false at another. He described this view in the following way:

> ... the statement or opinion that someone is sitting will be true so
> long as the person in question is in fact seated, and will become false

— if it is persisted in — as soon as he rises [Prior, 1976a, p. 38].

Even though Prior did not begin the development of temporal logic proper before 1953, the above remark makes it clear that already around 1950 he realised that there must be some relation between time and logic. In December 1951 Prior sent his manuscript to Clarendon Press. In the beginning of 1953 Clarendon accepted to publish *The Craft of Logic* on the condition that Prior made a number of rather substantial changes. As a result, Prior wrote a completely different book, *Formal Logic*, which was published in 1955 with a second edition in 1962.

In *Formal Logic*, Prior wanted to use neither Hilbert and Ackermann's notation (in which, for instance, the conjunction is represented as $p \& q$) nor the notation suggested by Russell and Whitehead (in which conjunction is represented as $p.q$). Instead he adopted Łukasiewicz's so-called Polish notation, in which conjunction is represented as Kpq. He used Polish notation in most of his writings throughout his life. He emphasised that this prefix notation "obviates the necessity of using brackets", so that "no special rules about bracketing and rebracketing need to be included among the rules for proving one formula from another" [1955b, p. 6]. Polish notation was rather common during Prior's lifetime. Apart from its theoretical appeal it also had the significant practical advantage that proofs etc. could be written directly on a typewriter. (In personal communication with the authors, Mary Prior has told us how Arthur Prior would time and again express his appreciation of this practical gain.) Nevertheless, there is no doubt that Prior also was quite convinced about the syntactical superiority of Polish notation, for which he campaigned throughout his career as a logician. But — as Dr. Mary Prior has put it — the battle between Polish and Russellian notation is over, and Russellian notation has clearly won. She has therefore approved that the notation be changed in new editions of her husband's works. This will no doubt make Prior's work accessible to a larger audience.

Prior not only preferred to use Polish notion for his works within symbolic logic. In fact he highly valued various parts of Polish logic, and he corresponded with several Polish logicians. In 1961 he even went to Poland to give a lecture (see [Prior, 1962b]) and to take part in the 1961 'International Colloquium on Methodology of Science', Warsaw. In particular, Prior found Łukasiewicz's three-valued logic very interesting [1920; 1930], and he carried out some careful studies of this logic (see [Prior, 1952]).

Prior had a strong belief in the value of formal logic. On the other hand, he also emphasised that logic has to do with real life. He wanted a logic that would take full advantage of formal methods, while also being sensitive to the reality of human experience. In an unpublished paper, he described this view:

> Perhaps you could call my logic a mixture of Frege and Kolakowski. — I want to join the formal rigorism of the one with the vitalism of the other. Perhaps you regard this as a bastard mixture — a mesalliance. — I think it is a higher synthesis. And I think it important that people who care for rigorism and formalism should not leave the basic flux and

flow of things in the hands of existentialists and Bergsonians and others who love darkness rather than light, but we should enter this realm of life and time, not to destroy it, but to master it with our techniques [Prior, Unpublished f].

A remarkable and pervasive feature of Prior's work is an unusually strong interest in — and huge knowledge of — the history of logic. Indeed, Prior took an interest in the history of logic not only as a subject in its own right, but also because he saw the works of ancient and medieval logicians as a significant contribution to the contemporary development of logic. From 1952 to 1955 he had seven articles published on the history of logic. Four of these were concerned with medieval logic and one with Diodorean logic. His interest in the history of logic is also evident in *Formal Logic*. Prior was particularly interested in Aristotle, Diodorus, and the Scholastics, but his interest also extended to more recent logicians such as Boole and Peirce, the latter of which he called "the greatest of all symbolic logicians" [1957a].

In 1954, Gilbert Ryle visited Christchurch. He brought with him an invitation to Prior to give the 'John Locke Lectures' in Oxford. (Cf. [Hasle, 2003, p. 299].) In 1956 the Priors went to Oxford for this purpose.

In Oxford, Prior made some important and lasting friendships and professional associations, especially with John Lemmon, Ivo Thomas, P.T. Geach, Elizabeth Anscombe, Carew Meredith, David Meredith, and C. Lejewski.

The John Locke lectures gave Prior an excellent opportunity to present his new findings regarding time and modality. The lectures were held on Mondays. Among the participants were John Lemmon, Ivo Thomas, and Peter Geach [Kenny, 1970, p. 337]. The lectures were later published as the book *Time and Modality* [1957b]. It was this work which made Prior internationally known. After the publication of *Time and Modality*, he received a number of important and interesting letters from various logicians. One of these logicians was Saul Kripke. In two letters to Prior in September and October 1958, Kripke put forth some very stimulating ideas regarding temporal logic. In particular, this correspondence led Prior to the development of the idea of branching time.

In December 1958, the Priors left New Zealand, Arthur Prior taking up a professorship at the University of Manchester. This transition brought with it another change, namely Prior's abandonment of religion. While in Oxford in 1956, he had continued to be an active member of the Presbyterian community, but during the last few years in New Zealand, he had become an ever more reluctant member. After coming back to England in late 1958, he refrained from joining the local Presbyterian community. Prior had become agnostic. There were probably several reasons for this development, but the main thing was without doubt (cf. [Hasle, 1999]) the tension between the idea of predestination and those ideas of free will which Prior constantly associated with the development of temporal logic. However, he continued to treasure his theological library and to study problems related to theology [Kenny, 1970, p. 326].

In the early winter of 1962 Prior was visiting professor at the University of Chicago. During this stay he made some thorough studies of parts of Charles Sanders Peirce's logic. He became aware of Peirce's attempt at developing a graphical logic, the so-called existential graphs (see [1967a]), but it seems that this graphical approach had almost no appeal to Prior's intuition. He seemingly wanted to maintain a traditional algebraic approach to logical formalism. Nevertheless, his admiration for Peirce was enhanced by his realisation that Peirce had had an idea of a tempo-modal logic, embodied in a rudimentary manner in the so-called Gamma graphs. In Chicago, Prior introduced Jay Zeman to the existential graphs of C.S. Peirce and suggested the topic to Zeman for his doctoral thesis.

From September 1965 to January 1966 Prior was a visiting Flint professor at the University of California. In addition to lecturing there, he read papers in various Californian universities including Berkeley. During his stay in California, Prior made some important professional associations, especially with Dana Scott, Davidson, David Lewis, and Richard Montague. In this period, the later book *Past, Present, and Future* — often regard as Prior's most important book — was drafted. Apparently Prior's California lectures contributed significantly to the flourishing development in logic there at that time, and especially it seems to have sparked off a great interest in tense logic in the USA.

The Priors stayed in Manchester for seven years. In 1966 Anthony Kenny recommended Prior for a fellowship at Balliol College. Prior was offered this position. He accepted and the family moved to Oxford, where Prior worked until his death in 1969.

In 1967 Prior published the aforementioned major work, *Past, Present and Future*, in which his approach to tense logic had reached a very convincing form. The decade of intense work in the field since the John Locke lectures had brought him a lot further. Also, he had been able to benefit greatly from the correspondence with logicians like Saul Kripke and Charles Hamblin.

As a teacher, Prior was very inspiring, and the style in his books and papers is often very entertaining. One example could be taken from the acknowledgements in *Logic and the Basis of Ethics*:

> It needs also to be said that the logic of ethics owes much to those who have put forward the fallacious arguments which it is its business to expose ... Of those who have performed this negative service for the logic of ethics, the two who seem to me to be most deserving of our gratitude are William Wollaston and Adam Smith [1949, p. xi].

It seems clear that he very much liked teaching and lecturing. Prior was not 'the Oxford type,' but it appears that he almost immediately built up a reputation as one of the best lecturers in Oxford.

Prior died on October 6th, 1969, whilst on a lecture tour in Scandinavia. On the day of his death he was visiting Trondheim in Norway. Prior had by then accomplished an impressive production. The bibliographical overview of Prior's philosophical works comprises about 200 titles [Prior, 2003, pp. 311–328]. In this

overview one can follow how Prior's interests developed during the course of his work.

Prior's work on philosophical logic includes an analytical and modern component as well as a historical component. Nevertheless, there is no sharp distinction between Prior's analytical and historical concerns on the one hand and his work as a formal logician on the other.

Summarising the main trends of Prior's works, it can be said that his work until the middle of the 1950s was characterised by a preoccupation with a logical approach to theology. After the war, the investigations into the logic of ethics caught his interest. He kept his interest of theology and ethics throughout his life, but, from the mid-1950s onwards, he mainly devoted himself to the study of the relation between time, modality, and logic. This should be seen as a natural consequence of his endeavour to develop a formal calculus of tense logic, a task which he took up in the early 1950s. He demonstrated that temporal logic can in fact be a very powerful tool in philosophical analysis — also in relation to many of the questions to which his earlier studies in theology and ethics had given rise. During the 1960s, Prior demonstrated that some very important contributions to the understanding of the concept of time can be obtained from the study of temporal logic.

In the following sections we intend to concentrate on these main trends in Prior's philosophical logic: (1) The logic of ethics; (2) How temporal logic began: Human Freedom and Divine Foreknowledge; (3) The logic of existence; (4) The syntax of tempo-modal logic; (5) The semantics of tempo-modal logic; and (6) Four grades of tense-logical involvement.

1   THE LOGIC OF ETHICS

In the introduction to his book *Logic and the Basis of Ethics* [1949], Prior pointed out that Aristotle divides the possible subjects of inquiry and dispute into three broad sorts — 'natural', 'ethical', and 'logical' [*Topics*, 105b19–29]. This is a worldview which Prior accepted. In his view, it is important to distinguish between natural, ethical, and logical statements. In accordance with this view, Prior rejected ethical naturalism i.e. the view that ethical propositions are just a subspecies of natural propositions. He agreed with G.E. Moore, who had criticised the deduction of 'ought' from 'is' (i.e. the so-called naturalistic fallacy), but Prior maintained that it would be a larger error to deny the autonomy of ethics [1949, p. 107]. Any of the fundamental Aristotelian sorts of inquiry can stand alone and none of them can be reduced to the others. However, this does not mean that the three sorts of inquiry are completely independent. In fact, the inquiry into any of them can benefit from the studies of the two others. Being a logician, Prior wanted to demonstrate that logic can be used in the study of ethics as well as in the study of nature.

Prior pointed out that the 'logic of ethics' is not a special kind of logic, nor a special branch of logic, but an application of it [1949, p. ix]. He maintained

that categorical obligations must lie on particular persons at particular moments. However, such (particular) obligations can be derived from general rules by adding a categorical premise, as it is done in the following argument: "If any debt falls due at any time, it ought to be paid at that time, And this debt falls due now; Therefore, this debt ought to be paid now" [1949, p. 41].

Prior is obviously not suggesting that we normally come to learn about our moral obligations by making this kind of inferences, but he maintains that we are sometimes led to a mistaken view of our present obligation because of making a mistake regarding the kind of situation we are in, or because of a mistake concerning a general moral rule.

For Prior, as for many other working with ethics, the notion of 'duty' is rather basic. In the 1950s G.E. Moore's definition of 'duty' in *Principia Ethica* was very influential. In this work Moore repeatedly affirms that our duty is that action which, of all the alternatives open to us, will have the best total consequences. In his paper, 'The Consequences of Actions' [Prior, 2003, pp. 65–72], which was originally presented at the "Joint Session of the Mind Association and the Aristotelian Society at Aberystwyth" in 1956, Prior argued that this definition turns out to be very problematic. Moore was clearly aware of the fact that in many cases it might turn out to be very difficult and perhaps even practically impossible to find out with any certainty what our duty is, given his definition of 'duty'. Obviously, these practical problems do not qualify as a logically compelling argument against the use of Moore's definition in ethical discourse. However, Prior's criticism of Moore's position was much more fundamental. Prior maintained that the very uncertainty of the future necessarily gives rise to a serious criticism of a utilitarian theory such as Moore's. Prior analysed the problems of the idea of consequences in a very entertaining manner referring to Mother Goose [2003, p. 68 ff.]:

> For want of a nail the shoe was lost;
> For want of a shoe the horse was lost;
> For want of a horse the rider was lost;
> For want of a rider the battle was lost;
> For want of a battle the kingdom was lost;
> And all for the want of a horse-shoe nail.

According to this text, the fate of the kingdom depended on a 'decisive' battle. If the cavalryman Bayard Bloggs had been in the field, the army would not have lost the battle. He would have been there if his horse hadn't been crippled through the loss of a shoe. The shoe would not have been lost if it had had one more nail in it. So the lost kingdom was the consequence of the missing nail.

This example is a very clear illustration of the fact that it is not always very clear what should be accepted as a consequence of a given act or behaviour. Maybe a defender of Moore's definition could answer that we should only take necessary consequences into account and that there is no necessary connection between the missing nail and the lost kingdom. But then it is not very clear what Moore

would want to take into account when we want to establish the totality of the consequences.

However, Prior's criticism of Moore's definition goes much deeper than to an analysis of the idea of a consequence. In fact Prior argued that there is a logical impossibility in there *being* such a thing as a duty in Moore's sense. Supposing that determinism is *not* true, Prior formulated his main criticism in the following way:

> Then there may indeed be a number of alternative actions which we could perform on a given occasion, but none of these actions can be said to have any 'total consequences', or to bring about a definite state of the world which is better than any other that might be brought about by other choices. For we may presume that other agents are free beside the one who is on the given occasion deciding what he ought to do, and the total future state of the world depends on how these others choose as well as on how the given person chooses ... [Prior, 2003, p. 65].

This is a very interesting argument. Although it was formulated already in 1956, Prior seems to be aware of the importance of the future choices of individuals in the future. It seems that he dealt with this problem when his analysis of the unstatability of the future (i.e. his system Q, which Prior introduced in *Time and Modality*) was still immature. Later he would probably have put even more emphasis on this way of criticism. Although one can spoil one's calculations of the future consequences alone with one's own future choices, it seems even more problematic when the influence of future individuals has to be taken into account. Because of the uncertainty related to the unstatability of the future, it will not be possible, even in principle, to calculate the totality of future consequences of a certain choice. From Prior's (Peircean) position this simply means that there is no such totality. For this reason he rejects Moore's idea of 'duty' as incoherent.

According to Prior, the only way out this problem which is open for the utilitarian involves another definition of 'duty'. Following this alternative definition, the 'duty' is to do what will *probably* have the best total consequences of all the actions open to us. Maybe there is no need to take the actual consequences of various possible choices into account, i.e. maybe there is no need to refer to what is in fact going to be case under various assumptions. Instead we might do the calculation based on probabilities. This means, however, that we have to be prepared to talk about objective probabilities, if we want 'duty' to be objectively defined. Prior suggests that 'p is probable' may mean something like 'p is not yet either going to be the case or not going to be the case, but is more like going to be the case than not'.

Prior's criticism of utilitarian theory should also be seen in the light of the fact that he wanted ethics to be treated theoretically in another way. His own contribution to ethical theory was mainly the formulation of a deontic logic involving operators corresponding to obligation and permissibility.
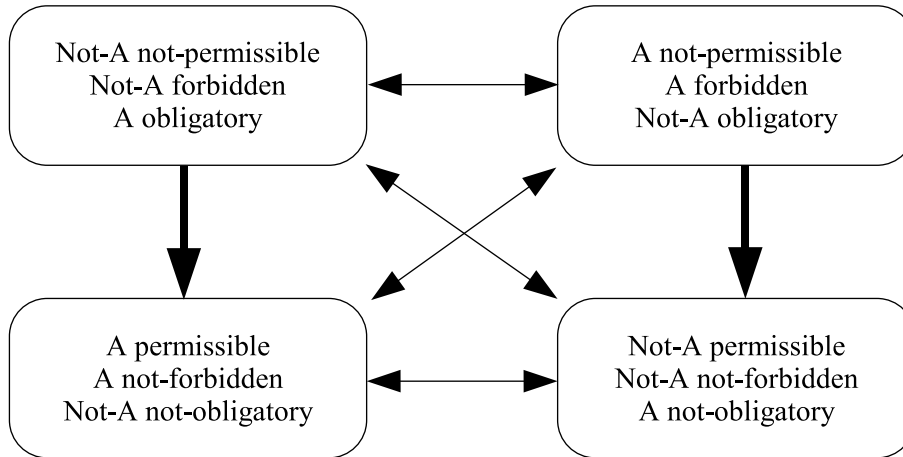
Figure 1. Mutual relations between some basic notions in deontic reasoning

In his *Formal Logic* [1955b], he made a number of interesting suggestions as to how the logic of ethics can be formalised. In fact the book contained a short chapter entitled 'deontic logic' — a name suggested by Henrik von Wright [1951]. Prior defended von Wright's view that the logic of obligation can be handled very much like the logic of necessity. He was, however, aware of the fact that many philosophers would resist this very much, insisting that moral philosophy has very little to do with logical deduction. In the interesting but still unpublished paper, *The Logic of Obligation and The Obligation of the Logician*, Prior wrote;

> To the moralist, the logician — especially when he talks about obliga-tion — is irresponsible; to the logician, the moralist is puritanical. I am frankly on the logician's party, and am anxious that moralists should understand a little better what our standards are [Prior, Unpublished c].

Prior's point was that although ethics cannot be deduced from logic, ethical argumentation has to live up to certain formal standards, which are certainly worth studying for their own sake.

Prior constructed the diagram in Figure 1, corresponding to the usual Aris-totelian logical square for syllogisms, explaining the mutual relations between some basic notions in deontic reasoning:

In his deontic logic Prior used P for 'it is permissible that (such-and-such an act be done)'. From this operator we may construct the operator $O = \sim P \sim$ corresponding to 'it is obligatory that . . .'. A deontic logic can be constructed by adding the following two axioms to propositional logic:

AD1: $Oa \supset Pa$
AD2: $P(a \vee b) \equiv (Pa \vee Pb)$

together with the rule

RD1: $\vdash \alpha \equiv \beta \quad \longrightarrow \quad \vdash P\alpha \equiv P\beta$

Prior demonstrated that in this axiomatic system it is possible to derive the following rule:

RD2: $\vdash \alpha \quad \longrightarrow \quad \vdash P\alpha$

Proof:

(1)	$\vdash \alpha$	[assumption]
(2)	$\vdash Oa \supset Pa$	[AD1]
(3)	$\vdash \sim P \sim a \supset Pa$	[from 2 and def.]
(4)	$\vdash P \sim a \vee Pa$	[from 3 and propositional logic]
(5)	$\vdash P(\sim a \vee a)$	[from 4 and AD2]
(6)	$\vdash \alpha \equiv (\sim a \vee a)$	[from 1 and propositional logic]
(7)	$\vdash P\alpha \equiv P(\sim a \vee a)$	[from 6 by RD1]
(8)	$\vdash P\alpha$	[from 5 and 7]

Q.E.D.

RD2 means that if α expresses a logical law, then it is a law that α is permissible. Prior renders this more freely as 'what I cannot but do, I am permitted to do'. This also amounts to 'what I cannot but omit, I am permitted to omit' and consequently also to the Kantian principle 'what I ought, I can'. A number of other interesting theorems can be proved in Prior's system, for instance:

$(Oa \wedge O(a \supset b)) \supset Ob$

(If doing what we ought commits us to doing something else, then we ought to do this something else.)

$\sim Pa \supset O(a \supset b)$

(Doing what is not permitted commits us to doing anything whatever.)

The latter example corresponds to one of the paradoxes of the strict implication.

In an appendix in his book *Time and Modality* [1957b], Prior discussed a different approach to deontic logic based on an idea from Alan Ross Anderson. According to this idea, a deontic logic can be established from modal logic by the addition of a propositional constant $\Re$ corresponding to the reading 'the world will be worse off', 'punishment ought to follow' or something of that sort. Given a modal propositional logic with a possibility operator $\diamond$ and the propositional constant $\Re$, we may define 'permissible' in the following way:

$Pa = \diamond(a \wedge \sim \Re)$

In accordance with this definition Oa should be seen as an abbreviation of $\square(\sim a \supset \Re)$, where $\square$ is the necessity operator defined as $\sim \diamond \sim$. In short this means that a is permissible if it is possible that a is the case without 'the bad thing' ($\Re$) being the case. Similarly, a is obligatory if $\Re$ necessarily follows from its negation.

Using these definitions, (AD2) can be immediately derived in most modal systems. Prior demonstrates that $\diamond a \supset Pa$ is equivalent to $\sim \diamond \Re$. Since it cannot

be accepted that all possible acts are permissible, Alan Ross Anderson suggested
the assumption of $\diamond \Re$. In fact, he proposed the axiom

$$\diamond \Re \wedge \diamond \sim \Re$$

which simply states that $\Re$ is contingent. Prior showed that the second part of
the axiom is deductively equivalent in most modal systems to AD1, ie. $Oa \supset Pa$.
He also demonstrated that in most modal systems it is possible to derive the
Kantian principle $Oa \supset \diamond a$ as well as the principle $(Oa \wedge O(a \supset b)) \supset Ob$.
Furthermore, he discussed the question of validity in various systems of more
complicated theorems such as $O(Oa \supset a)$ as well as the paradoxical $\sim Pa \supset O(a \supset b)$.

One of the few modal systems that may cause problems for the development of
deontic logic in Anderson's style is in fact Prior's own system Q, in which proper
becoming is taken into account. For instance, we may imagine that $\Re$ is true in all
possible states of affairs in which A.N. Prior exists and false in all other possible
states. In this example $\Re$ is obviously contingent, i.e. Anderson's axiom holds.
However, it means that according to Anderson's definitions whatever Prior does
will be forbidden, and yet it will be obligatory also — even impossible acts will be
obligatory! Prior calls this possible scenario "rather sombre and hyper-Calvinistic"
[1957b, p. 143]. It should be admitted that this example is odd, but the very fact
that Prior discussed how deontic logic could be treated in the context of Q, shows
that he was aware of the importance of integrating deontic and temporal notions.

Prior wanted to study the logical machinery involved in the theoretical deriva-
tion of obligation. He claimed that this study involves

      (a)   the description of the actual situation, and
      (b)   relevant general moral rules.

Prior stated his fundamental creed regarding deontic logic in the following way:

> . . . our true present obligation could be automatically inferred from
> (a) and (b) if complete knowledge of these were ever attainable [1949,
> p. 42].

Obviously, Prior wanted to present ethical argumentation as an axiomatic sys-
tem. But in doing so he obviously understood that something extra-logically has
to be taken for granted. In his unpublished draft *Logical Criticisms of the Theory
Identifying Duty with Self-interest* [Prior, Unpublished d], which he apparently
wrote from a lecture on ethics in 1947, he quoted C.S. Lewis, in "The Abolition of
Man" [1943, p. 21]: "If nothing is self-evident, nothing can be proved. Similarly
if nothing is obligatory for its own sake, nothing is obligatory at all." [Prior's
emphasis.] It seems to have been Prior's position that the axioms of deontic logic
have to be given extra-logically since they obviously cannot be deduced as long as
they are viewed as axioms.

It evident that Prior's long term ambition was to incorporate the logic of ethics
into a broader context of time and modality. Unfortunately, he was never able
to pursue this goal in detail, but he certainly managed to establish the broader
context of time and modality into which the logic of obligation has to fit.

2 HOW TEMPORAL LOGIC BEGAN: HUMAN FREEDOM AND DIVINE FOREKNOWLEDGE

Prior made a great and lasting contribution to philosophical logic; however, nothing similar can be said about his early work as "a predestinarian theologian", to use Kenny's term. This is one obvious reason why Prior's theological starting point as a thinker was ignored for many years. Moreover, in the course of the second half of the 1950s he himself gave up his religious beliefs and became agnostic. It should be noted, though, that also after this change he considered theological problems to be worthy of intellectual treatment, a point which can be seen not least in his paper 'The Formalities of Omniscience' from 1962.

The theological starting point is, however, of significance for Prior's work. First, his huge knowledge of theology, and in particular Medieval theologians, remained a source of inspiration for him throughout his career — many are the references to Ockham, Aquinas, Buridan and others even in his mature work. Second, his preoccupation with free will and human choice was evidently motivated by his struggle with the Presbyterian doctrine of predestination. Thus, an adequate understanding of Prior's philosophical logic is only possible if his background in Presbyterian theology is kept in mind.

As a Christian philosopher Prior was very much in favour of the use of logical argumentation in theology. This approach was consistent with his preference for Presbyterianism over Methodism, on account of the better systematicity of the former. Clearly, Prior held that the same standards of rationality must be applied to all realms of life. In this way theology is challenged — as well as enriched. Prior's position on the classical proofs of the existence of God was in line with these convictions. He was not against the very idea of trying to make proofs of God's existence, but he did emphasise that for a valid argument to constitute a proof, "it is requisite that those to whom it is addressed should be convinced of the truth of its premises ..." [1976b, p. 56]. He gave the following example as an illustration of the possible function of a proof of God's existence:

> A man may be absolutely convinced that only if God exists would he be obliged to live in a certain way — for example, to respect certain freedoms in other people even when violation of these freedoms seems the only way to avoid some grave social disaster — and may also be absolutely convinced that he is obliged to live in this way; and a man in this state of mind would surely be not only rational in drawing the conclusion that God exists, but positively irrational in not drawing it [1976b, p. 58].

However, the discussion of proofs of God's existence was not Prior's primary concern in theological argumentation. Rather, he concentrated on other parts of Presbyterian theology mainly related to ethics, predestination and time.

The roots of the Presbyterian denomination are Calvinist. The central tenet of the Reformation was that man could not save himself through his deeds. Rather,

salvation was pure grace, a gift from God, demanding only faith. However, this immediately raises the question whether faith is something man is free to accept or reject, or whether some are 'elected' to be believers — receiving passively the gift of faith — while others are not accorded that gift. The reformers differed on this point, but Calvin, at any rate, took a firm stand: indeed that there is no such thing as a free choice with respect to faith. Every person is predestined either to belief or disbelief, and thus to salvation or damnation. The most marked feature of Presbyterianism was, therefore, its teaching concerning predestination.

One remarkable defence of predestination, in effect determinism, is given in Prior's paper *Determinism in Philosophy and Theology* [Prior, Unpublished a]. The paper is kept in the Bodleian Library, and we have not been able to determine whether it has ever been published. The paper is difficult to date, but it was probably written in the mid-1940s. As the title suggests, the paper thematically compares the doctrine of predestination with philosophical determinism, respectively, indeterminism. The paper opens by observing that in 'modern discussions', determinism is often seen as a 'scientific creed' as opposed to the idea of free will, which is considered to be religious. But this perception is immediately countered [Prior, Unpublished a, p. 1]:

> It is exceedingly rare for philosophers to pay any great attention to the fact that a whole line of Christian thinkers, running from Augustine (to trace it back no further) through Luther and Calvin and Pascal to Barth and Brunner in our own day, have attacked freewill in the name of religion.

The paper then proceeds in four major steps:

First, it is emphasised that philosophical or scientific determinism is in part different from the idea of predestination: the Calvinism expounded by Barth and Brunner is not pure determinism, but a paradoxical mixture of determinism and free will [Prior, Unpublished a, p. 1]. They wish to replace the 'secular mystery of determinism', respectively, indeterminism, by the 'holy and real mystery of Jesus Christ.' Man is seen as unable to perform by himself an act of faith, but when, by the grace of God, he does perform it, that is an act of real freedom, 'free will for the first time'.

Second, it is argued that the ordinary ideas of free will, when understood as moral accountability and general indeterminism, are at least as absurd as the idea of predestination:

> We are guilty of that which we are totally helpless to alter; and to God alone belongs the glory of what we do when we are truly free. — Absurd as these doctrines appear, they are in the end no more so than the ordinary non-Augustinian concept of 'moral accountability' ... [Prior, Unpublished a, p. 2].

Third, Prior goes on to describe how certain human experiences actually are compatible with the notion of predestination, observing that

> Even those of us who accept a straightforward determinism have to
> give some account of men's feeling of freedom, and their feeling of guilt;
> and it is at least conceivable that the 'absurdities' of Augustinianism
> contain a more accurate psychological description of the state of mind
> concerned, than does the 'absurdity' of the ordinary non-Augustinian
> concept of 'moral accountability' [Prior, Unpublished a, p. 3].

Prior argues the plausibility of Augustinianism, that is, his doctrine of predestination, in the face of human experience. Up to this point, the paper — even if brief in its analysis — is a vivid and convincing defence of predestination, or determinism in an Augustinian sense. But this perception is modified in the final step of the analysis. In the fourth and concluding part, Freudian psychoanalysis is brought into the picture. It is argued that religious determinism is concerned with "particular inward compulsions and dependences", from which we can be released through (psycho)analysis [Prior, Unpublished a, p. 4].

The doctrine of sin and salvation in St. Paul and Augustine is seen as a partial psychoanalysis, leading to the conclusion that "The theological doctrine of predestination is a 'Theory of Obsessions', prefaced to the analysis of a particular case" [Prior, Unpublished a, p. 4]. Nevertheless, it is not quite clear whether this means that Christianity, and especially the doctrine of predestination, are 'subjected' to a psychoanalytical viewpoint, or whether it rather implies that evidence from psychoanalysis corroborates the idea of predestination within (Presbyterian) Christianity. The final remarks point in the former direction, the overall context rather points in the latter direction. We are not here dealing with a case of outright inconsistency, but there is a tension which may well reflect Prior's own state of mind at the time of writing.

Prior's logical studies increasingly led him away from what he regarded as indispensable parts of the Christian faith. In 1959, when he took up the professorship at the University of Manchester, he had become agnostic. He never declared himself an atheist, though. He remained respectful in his treatment of Christian belief as an intellectual possibility, but at least one unusually sharp remark in *Creation in science and theology* on Karl Barth reveals how Barth's theology, acknowledged as a pinnacle of theological thought in the Twentieth Century, had ceased to be of any value for him:

> One silly thing it's only too easy to do ... is to talk as if 'nothing'
> were the name of some kind of stuff out of which the world was made.
> I've even read a theologian (Barth) who [in his Dogmatics in Outline,
> 1949] talks as if 'nothing' were a sort of hostile power from which God
> rescued the world in giving it being [1959a, p. 89].

Even so, a modernist-liberal Christianity was not an option which lay open to Prior. He obviously saw such an approach as an almost dishonest and at any rate inconsistent way of thinking (very much the same spirit as his insistence on applying the full rigour of formal logic to theology, that is, that religion should

be rational if it were to be believed in at all). It is perhaps not too difficult to follow Prior in this on a general level, but maybe there is also a paradox here. Prior gave up Methodism in favour of Presbyterianism, finding the former 'unruly', but the latter consistent and well worked out. As mentioned in the previous section, an even more important reason for this shift had been his lack of any 'conversion experience', an ingredient of Christian faith which is strongly emphasised in Methodist theology. But at least as regards that troublesome point of predestination, Methodism is more congenial with the spirit of Prior's later indeterminist conviction. Methodism traces its roots to Jacob Arminius (1560–1609), who sought to modify the reformed faith exactly on the points appertaining to predestination: in particular, he taught that men were free to choose to believe.

At any rate, the founder of Methodism, John Wesley (1703–1791), was strongly influenced by Arminius, not least on this point. Thus, in a sense, Prior of his own accord left one interpretation of Christianity in favour of another one, whose most distinctive feature was that doctrine of predestination, which appears to have been a main motive for his later becoming agnostic.

Such observations can, of course, in no way detract from A.N. Prior's arguments. He has, perhaps more clearly than any other thinker, pointed out the *logical* limitations of foreknowledge. Likewise, he has shown and developed the logical possibilities for indeterminism.

Prior's stance on determinism was to change from the early fifties and onwards. Throughout the 1940s, he was interested in logic — mainly classical and non-symbolic logic — but apparently even more interested in philosophical and historical issues within theology. His first interest in modal logic was aroused in [1951], leading to the publication of *The Ethical Copula*. At this time he also developed into an adherent of indeterminism, and indeed, of free will.

Around 1953, Prior began to work on the development of a formal calculus of tenses. Mary Prior has described the first occurrence of this idea: "I remember his waking me one night, coming and sitting on my bed, and reading a footnote from John Findlay's article on Time, and saying he thought one could make a formalised tense logic." This must have been some time in 1953. The footnote which Prior studied that night was the following:

And our conventions with regard to tenses are so well worked out that we have practically the materials in them for a formal calculus ... The calculus of tenses should have been included in the modern development of modal logics. It includes such obvious propositions as that

x present = (x present) present;

x future = (x future) present = (x present) future;

also such comparatively recondite propositions as that

> (x).(x past) future; i.e. all events, past and future will be
> past [Findlay, 1941].

Findlay's considerations on the relation between time and logic in this footnote were not exactly elaborated, but it apparently gave the final impulse to Prior's idea of developing a formal calculus which would capture this relation in detail. For this reason Prior called Findlay "the founding father of modern tense logic". But there are, in our opinion, certainly not sufficient reasons for viewing Findlay as the founder of tense logic. The honour of being the founder must without doubt be ascribed to Prior himself. With his many articles and books on questions in tense logic he presented a very extensive and thorough corpus, which still forms the basis of tense logic as a discipline. Findlay's major merit in tense logic is to have had the luck of inspiring Prior to initiate the development of formal tense logic.

It seems that a short article by Benson Mates in particular made Prior even more aware of the interesting relation between time and logic. The paper in question was 'Diodorean Implication' [Mates, 1949]. The paper was concerned with Diodorean logic, primarily Diodorus' definition of implication. Prior realised that it might be possible to relate Diodorus' ideas to contemporary works on modality by developing a calculus which included temporal operators analogous to the operators of modal logic.

Prior believed that the problems of future contingents can be analysed and much better understood by the use of temporal logic. In his earliest attempt to deal with these problems he used Łukasiewicz's three-valued logic, in which the third value, $\frac{1}{2}$, was supposed to represent 'indeterminate' (see [Prior, 1953]). He suggested that this is the case for contingent statements such as the Aristotelian 'there is a sea-fight tomorrow' i.e. contingent statements of the form $F(1)p$.

Prior realised, however, that there is a serious problem with this approach. In fact, the usual truth-functional technique breaks down for these theories. For instance, if $F(1)p$ and $\sim F(1)p$ are both 'indeterminate' ($\frac{1}{2}$), it is very hard to explain how statements like the conjunction $F(1)p \wedge \sim F(1)p$ and the disjunction $F(1)p \vee \sim F(1)p$ could come out as anything else than 'indeterminate', when treated according to Łukasiewicz's three-valued logic [Prior, 1967b, p. 135]. Such results are, however, highly counter-intuitive, and they give rise to serious formal problems too. It turns out that the introduction of this kind of 'indeterminate' statements is an unnecessary complication. Evidently, Prior realised that Łukasiewicz's three-valued logic could not provide a satisfactory solution of the problem of future contingents.

Prior's early work on the logic of time also led to the paper *Diodoran Modalities* [1955a]. (Prior later changed his spelling into 'Diodorean', in accordance with Mates.) From the very outset of Prior's development of tense logic, the problem of determinism was dealt with in parallel with the logic of time. It is clear that the determinism-issue has roots in the problem of predestination, and that Prior's dealing with this issue was a natural continuation of his earlier preoccupation with

predestination. As Jack Copeland has argued "there can be no doubt that Prior's interest in tense logic was bound up with his belief in the existence of real freedom" [1996, p. 16]. In fact, his paper on Diodorean modalities, which was his very first proper study of tense logic, was an analysis of an ancient argument in favour of determinism, the Master Argument of Diodorus [1955a]. Interpreted with respect to its theological implications, this argument calls into question whether the idea of free will can ever be reconciled with the doctrine of divine foreknowledge, and hence, with the doctrine of divine omniscience. Accordingly, and in order to preserve the possibility of real freedom, Prior rejected the traditional version of the doctrine of God's foreknowledge. Prior concluded from his analysis of the Diodorean argument that for some (contingent) p, which is assumed to be true now, God has never known that p would be the case. It is obvious that this position is very far from Presbyterian theology. In section 4, we shall study Prior's work on the Master Argument in somewhat greater detail.

In *The Formalities of Omniscience* [1962a] he further investigated the problems of determinism and foreknowledge. The paper examines the idea of omniscience, especially in the form of the statement "God is omniscient", and some putative consequences of it, such as:

> (7) It is, always has been, and always will be the case that for all p, if p then God knows that p [2003, p. 43]

and

> (8) For all p, if (it is the case that) p, God has always known that it would be the case that p [2003, p. 43].

Various interpretations of such statements are discussed, especially with reference to St. Thomas Aquinas. It is argued that for logical reasons future contingents cannot be 'known' at all, leading to the observation: "I don't think we get my proposition '8' ... except in the weak sense that He [God] knows whatever is knowable, this being no longer co-extensive with what is true" [1962a, p. 122]. (This is inconsistent with Prior's former remark that future contingents are not "strictly speaking true", but the point that a truth value for such propositions cannot be known is clear in any case). Prior concludes with the following statement (which may be indicating not an atheist, but rather an agnostic position):

> I agree also with the negative admission of Thomas ... that God doesn't know future contingencies literally ... But (and this is what Thomas himself says) this is only because there is not then any truth of the form 'It will be the case that p' (or 'It will be the case that not p') with respect to this future contingency p, for Him to know; and *nihil potest sciri nisi verum* [nothing can be known except (what is) true] [2003, p. 58].

To be true, Prior argued against Thomas' view that God's knowledge is in some way beyond time, but otherwise he consented to most of what Thomas had

said about tense-logical reasoning. According to Prior's interpretation of Thomas' philosophy, Thomas would even agree on the rejection of the following Diodorean assumption:

> necessarily, if p, then it has always been the case that in the future p would be the case.

On the basis of his studies of medieval logic Prior developed an argument regarding the contingent future and divine foreknowledge. This argument was often formulated in terms of metric tense logic, i.e. by the use of the following to operators:

$F(x)$ "in x time units it will be the case that ..."
$P(x)$ "x time units ago it was the case that ...".

In the argument two other operators are also needed, namely

$\square$ "it is necessary that ..."
D   "God knows that ...".

In *The Formalities of Omniscience* [1962a] as well as other writings Prior presented several versions of the argument. The most interesting version can be rephrased by using the following 5 principles:

(P1) $F(y)A \supset P(x)DF(x)F(y)A$ (Divine Foreknowledge)
(P2) $\square(P(x)DF(x)A \supset A)$ (Infallibility of God's knowledge)
(P3) $P(x)A \supset \square P(x)A$ (The fixity of the past)
(P4) $(\square(A \supset B) \wedge \square(A)) \supset \square B$ (Basic assumption about modality)
(P5) $F(x)A \vee F(x){\sim}A$ (Principle of the excluded middle)

Here A and B represent arbitrary well-formed statements within the logic. Let q stand for some atomic statement so that $F(y)q$ is a statement about the contingent future.

(P1) states that if something is going to happen, God has already known for some time that it is going to happen. According to (P2), if it was the case x time units ago that God knew that A would be the case x time units later, then it necessarily follows that A is the case now. The principle (P3) means that if A was the case x time units ago, then it is necessary that it was the case x time units ago. (P4) is a basic assumption in modal logic, and (P5), which is about the determinateness of the future, states that either A is going to be the case in x time units or ${\sim}A$ is going to be the case in x time units.

The argument proceeds in two phases: first from divine foreknowledge to necessity of the future, and from that argument to the conclusion that there can be no real human freedom of choice. Formally, the argument goes as follows:

(1) $F(y)q$ (assumption)
(2) $P(x)DF(x)F(y)q$ (from 1 & P1)
(3) $\square P(x)DF(x)F(y)q$ (from 2 & P3)
(4) $\square(P(x)DF(x)F(y)q \supset F(y)q)$ (from P2)
(5) $\square F(y)q$ (from 3, 4, P4)

In this way it is proved that

(6) $F(y)q \supset \Box F(y)q$

and similarly it is possible to prove

(7) $F(y)\sim q \supset \Box F(y)\sim q$

The second part of the main proof is carried out in the following way:

(8) $F(y)q \vee F(y)\sim q$ (from P5)
(9) $\Box F(y)q \vee \Box F(y)\sim q$ (from 6, 7, 8)

Here (9) is equivalent to a denial of the dogma of human freedom. Therefore, if one wants to save this dogma (and escape fatalism) at least one of the above principles (P1–5) has to be rejected. Prior realised that this can be obtained in several ways. He argued, however, that two of them are particularly important, i.e. the denials of (P3) and (P5). The solution based in the denial of (P3) is called the Ockhamistic solution. According to this view, not all propositions formulated in the past tense should be treated as statements properly about the past, and (P3) should only be accepted if $P(x)A$ is a statement about the proper past. Obviously, this would rule out the use of (P3) to deduce (3) from (2), since $P(x)DF(x)F(y)q$ is clearly not a statement about the proper past.

Prior's own position was that (P3) should in fact be accepted, whereas (P5) should be rejected. His view on future contingents was that their truth value cannot be known now, not even by God, that is, there are no true statements about future contingents. On this view, the statement 'there will be a sea-battle tomorrow' (this example being taken from Aristotle's classical discussion of future contingent in *Prior Analytics*) cannot be true today, and the same is the case for the statement 'there will be no sea-battle tomorrow'. Prior would maintain that both of these statements are in fact *false* today, and suggested the following condition of truth with respect to future statements:

> ... nothing can be said to be truly 'going-to-happen' (futurum) until it is so 'present in its causes' as to be beyond stopping; until that happens neither 'It will be the case that p' nor 'It will not be the case that p' is strictly speaking true [2003, p. 52].

Prior held that the proposition $F(x)p$ can only be true if it is in principle possible to verify it from facts known at the time of utterance. Obviously, the same can be said about $F(x)\sim p$. According to his view, future tense propositions are false if they cannot be verified. As a consequence, the proposition $F(x)p \vee F(x)\sim p$ is false according to this view, if $F(x)p$ is a statement about the contingent future.

As indicated above it was Prior's conviction that St. Thomas Aquinas also held these ideas. Prior also pointed out that this position regarding the contingent future is quite essential in Peirce's philosophy. In fact, Prior called the way of answering the problems of arguments like the one presented above the Peircean

solution. This view means that he had to reject $q \supset P(x)F(x)q$ as a thesis. If q is true now, but not something which had to be true (by necessity), then the Peircean solutions implies that $F(x)q$ was false x time units ago, for some x.

The view that statements about the contingent future are false was expressed rather early in Prior's writings. For instance, in *Some Free Thinking About Time* [Prior, 1996a], which is written sometime during the 1950s, he stated his belief in indeterminism as well as the limitations of divine foreknowledge very clearly:

> I would go further than Duns Scotus and say that there are things about the future that God doesn't yet know because they're not yet to be known, and to talk about knowing them is like saying that we can know falsehoods [Copeland, 1996, p. 48].

So even God cannot know the contingent future for the simple reason that knowing the contingent future would turn out to be the same as knowing a falsehood. This view was obviously in conflict with his former Presbyterian belief, but he saw this position as a necessary consequence of his belief in human freedom of choice. He explained this belief in the following way:

> I believe that what we see as a progress of events is a progress of events, a coming to pass of one thing after another, and not just a timeless tapestry with everything stuck there for good and all ... This belief of mine ... is bound up with a belief in real freedom. One of the big differences between the past and the future is that once something has become past, it is, as it were, out of our reach — once a thing has happened, nothing we can do can make it not to have happened. But the future is to some extent, even though it is only to a very small extent, something we can make for ourselves ... if something is the work of a free agent, then it wasn't going to be the case until that agent decided that it was [Copeland, 1996, p. 47–48].

For many years, Prior saw no conflict between his faith and his insistence on the freedom of inquiry and criticism. But as we have seen, he gradually came to doubt the dogmas of Christianity. One is tempted to formulate a 'trilemma':

- the doctrines of predestination and foreknowledge are integral parts of the Christian faith,

- the doctrine of foreknowledge is untenable for intrinsic logical reasons, and the doctrine of predestination is incompatible with a belief in indeterminism and free will,

- any convenient 'abbreviation' of Christianity is dishonest and untrustworthy.

The last paper, wherein Prior seems to be endorsing Christian faith, if only vaguely, is *The good life and religious faith* [1958]. This is a discussion between

Prior and a few other philosophers on religion — among them John Mackie. Prior seems at this point to be still 'defending' religion (Christianity) in replies to Das and Mackie. However, one statement by Mackie seems to anticipate an essential reason why Prior became agnostic. The statement Mackie makes is this:

> In fact I think it [religion] hostile to the good life, because of the value it always puts upon firm belief for inadequate reasons. It blocks inquiry, which is a principal ingredient of the good life [Prior, 1958, p. 10].

Prior became agnostic because he came to see Christianity as an obstacle to the freedom of inquiry — in particular with respect to the doctrine of predestination, but also at a general level. According to Mary Prior [Hasle, 2003, p. 301–302], he was preoccupied by the problem of free will, and he was certainly aware of the dilemmas Calvinism posed. Mary Prior has suggested that his failure to resolve them was a reason why despite so much preparation the book on Scottish Theology never came to anything. His logic led him to the conclusion that the future must be open to choice. The idea of free choice also seems to have been very important for him emotionally.

Prior's commitment to a genuine freedom of choice clearly had an ethical dimension, too. Freedom of choice is often seen as a precondition of human moral accountability. Even if freedom of choice is not a necessary condition of moral accountability — as it is asserted in so-called compatibilism regarding determinism and human freedom — it is clearly a sufficient condition (at least if taken together with the condition that one is not by force prevented from exerting it). This observation establishes a connection between Prior's work on tense logic and his investigations into the logic of ethics.

3   THE LOGIC OF EXISTENCE

As is obvious from his studies of the logic of foreknowledge and divine omniscience, Prior wanted to see the future as open and certainly not as completely determined and settled. Some things are evidently the works of free agents. In fact, by such works things can come into existence. In dealing with this view, Prior often considered fundamental questions concerning the logic of quantification. He wanted a deeper understanding of how quantification and modality can be combined. In particular, he wanted to describe the relation between existence in time and quantification.

In their interesting essay on Prior's philosophy, Philip Hugly and Charles Sayward [1996, p. 240] have argued that according to Prior there are non-eliminable, non-substitutional, non-objectual, non-referential kinds of quantification. They have suggested that following Prior's ideas, quantification can be presented as "a method for constructing general sentences applicable to virtually any type or category of term" [1996, p. 265]. This is very well put. Prior's view on quantification was obviously different from that of Quine. Prior explained the difference in the following way:

Quine says in effect that non-existents cannot figure as the values of bound variable. I would suggest that, on the contrary, this is the only way in which non-existents of this sort can figure. I cannot directly refer to what does not exist but is merely imagined to exist, or is merely going to exist; but I can make purely general (i.c. quantified) statements about the imaginary or future denizens of the world. The quantification, however, must occur within a 'modality' [Prior, 2003, p. 220].

Prior's intuition seems rather convincing. He wanted to maintain a clear logical difference, which he illustrated using the following example [1957b, p. 26]:

    (a)   It will be the case that someone is flying to the moon.

    (b)   There is someone who will fly to the moon.

Here Prior obviously understands (b) as "There is someone presently existing who is going to fly to the moon". If F stands for the future operator, the structure of (a) is obviously $F(\exists x : p)$ (i.e. a quantification "within a modality"), whereas the formal structure of (b) is $\exists x : Fp$. — The relation between statements like (a) and (b) had been studied by Ruth Barcan Marcus already in 1946 in an attempt to combine modal logic with quantification theory. In particular Ruth Barcan Marcus [Barcan, 1946] had studied systems in which the following formula holds:

$$F(\exists x : p) \supset \exists x : Fp$$

This formula is now known as Barcan's formula and it can of course be discussed for all kinds of modal operators. Prior maintained that Barcan's formula should not hold in general for the future operator. He wanted a clear logical distinction between quantification "within a modality" and quantification outside the scope of a modality.

However, Prior realised that for formal reasons it is rather difficult to keep the quantification within a modality. With just a few seemingly quite straight forward axioms of tense logic and Prior's own general theory of quantification, Barcan's formula for the future operator becomes provable. In dealing with this logical problem, Prior found that he needed a logical system, in which the notion of statability is taken into account. The reasoning is that because new things have been brought into existence today, there are some statements which can be stated today, but which could not be stated yesterday. This was probably Prior's main motivation for his proposal in 1957 of the modal system Q wherein it is assumed that in certain possible worlds, some propositions simply cannot occur. An obvious example could be propositions directly concerned with individuals, which are absent from those worlds. According to Prior, no facts can be stated about an individual x except when x exists.

In 1959 Prior described the basic idea of the system Q in the following way:

Nothing can be surer than that whereof we cannot speak, thereof we must be silent, though it does not follow from this that whereof we could not speak yesterday, thereof we must be silent today [Prior, 1959b].

When translated into tense logical terms, the system Q offers an interesting example of a logical system which is among other things designed to solve problems associated with non-permanent or contingent existents.

It is interesting to study the problem of statability and its implications for the philosophy of time. It turns out to be a very difficult task to establish a tense logical formalism within which we can deal with the temporal aspects of statability in a satisfactory way (see for instance [Wegener and Øhrstrøm, 1997]). However, the basic idea is rather obvious. In particular, it becomes evident when we are dealing with identifiable individuals. The very fact that individuals come into being makes it impossible for us to formulate crucial statements about such individuals in a satisfactory way before they have actually been brought into being. As Prior has pointed out, the statement 'It is not the case that Julius Caesar existed in 200 BC' makes sense, but here it is important that the main verb is in the past and not in the present tense [Prior, 2003, p. 92]. In 200 BC a statement like 'Julius Caesar does not exist' would not make any sense. It was simply not statable then.

It may be argued that many future tense statements are not about particulars, but rather about types. However, this observation certainly does not solve the problem of statability. Prior's claim regarding non-statability is not only about the non-existence of subjects of predication. It is also a question about other parts of the vocabulary. The point is that new concepts, i.e. new predicates, may arise. This means that the language of specification may be growing in a very radical manner.

Reflecting on the temporal aspects of statability, Prior maintained that the passage of time not only means that more and more possibilities are lost. It also gives rise to new possibilities for us as new individuals come into being. In his own words:

> Hence, while the passage of time may eliminate 'possibilities' in the sense of alternative outcomes of actual states of affairs, and cause that to be no longer alterable which once might have been otherwise, with 'logical' possibilities the opposite change occurs. For as new distinguishable individuals come into being, there is a multiplication of the number of different subjects to which our predications can be consistently attached, and so a multiplication of distinguishable logical possibilities [Prior, 2003, p. 91].

This means that we have no way of dealing with all future possibilities — not even in principle. Some states of affairs, which we may in fact later regard as very important, cannot be incorporated in a satisfactory way in the present scope of possibilities. The problem is that these states of affairs simply cannot be described in a sufficiently precise manner. For this reason they cannot be taken into serious account today. This means that we cannot even discuss the probabilities of such non-statable possibilities.

However, even if we do not take the question of statability into account there will still be serious problems regarding time and existence. In particular, Prior

was interested in the questions concerning identity of things over time. How can one thing at one time be the same as another thing at another time? How can a thing keep its identity over time? How can we be sure that individual things never split up into two (or more) identical individual things?

In a quite entertaining story called "The Fable of the Four Preachers" [Prior, Unpublished b] Prior illustrated the problems regarding identity over time. The story is about four churches (sects) and their preachers in a fictive city in Massachusetts and the beliefs in these sects regarding life after death. In what Prior called Sect A it is believed that "when this life is over we go to another place, where our happiness and misery depend on whether we have behaved well or ill down here". However, the adherents of Sect A also believe "that in the other world we have no memories at all of the present one". Sect B is more modernistic, since its adherents hold that death is in fact the end, although they do believe in the existence of the other world and that we can in fact in this life influence life in the other world. They believe "that as soon as anyone in our own world dies, another — *quite* different person — comes into being in the other world; and that Providence has so arranged it that the happiness or misery of this other person depends on whether the person who has just died has behaved well or ill during his life (his *only* life, of course)".

The adherents of Sects C and D agreed in holding, "not only that there is a life after death, but also that in the other world we *do* remember a great deal about what we did and experienced here below". According to Sect C, however, "the other world is a *much* vaster place than this one, with many wide open spaces to be filled up, so that God has decreed that when each person moves from this world to the next, he turns up there not as one but as several, each of whom clearly remembers *having been* the person who died, and each of whom indeed was the person who died. And all of them suffer for his sins — and justly, for as they very well know, they were *their* sins". The adherents of Sect D argued that such claims of the preacher of Sect C were rather absurd. They believed their preacher, on the contrary, who explained that "the other world is bothered with a population problem — generation after generation keep pouring into it, as they die, from here, and if steps were not taken it would soon be quite intolerably crowded. Steps *are* taken, however; what God has arranged is that when several people down here die simultaneously, they all become a single individual in the other world, who remembers perfectly well having been all of them, and who indeed was all of them".

Prior writes in the fable that the local sceptics in Massachusetts were inclined to regard the tenets of Sects C and D as logical impossibilities. However, they "for reasons which they found it very difficult to make clear even to themselves, found the 'fusion' doctrine of Sect D appreciably more impossible to stomach (if there can be degrees of impossibility) than the 'fission' doctrine of Sect C".

The fable at the same time raises questions concerning 'temporal identity' and theology. It is clear that the idea of (temporal) trans-world identity is at stake here. What is the moral implication, for instance, of Sect A's view that a person

without memories of his previous existence is nevertheless punished or rewarded for things of which he has no knowledge? (Indeed, what sense does it make to say that it is the same person?) What does the multiplication of an individual according to Sect C mean? (And does this idea suggest a branching-time-like picture?) As for Sect D, their conception can seem unintelligible. It has, however, a possible affinity to some interpretations of Calvinism, wherein the Elect are elected only in Christ — and not at all in themselves — and in a sense, live on only in Christ. If that is what is here hinted at, the paper may be seen as a very interesting holding-together of some classical Christian ideas and the (temporal study of) questions concerning time and identity. A limitation to this interpretation is that the sects — at least A, B, and C — all hold that the states in the "after-world" somehow depend on deeds, as opposed to the Protestant and Calvinist emphasis on salvation as dependent on faith and the sheer "grace of God". But it may be at telling fact that Sect D is the one which keeps silent on the question of how life may be in the other world.

"The Fable of the Four Preachers" may be seen as a nice illustration of the logical problems Prior was trying to solve in his search for a logic of identity. This turned out to be a rather complicated matter, among other things because it seems to be almost impossible to explain how the meaning of what the preacher of Sect A is saying differs from the meaning of what the preacher of Sect B is saying. However, Prior was in particular interested in the problems to which a position like that of Sect C can give rise. In the paper *Time, Existence, and Identity* [1965 1966] (republished in [Prior, 2003]) he analysed the crucial question: 'Can one thing become two?'. He stated:

> There do seem to be at least approximations to this in nature, e.g. the 'multiplication by division' of unicellular organisms, and still closer approximations to it seem to be easily imaginable, e.g. conscious organisms which divide in two and retain after division a clear memory of their undivided state [2003, p. 96].

Suppose that x and y are two different individuals which were identical n time units ago, i.e. $P(n)(x = y)$. Prior assumed that the following propositions hold:

(1) $q \supset P(n)F(n)q$
(2) $P(n)F(n)q \supset q$
(3) $(x = y) \supset (\phi x \supset \phi y)$.

Prior showed that (3) is in fact equivalent with Leibniz' principle ('the identity of the indiscernibles'). From this principle we can prove:

(4) $P(n)(x = y) \supset (P(n)\phi x \supset P(n)\phi y)$.

Let us assume that the object x has the property ϕx, i.e. ϕx holds now. From (1) follows that $P(n)F(n)\phi x$ also holds. Substituting $F(n)\phi x$ for ϕx (4), this leads to $P(n)F(n)\phi y$. Using (2) we deduce that ϕy also holds now. This means that x and y have exactly the same properties now. But how can x and y then be different now given Leibniz' principle? It appears that this deduction has led us to a contradiction.

Prior suggested that we have to drop (1) which, as we saw in the previous sections is also dubious for other reasons. However, he also realised that this move is not enough to prevent us from encountering other troubles. In fact Leibniz' principle itself can also be questioned, since it may easily be seen that it leads to the principle of transitive identity:

(5) $(x = y) \supset (z = x \supset z = y).$

Prior argued that (5) gives rise to a contradiction, if one can become two. He wrote:

> Let us suppose that the single individual x has become the two individuals y and z. If x has really *become* these two individuals, and has not simply ceased to exist and been in some sense replaced by them, then if anyone were to ask 'Where is x now?', one correct answer would be to say 'Here he is' and point to y. In other words, x is now y, and it would perhaps also be true to say that it is y who is now x, i.e. y is now x. But it would be equally correct to answer the question 'Where is x now?' by saying, 'Here he is', and pointing to z. In other words, x is now z ... [Prior, 2003, p. 98].

From this it can obviously be concluded that 'y is now z', which clearly contradicts the assumption. The only satisfactory way out seems to be denial of Leibniz' principle. Prior suggested that we at least have to weaken (5) to the following, where 'I' stands for a general identity relation [Prior, 2003, p. 100]:

$$Ixy \supset (Ixz \supset (Iyz \lor PIyz \lor FIyz).$$

Regarding the 'fusion' doctrine of Sect D Prior was like the sceptics in Massachusett inclined to regard it as even more problematic and unlikely than the 'fission' doctrine of Sect C'.

4 THE SYNTAX OF TEMPO-MODAL LOGIC

A persistent feature throughout his works is a clear interest in the history of logic. Indeed, Prior took an interest in the history of logic not only as a subject in its own right, but also because he saw the works of ancient and medieval logicians as a significant contribution to the contemporary development of logic. In fact, Prior revived the medieval attempt at formulating a temporal logic for natural language. In a short but thought-provoking sketch of the history of logic with a special emphasis on tense-logic, Prior has argued that the central tenets of medieval logic with respect to time and tense can be summarised in the following way:

(i) tense distinctions are a proper subject of logical reflection,

(ii) what is true at one time is in many cases false at another time, and vice versa [1957b, p. 104].

Prior observed that ancient and medieval logicians took these assumptions for granted, but that they were eventually denied (or simply ignored) after the Renaissance. In fact the waning of tense logic began with a gradual loss of interest in

temporal structures, that is, it was (i) which was first abandoned by the different schools of logic, and (ii) came to be rejected only afterwards.

Prior can be said to have realised the possibility of (re)formulating a logic based on these old assumptions. His first hint at the possibility of a logic of time-distinctions is found in the unpublished manuscript *The Craft of Logic* 1951 (cf. [Copeland, 1996, p. 15]). In 1953, when he was reading a paper of Findlay titled "Time: A Treatment of Some Puzzles" [1941], he decided to take up Findlay's challenge of working out a calculus of tenses. Major sources for him were also Łukasiewicz' discussion of future contingents [1920], which was inspired by Aristotle's *De Interpretatione*, and the Diodorean *Master Argument*, which he came to study via a paper by Benson Mates on *Diodorean Implication* [1949]. As we have seen, he very early demonstrated that tense logic can be used as a powerful tool in the analysis and reconstruction of the Master Argument.

In fact, one of his very first proper studies in tense logic was an analysis of an ancient argument in favour of determinism, the Master Argument of Diodorus [1955a]. This argument was constructed by Diodorus Cronus (ca. 340–280 BC), who was a philosopher of the Megarian school, and who achieved wide fame as a logician and a formulator of philosophical paradoxes [Sedley, 1977]. Unfortunately, only the premises and the conclusion of the Master Argument are known. We know almost nothing about the way in which Diodorus used his premises in order to reach the conclusion. It is, however, known that the Master Argument was presented as a trilemma. According to Epictetus, Diodorus argued that the following three propositions cannot all be true [Mates, 1961, p. 38]:

(Dl)   Every proposition true about the past is necessary.
(D2)   An impossible proposition cannot follow from (or after) a possible one.
(D3)   There is a proposition which is possible, but which neither is nor will be true.

Diodorus used this incompatibility combined with the plausibility of (D1) and (D2) to justify that (D3) is false. Assuming (D1) and (D2) he went on to define possibility and necessity as follows:

(D◇)   The possible is that which either is or will be true.
(D□)   The necessary is that which, being true, will not be false.

The reconstruction of the Master Argument certainly constitutes a genuine problem within the history of logic. It should, however, be noted that the argument has been studied for reasons other than historical. First of all, the Master Argument has been read as an argument for determinism. Secondly, the Master Argument can be regarded as an attempt to clarify the conceptual relations between time and modality.

Prior's reconstruction [1967b] of the Master Argument is based on the assumption that the statements in question are in fact propositional functions whose truth-values can vary from time to time. Thus it basically adopts the same understanding of 'proposition' and consequence as we have been arguing for above. Prior uses his tense- and modal operators in the reconstruction:

$$P: \quad \text{"it has been the case that } \ldots \text{"}$$

$$F: \quad \text{"it is going to be the case that } \ldots \text{"}$$

$$H(= \sim P \sim): \quad \text{"it has always been the case that } \ldots \text{"}$$

$$G(= \sim F \sim): \quad \text{"it will always be the case that } \ldots \text{"}$$

$$\Diamond: \quad \text{"it is possible that } \ldots \text{"}$$

$$\Box(= \sim \Diamond \sim): \quad \text{"it is necessary that } \ldots \text{"}.$$

On these assumptions it is possible to restate the reconstruction problem. Using symbols, (D1–3) can be formulated in the following way:

(D1′) $Pq \supset \Box Pq$

(D2′) $((p \rightarrow q) \wedge \Diamond p) \supset \Diamond q$

(D3′) $(\exists r)(\Diamond r \wedge \sim r \wedge \sim Fr)$

where $\rightarrow$ is the strict implication defined as

$$p \rightarrow q \equiv \Box(p \supset q).$$

We are now ready to reformulate Prior's reconstruction. It is, however, clear that Prior is not able to reconstruct the argument only using (D1), (D2) and (D3). In addition to these, he needs two extra premises. He must assume the thesis

$$(\sim q \wedge \sim Fq) \supset P \sim Fq$$

or, to put it in a general form:

(D4) $(p \wedge Gp) \supset PGp$

where $G \equiv \sim F \sim$ ('it will always be the case that . . .'). Furthermore, he must assume that

(D5) $\Box(p \supset HFp)$

is valid in general.

Prior's proof that the three Diodorean premises (D1′, D2′, D3′) are inconsistent given (D4) and (D5) can be summarised as a *reductio ad absurdum* proof in the following way:

(1)	$\Diamond r \wedge \sim r \wedge \sim Fr$	(from D3′)
(2)	$\Diamond r$	(from 1)
(3)	$\Box(r \supset HFr)$	(from D5)
(4)	$\Diamond HFr$	(from D2′, 2 & 3)
(5)	$\sim r \wedge G \sim r$	(from 1)
(6)	$PG \sim r$	(from 5 & D4)
(7)	$\Box PG \sim r$	(from 6 & D1′)
(8)	$\sim \Diamond HFr$	(from 7; contradicts 4)
	Q.E.D.	

O. Becker [1960] has shown that the extra premises (D4) and (D5) can be found in the writings of Aristotle. For that reason Becker concludes that it seems reasonable to assume that the extra premises were generally accepted in antiquity.

However, for historical reasons Prior's addition of (D4) and (D5) is nevertheless problematic. (D4) is in fact a rather complicated statement and not so innocuous as it may seem at first glance — observations which will indeed become clear when we are going to discuss the Ockhamist and Peircean systems. It is not very

likely that Diodorus would involve such an argument without making it an explicit premise in the Master Argument. As regards (D5), we know that Diodorus used the Master Argument as a case for the definitions (D◇) and (D□). That is, in the argument itself ◇ (or □) should in a sense be regarded as primitive. It is hard to believe that Diodorus would involve a premise about □ without stating it explicitly. As we have demonstrated elsewhere [Øhrstrøm and Hasle, 1995, pp. 23–8], there is another possible reconstruction of the Master Argument, which for historical reasons should be considered to be more likely than Prior's. But obviously Prior's suggested reconstruction is interesting in its own right as an argument in favour of determinism. Being an indeterminist, Prior obviously could not accept the deterministic conclusion of the argument he had reconstructed. Since he accepted the derivation of the conclusion from the premises he had to reject at least one of the premises. In fact, he questioned the validity of (D5) i.e.

(D5) $\Box(p \supset HFp)$.

If we understand 'will be' as 'determinately will be', then according to Prior (D5) should certainly be denied. As explained in section 1 Prior based this denial on metaphysical reasoning. He claimed that the conjunction $p \wedge \sim HFp$ is in fact possible i.e. something may be the case right now (p) although it was not always true to say that it would be the case ($\sim HFp$).

During the 1950s and the 1960s Prior developed his calculus of tenses into a rather sophisticated formalism. In particular he was interested in a system as weak as possible, i.e. a system in which no assumptions are made regarding the structure of time. He formulated this minimal tense logic K_t, which was also studied by John Lemmon, in the following way [1967b, p. 176]:

Axioms:

(A1) p, where p is a tautology of the propositional calculus
(A2) $G(p \supset q) \supset (Gp \supset Gq)$
(A3) $H(p \supset q) \supset (Hp \supset Hq)$
(A4) $p \supset HFp$
(A5) $p \supset GPp$.

Rules:

(RMP)   If $\vdash p$ and $\vdash p \supset q$, then $\vdash q$.
(RG)   If $\vdash p$, then $\vdash Gp$.
(RH)   If $\vdash p$, then $\vdash Hp$.

In 1958 he entered into a very interesting correspondence with Charles Hamblin of The New South Wales University of Technology in Australia. Their correspondence led to important results, especially on implication relations among tensed propositions. Prior and Hamblin discussed two central issues in tense logic: the number of non-equivalent tenses, and the implicative structure of the tense operators. In 1958 Hamblin suggested a set of axioms with P and F as monadic operators, corresponding to "a simple interpretation in terms of a two-way infinite continuous time-scale". Hamblin's axioms are:

Ax1: $F(p \vee q) \equiv (Fp \vee Fq)$
Ax2: $\sim F \sim p \supset Fp$
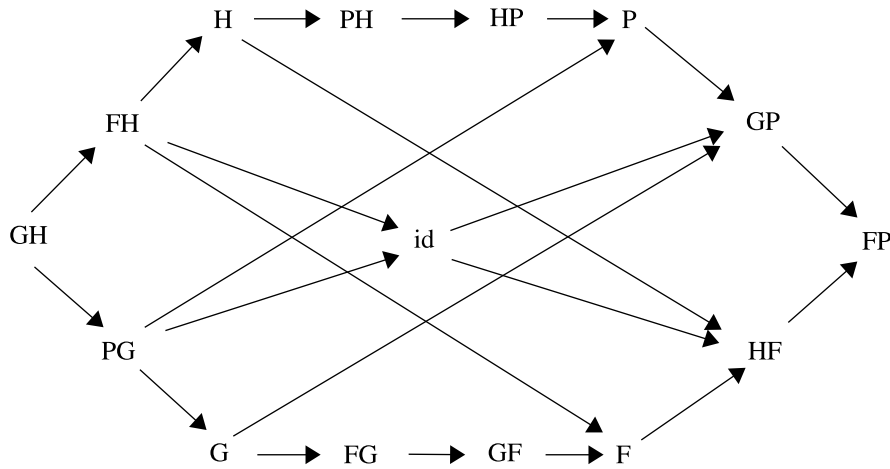
Figure 2. Hamblin's and Prior's implicative structure for the non-metrical tense-operators

$$
\begin{array}{ll}
\text{Ax3:} & FFp \equiv Fp \\
\text{Ax4:} & FPp \equiv (p \vee Fp \vee Pp) \\
\text{Ax5:} & \sim\! F \!\sim\! Pq \equiv (q \vee Pq).
\end{array}
$$

Hamblin also assumed 3 rules of inference:

> R1:   If A is a thesis, then $\sim\! F \!\sim\! A$ is also a thesis.
>
> R2:   If $A \equiv B$ is a thesis, then $FA \equiv FB$ is also a thesis.
>
> R3:   If A is a thesis, and A' is the result of simultaneously replacing each occurrence of F in A by P and each occurrence of P in A by F, then A' is also a thesis. (A' is the so-called mirror-image of A.)

When these axioms and rules are added to the usual propositional calculus a number of interesting theorems can be proved. In fact, Hamblin could prove that "there are just 30 distinct tenses", which can be formed using only P, F and negation.

Prior defined G ('is always going to be') as $\sim\! F \!\sim$, and H ('has always been') as $\sim\! P \!\sim$. Using this formalism Hamblin and Prior studied the implicative structure of the tenses given Hamblin's axiomatic system. In 1965 they ended up with the nice implicative structure for the tense-operators shown in Figure 2, which according to Hamblin is "a bit like a bird's nest" (see [Øhrstrøm and Hasle, 1995, p. 178]).

This system is obviously much stronger than the minimal tense logic K_t. It may be said to correspond to an intuitive idea of the structure of time. Prior and his followers in tense logic presented several other axiomatic systems. We shall comment on some of them in our paper on the history and philosophy of temporal logic after Prior (elsewhere in this volume).

In addition to the four basic tense operators, Prior also found it useful to introduce metrical tense operators, $F(x)$ (corresponding to 'in x time units it will be the case that') and $P(x)$ (corresponding to 'x time units ago it was the case that').

In 1967 Prior published his major work, *Past, Present and Future*, in which his approach to tense logic had reached a very convincing form. It turned out that several interesting tense logical systems could be established. Some of these systems incorporate not only tense operators but also an independent modal operator, $\diamond$ (corresponding to 'possibility'). Later, Prior even considered a logic integrating an operator, I, standing for 'the present' (now see [Prior, 2003, pp. 171–93]).

5   THE SEMANTICS OF TEMPO-MODAL LOGIC

According to Peter Geach, Prior regarded his own research into the logic of ordinary language constructions as a continuation of the medieval tradition [Geach, 1970, p. 188]. His attitude was congenial to that of the young Russell in *Principles of Mathematics*: ordinary language is not a logician's master, but it must be his guide [Geach, 1970, p. 187]. After all logic in Prior's opinion "is not primarily about language, but about the real world" [Prior, 1996b, TR]. For this reason he strongly opposed the formalistic view on logic:

> Formalism, i.e. the theory that logic is just about symbols and not about things, is false [Copeland, 1996, p. 45].

> I cannot see how any statement whatever can be made true simply by using language in a particular way, except, of course, the statement that we are using language in the way in question, and nobody would contend that a statement to this effect would be logically true — it is not logically necessary that we should speak in such and such a way [Prior, 1976c, p. 123].

Prior adopted the Stoic view of logic according to which the logic of propositions is basic, and according to which "the rest of logic is built upon it" [1955b, p. 3]. In short, his own answer to the question about the nature of logic ran as follows:

> Logic deals, at bottom, with statements — it enquires into what statements follow from what — but logicians aren't entirely agreed as to what a statement is. Ancient and medieval logicians thought of a statement as something that can be true at one time and false at another [Copeland, 1996, p. 47].

But what does it mean that a statement is true (or false)? In other to answer this question Prior worked out important theories of truth. In their excellent analysis of some of Prior's fundamental ideas of 'truth', Hugly and Sayward have distinguished between four categories of sentences in which the word 'true' is used [1996, p. 333]:

C1:   Sentences of the form 'It is true that S',

C2:   Sentences of the form 'The proposition (statement, belief) that S is true',

C3:   Sentences in which 'true' is applied independently of sentences specifying what is true, and not as a predicate of sentences or utterances, e.g. 'Some of Bill's beliefs are true',

C4:   Sentences in which 'true' is predicated of linguistic items, e.g. 'Bill's utterance is true'.

According to Hugly and Sayward, four theses corresponding to these four categories comprehend Prior's theory of truth [Hugly and Sayward, 1996, p. 389]. The first one says that in category 1 sentences, 'true' functions as a connective and not as a predicate. The second thesis says that the connective is null. The third thesis says that 'true' is analysable in terms of that connective in category 2 sentences and in category 3 sentences. According to the fourth thesis, 'true' is not analysable in terms of 'it is true that' in category 4 sentences.

In particular, Prior was interested in modal logic, and in consequence he wanted to explain what it means for a proposition in modal logic to be true. This interest led him to the very first formulation of the answer which is now normally given, i.e. the answer in terms of accessibility between possible worlds. In fact, already in 1951 he had suggested to deal with modal logic using 'state-descriptions' (see [Copeland, 1996, p. 11]). A few years later, he showed how tense logic can be studied using instants as state-descriptions, which are ordered by an earlier-later relation. Together with Carew Meredith, these ideas were later further developed, and they were thereby led to the significant invention of the possible world semantics (see [Copeland, 1996, p. 8 ff.]. In 1956 Prior and Meredith wrote up a brief joint paper entitled "Interpretations of Different Modal Logics in the 'Property Calculus' " [Meredith and Prior, 1956]. The paper was circulated in mimeograph form, and it contained the essential elements of the possible worlds semantics for propositional modal logic. It seems that Jack Copeland [2002] is right in holding that in this paper a binary relation appeared for the first time as an accessibility-like interpretation of the relation in an explicitly modal context. In this paper the authors do not suggest any philosophical explanation of the relation or of the related object. Nevertheless, there can be no doubt that they had a relation between possible worlds in mind. As Jack Copeland has pointed out, Meredith in a letter to Prior dated 10 October 1956 in fact uses the term 'possible world' and Prior in 'Computations and Speculations' [Meredith and Prior, Unpublished, p. 119] used the same term. Later Prior wrote:

I remember . . . C.A. Meredith remarking in 1956 that he thought the only genuine individuals were 'worlds', i.e. propositions expressing

total world-states, as in the opening of Wittgenstein's *Tractatus* ('The world is everything that is the case') [Prior, 2003, p. 219].

In order to introduce a logic of instants or dates, we need a set $TIME$ of instants (or dates) with a relation, $<$, which attributes to $TIME$ some structure. The relation '$<$' is called the before-after-relation. For any temporal instant t and any statement p, $T(t, p)$ is a new statement, which can be read 'p is true at t'. It is assumed that

(T1) $T(t, p \wedge q) \equiv (T(t, p) \wedge T(t, q))$

(T2) $T(t, \sim p) \equiv \sim T(t, p)$.

Note that in principle we should make a distinction between two kinds of conjunction (and also between two kinds of negation) in (T1–2). The reason is that p and q are treated as propositional functions rather than full-fledged propositions such as $T(t, p)$. This means that the two kinds of expressions would be of different types. On the other hand, it is also possible to put both types of expressions syntactically on a par, as we shall see in the next section. So we shall neglect this complication, since it is after all rather clear how the conjunctions, negations etc. should be read in each case.

Now, the definitions

(DF) $T(t, Fp) \equiv_{def} \exists t_1 : (t < t_1 \wedge T(t_1, p))$

(DP) $T(t, Pp) \equiv_{def} \exists t_1 : (t_1 < t \wedge T(t_1, p))$

would allow us to evaluate any tense logical formula p, in terms of $T(t, p)$. From the definitions $Hp \equiv_{def} \sim P \sim p$ and $Gp \equiv_{def} \sim F \sim p$ it immediately follows

(DG) $T(t, Gp) \equiv_{def} \forall t_1 : (t < t_1 \supset T(t_1, p))$

(DH) $T(t, Hp) \equiv_{def} \forall t_1 : (t_1 < t \supset T(t_1, p))$.

We shall say that a structure $(TIME, <, T)$ is an *instant-logical structure*, if T satisfies (T1–2) and the definitions (DF), (DP), (DG), and (DH). T is called the T-*operator* (or the valuation operator) of the structure.

Using an idea communicated to him from Saul Kripke in 1958 (see [Øhrstrøm and Hasle, 1995, p. 189], Prior showed that important differences between some of the systems can be illustrated graphically. Hamblin's system corresponds to a linear notion of time, whereas other systems presuppose a notion of branching time.

Prior discussed three different models of branching time. The main difference between these models has to do with the status of the future. The models fall into a small number of groups, where the basic ideas can be shown in a very intuitive way: consider once again the old Aristotelian example about the possible sea-fight tomorrow. How should we define truth for statements like $F(1)p$?

One particular line of answer to this question can be based on a simple but radical assumption, namely the rejection of the principle of bivalence. This may give rise to some serious formal problems as well as some highly counter-intuitive features. For instance, if $F(1)p$ and $\sim F(1)p$ are both 'indeterminate' (or 'undefined'), it is very hard to explain how statements like the conjunction $F(1)p \wedge \sim F(1)p$ and the disjunction $F(1)p \vee \sim F(1)p$ can be anything else than 'indeterminate' (or
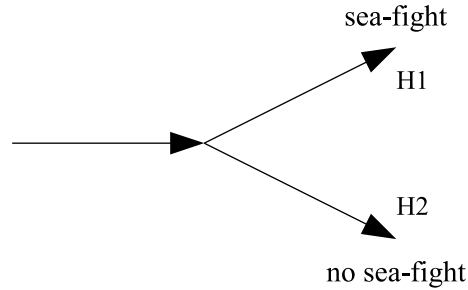
Figure 3. Branching time in the K_b system

'undefined') [Prior, 1967b, p. 135]. Prior came to believe that the introduction of 'indeterminate' or 'undefined' statements is an unnecessary complication. For this reason he in his later writings left aside solutions based on the rejection of bivalence and concentrated on bivalent answers. For the sake of simplicity, we shall use metrical time in our examples; but the results can be generalised into non-metrical tense-logic.

Let us consider three ways (a, b, and c below) of defining truth for statements like $F(1)p$:

> (a) The first answer is that the two possibilities, sea-fight and no sea-fight, are both part of the future, and that none of them has any superior status relative to the other. This answer can be represented graphically as in Figure 3.

The arrows on the ends of the two future branches indicate that the statements 'there is going to be a sea-battle (tomorrow)' and 'there is not going to be a sea-battle (tomorrow)' are both true in this picture of branching time. That is, if we let p stand for 'there is a sea-battle going on', and $F(1)p$ stand for 'there is going to be a sea-battle tomorrow', then

$$F(1)p \wedge F(1){\sim}p$$

is true. The corresponding tense-logical system is called K_b after Saul Kripke. We shall comment on this systems in more details in our paper on the history and philosophy of temporal logic after Prior (elsewhere in this volume).

> (b) Prior named the Ockham-model named after William of Ockham (c. 1285–1349), who in his logic had insisted that God knows the truth-value of every future contingent statement. According to this model only one possible future is the true one, although we as human beings do not know which of them it is. Let us assume that there is in fact going to be no sea-fight tomorrow. In this case the future should be represented graphically in the following way, where a line not ending in

sea-fight

H1

H2

no sea-fight
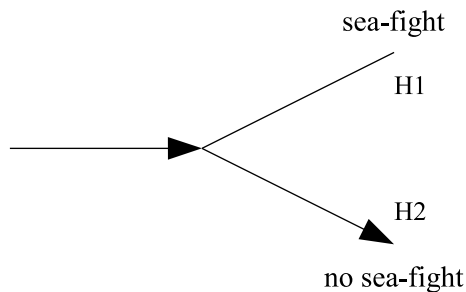
Figure 4. Branching time in the Ockham model

sea-fight

H1

H2

no sea-fight
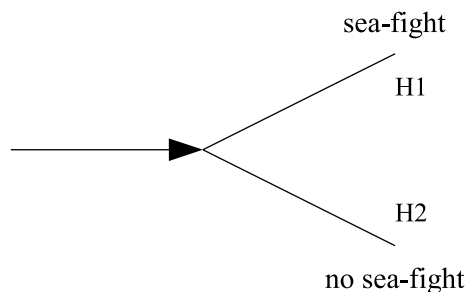
Figure 5. Branching time in the Peirce model

an arrow indicates that it will be false to assert that the corresponding
state-of-affairs will be the case tomorrow (see Figure 4).

So, $\sim F(1)p \land F(1)\sim p$ is the true description of this situation, even though
we may be unable to know this at the present moment (p etc. being defined as
above).

(c) Prior named the Peirce-model after Charles Sanders Peirce (1839–
1914). According to this model — which Prior himself adopted as
covering his own view — it makes no sense to speak about the true
future as one of the possible futures. There is no future yet, just a
number of possibilities. Hence, the future — or perhaps rather, the
'hypothetical future' — should be represented graphically as in Figure
5.

Neither $F(1)p$ nor $F(1)\sim p$ are true on this picture. However, if some proposi-
tion q holds tomorrow in all possible futures — that is, if the truth of q tomorrow
is regarded as necessary — then $F(1)q$ is true. In order to describe the semantics
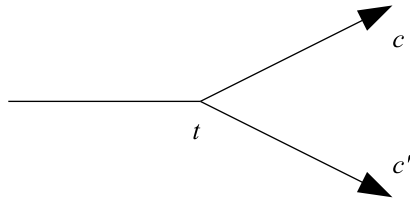
Figure 6. Branching time with chronicles

for these tempo-modal systems Prior [1967b, p. 126 ff.] needs a notion of temporal 'routes' or 'temporal branches' i.e. maximally ordered (i.e. linear) subsets in $(TIME, C, <, =)$. We prefer the term 'chronicle'. The set of all such chronicles will be called C (see Figure 6).

An Ockhamistic valuation operator, Ock, can be defined in the structure $(TIME, C, <, =)$, where $<$ is transitive and backwards linear. Given a truth-value for any propositional constant at any moment in $TIME$, $Ock(t, c, p)$ can be defined recursively for any moment in any chronicle, $t \in c$:

      (a) $Ock(t, c, p \wedge q)$ iff both $Ock(t, c, p)$ and $Ock(t, c, q)$
      (b) $Ock(t, c, \sim p)$ iff not $Ock(t, c, p)$
      (c) $Ock(t, c, Fp)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $t < t'$
      (d) $Ock(t, c, Pp)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $t' < t$
      (e) $Ock(t, c, \Box p)$ iff $Ock(t, c', p)$ for all c' with $t \in c'$.

$Ock(t, c, p)$ can be read 'p is true at t in the chronicle c'. A formula p is said to be Ockham-valid if and only if $Ock(t, c, p)$ for any t in any c in a branching time structure, $(TIME, C, <, =)$.

It may be doubted whether Prior's Ockhamistic system is in fact an adequate representation of the tense logical ideas propagated by William of Ockham. According to Ockham, God knows the contingent future, so it seems that he would accept an idea of absolute truth, also when regarding a statement Fq about the contingent future — and not only what Prior has called "prima-facie assignments" [1967b, p. 126] like $Ock(t, c, Fq)$. That is, such a proposition can be made true 'by fiat' simply by constructing a concrete structure which satisfies it. But Ockham would accept that Fq could be true at t without being relativised to any chronicle. And that actually brings us back to a two-place T-operator, like the ones we have previously discussed. In [Øhrstrøm and Hasle, 1995] we have shown that it is possible to establish a system which seems to be a bit closer to Ockham's original ideas. On the other hand, it should be noted that the question concerning the notion of truth is mainly philosophical. Prior's Ockhamistic system appears to comprehend at least all the theorems which should be accepted according to

Ockham's original ideas. Let us, for instance, consider one tense logical formula:
$$q \supset HFq.$$

It is obvious from the above definitions that $Ock(t, c, q \supset HFq)$ for any t and any c with $t \in c$. Therefore $q \supset HFq$ is a theorem in Prior's Ockhamistic system.

Now, let us turn to the Peirce system. In this system the truth-operator differs from the Ockhamistic operator when it comes to the evaluation of propositions on the form, Fp. In this case the Peircean truth-operator can be defined in the following way:

> $Peirce(t, Fp)$ iff
> for all c' with $t \in c'$:
> $Peirce(t', p)$ for some $t' \in c'$ with $t < t'$.

Prior put forward this tense logical system on the basis of his studies of Peirce's philosophy. He described the system in the following way:

> ... C.S. Peirce's description of the past (with, of course the present) as the region of the 'actual', the area of 'brute fact', and the future as the region of the necessary and the possible. That is why I call this system 'Peircean' [Prior, 1967b, p. 132].

There is hardly any doubt that Prior's rendition of Peirce's ambitions as regards the logic of time and modality is correct. By analysing Peirce's way of thinking and transferring this into the modern logic of time, Prior found that in the Peircean system the following formula must hold for any proposition p:
$$\sim(F(x)p \wedge F(x)\sim p),$$
whereas its 'excluded middle' analogue
$$F(x)p \vee F(x)\sim p$$
does not hold in general. — This is due to the fact that both assertions, $F(x)p$ and $F(x)\sim p$, can be false, if they represent a pair of statements about the contingent future. It turns out that in the Peircean system $F(x)p$ and $\Box F(x)p$ are equivalent. It is also obvious that in this system, $q \supset HFq$ does not hold in general.

The discussion regarding the Ockhamistic versus the Peircean system was crucial for Prior in his attempts to deal with philosophical arguments in favour of determinism. His careful analyses of these systems were, however, not his only contribution to the further development of tense-logic. In fact, he studied a number of tense-logical systems corresponding to various notions of time (for instance, dense time, circular time, discrete time). He dealt with many of his findings in the paper, "Recent Advances in Tense Logic", which was published shortly after his death in 1969 [Prior, 1969].

6   A- AND B-SERIES: FOUR GRADES OF TENSE-LOGICAL INVOLVEMENT

It was Peter Geach who sometime in the early 1960s made Prior aware of the importance and relevance of McTaggart's distinction between the so-called A-

and B-series conceptions of time [1967b, p. vi]. Since then, the notions and arguments in McTaggart's paper, "The Unreality of Time", [1908], have become a necessary ingredient of all major treatments of the philosophical problems related to temporal logic.

McTaggart's A-series conception is based on the notions of past, present, and future, as opposed to a 'tapestry' view of time, as embodied by the B-series conception of time. Prior later formally elaborated McTaggart's distinction, and showed that we can discuss time using either a tense logic, corresponding to the A-series conception, or using an earlier-later calculus, corresponding to the B-series conception. Prior's interest in McTaggarts observations was first aroused when he realised that McTaggart had offered an argument to the effect that the B-series presupposes the A-series rather than vice versa [1967b, p. 2]. Prior was particularly concerned with McTaggart's argument against the reality of tenses. Prior's studies brought renewed fame to this argument. In consequence, it has been very important in the philosophical debate about various kinds of temporal logic and their mutual relations. In our chapter on modern temporal logic (in this volume) we discuss the structure of McTaggart's argument and the philosophical debate to which it has given rise.

As we shall see in the chapter on modern temporal logic (in this volume) Prior rejected McTaggart's conclusion, and he held that the temporal world should in fact be described in terms of tenses (i.e. McTaggart's A-series). In his view, the alternative description of temporality in terms of earlier-later (i.e. McTaggart's B-series) was secondary. Prior clearly considered this tense-logical view (i.e. the A-series) to be the fundamental one when it comes to the study of time. On the other hand, Prior clearly found that the relations between the A-series and the B-series are crucial when it comes to a deeper understanding of logic and time. In his studies of the relations between the A-series and the B-series, Prior introduced four grades of 'tense logical involvement'.

The first grade defines tenses entirely in terms of objective instants and an earlier-later relation. For instance, a sentence such as Fp, 'it will be the case that p', is defined as a short-hand for 'there exists some instant t which is later than now, and p is true at t', and similarly for the past tense; these definitions are, of course,

$$(DF) \quad T(t, Fp) \equiv_{def} \exists t_1 : t < t_1 \wedge T(t_1, p)$$
$$(DP) \quad T(t, Pp) \equiv_{def} \exists t_1 : t_1 < t \wedge T(t_1, p).$$

Tenses, then, can be considered as mere meta-linguistic abbreviations, so this is the lowest grade of tense logical involvement. Prior succinctly described the first grade as follows:

> ... there is a nice economy about it ... it reduces the minimal tense logic to a by-product of the introduction of four definitions into an ordinary first-order theory, and richer [tense logical] systems to by-products of conditions imposed on a relation in that theory [Prior, 2003, p. 119–20].

In the first grade, tense operators are simply a handy way of summarizing the properties of the before-after relations, which constitute the B-theory. Hence, in the first grade B-theory concepts are seen to be determining for a proper understanding of time and reality; tenses are deemed to have no independent epistemological status. The basic idea is a definition of truth relative to temporal instants:

(T1) $T(t, p \wedge q) \equiv (T(t, p) \wedge T(t, q))$

(T2) $T(t, \sim p) \equiv \sim T(t, p)$.

In addition, there may be some specified properties of the before-after relation, like for instance transitivity:

(B1) $(t_1 < t_2 \wedge t_2 < t_3) \supset t_1 < t_3$.

In this way, instants acquire an independent ontological status. As we have seen, Prior rejected the idea of temporal instants as something primitive and objective.

In the second grade of tense logical involvement, tenses are not reduced into B-series notions. Rather, they are treated on a par with the earlier-later relation. Specifically, a bare proposition p is treated as a syntactically full-fledged proposition, on a par with propositions such as $T(t, p)$ ('it is true at time t that p'). The point of the second grade is that a bare proposition with no explicit temporal reference is not to be viewed as an incomplete proposition. One consequence of this is that an expression such as $T(t, T(t', p))$ is also well-formed, and of the same type as $T(t, p)$ and p. Prior showed how such a system leads to a number of theses, which relates tense logic to the earlier-later calculus and vice versa [Prior, 2003, p. 121]. The following crucial rule of inference makes this relation within the second grade especially obvious:

(RT) If $\vdash p$, then $\vdash T(t, p)$ for any t and any truth-operator T.

He also stated the following basic assumptions regarding the truth-operator:

(TX1) $(\forall t : T(t, p)) \supset p$

(TX2) $(\forall t_1 : T(t_1, p)) \supset T(t_2, \forall t_3 : T(t_3, p))$

(TX3) $T(t_1, p) \supset T(t_2, T(t_1, p))$.

The philosophical implication of this second grade of tense logical involvement is that one must regard the basic A- and B-theory concepts as being on the same conceptual level. Neither set of concepts is conditioned by the other.

The B-theory is sometimes considered as the semantics of the corresponding A-theory. This is not surprising if we again consider the first-grade formulation of Fp, 'it will be the case that p', as a short-hand for 'there exists some instant t which is later than now, and p is true at t' (cf. (DF)).

This is tantamount to stating a truth condition for Fp. On this view of the relationship between the A- and B-theories, it may be a bit puzzling that p and $T(t, p)$ can be treated as being on the same logical level — the former apparently belonging to the logical language (or object language) and the latter to the semantics (or meta-language). In Prior's opinion, however, this is not at all surprising. In a paper on some problems of self-reference he stated:

> In other words, a language can contain its own semantics, that is to say its own theory of meaning, provided that this semantics contains

the law that for any sentence x, x means that x is true [Prior, 1976b, p. 141].

It seems that this statement is exemplified exactly by the relation of the logic of tenses (the A-theory) to the logic of earlier and later (the B-theory), provided that we are willing to take the step of the second grade: syntactically conflating 'bare' p with $T(t, p)$.

The relation becomes even clearer in the third grade, a system which has crucial implications for the status of the indication of time. Prior introduced the third grade in the following way:

> What I shall call the third grade of tense logical involvement consists in treating the instant-variables a, b, c, etc. as representing propositions [Prior, 2003, p. 124].

Such instant-propositions describe the world uniquely at any given instant, and are for this reason also called world-state propositions. Like Prior we shall use a, b, c, ... as instant-propositions instead of t_1, t_2, t_3, ... In fact, Prior assumed that such propositions are what ought to be meant by 'instants':

> A world-state proposition in the tense-logical sense is simply an index of an instant; indeed, I would like to say that it is an instant, in the only sense in which 'instants' are not highly fictitious entities [Prior, 1967b, p. 188–189].

The traditional distinction between the description of the content and the indication of time for an event is thereby dissolved. From the properties of the logical language which embodies the third grade of tense logical involvement, Prior also showed that $T(a, p)$ can be defined in terms of a primitive necessity-operator. Then tense logic, and indeed, all of temporal logic can be developed from the purely 'modal notions' of past, present, future, and necessity.

In order to present the formalism of the third grade, Prior assumes the standard definitions of propositional and predicate logic, including the definition of $\exists a : \phi$ as $\sim\forall a : \sim\phi$. In the following, 'p' stands for an arbitrary well-formed formula in the system, whereas 'a' stands for an arbitrary instant proposition. The axioms of the system are the axioms of K_t together with the axiom

$\qquad$ (I1) $\quad \exists a : a$

and the rule:

$\qquad$ (RI) $\quad$ For any instant proposition a and any well-formed formula p: Exactly one of $\vdash a \supset p$ and $\vdash a \supset \sim p$ holds.

To this are added the rules included in Prior's quantification theory [Prior, 1955b, p. 76 ff.]:

$\qquad$ (Π1) $\quad$ If $\vdash \phi(x) \supset \beta$, then $\vdash (\forall x : \phi(x)) \supset \beta$.

$\qquad$ (Π2) $\quad$ If $\vdash \alpha \supset \phi(x)$, then $\vdash \alpha \supset \forall x : \phi(x)$, for x not free in α.

From (Π1–2) it is easy to deduce [1955b, p. 82] that

(Σ1)   If $\vdash \phi(x) \supset \beta$ then $\vdash (\exists x : \phi(x)) \supset \beta$, for x not free in β.

(Σ2)   If $\vdash \alpha \supset \phi(x)$ then $\vdash \alpha \supset \exists x : \phi(x)$.

It should be noted that (RI) is natural in the light of what it means to be a maximal consistent set. Intuitively, an instant proposition a may be viewed as the conjunction of the elements in the maximal consistent set. (I1) is also rather natural since it simply states that some instant proposition holds now. In addition, we assume the standard definitions from propositional and predicate modal logic, especially the definition of $\Diamond$ as $\sim\Box\sim$. The axiomatic system consists of the basic tense-logical system and the following axioms:

(L1) $\Box(p \supset q) \supset (\Box p \supset \Box q)$

(L2) $\Box p \supset p$

(L3) $\Box p \supset \Box\Box p$

(I2) $\sim\Box\sim a$

(I3) $\Box(a \supset p) \vee \Box(a \supset \sim p)$

(BF) $\Box(\forall a : \phi(a)) \equiv \forall a : \Box(\phi(a))$

(□G) $\Box p \supset Gp$

(□H) $\Box p \supset Hp$

along with the rule

(R□)   If $\vdash p$, then $\vdash \Box p$.

(L1), (L2), and (L3) are the Gödel postulates for (S5).

It is obvious that (RG) and (RH) follow from (R□), (□G), and (□H). (I2) means that any instant proposition should be regarded as possible. (I3) is in fact a consequence of (RI) together with the consistency and the maximality of a. (BF) is known as the Barcan formula after Ruth C. Barcan [1946], who was able to demonstrate it for modal logics which satisfy a few basic conditions. Now we want to construct a T-operator based on the full logic of instant propositions. That is, we wish to show how an entire earlier-later calculus can be developed — one might say boot-strapped — from definitions in the tense-logical theory.

Let W denote the set of instant propositions. For arbitrary elements a and b in W we introduce the following definitions:

(DB) $a < b \equiv_{def} \Box(a \supset Fb)$

corresponding to 'the instant a is earlier than the instant b', and

(DT) $T(a,p) \equiv_{def} \Box(a \supset p)$

corresponding to 'it is true at time a that p'. Using these assumptions and definitions we can prove the theorems (T1–2), as well as (DG) and (DH). In turn, this means that $(W, <, T)$ is a B-logical structure (with T defined as above). The following theses can also be proved in the system:

(DL) $\forall a : T(a,p) \equiv \Box p$

(DB) $\Box(b \supset Pa) \equiv a < b$

(TX1) $(\forall a : T(a,p)) \supset p$

(TX2) $(\forall a : T(a,p)) \supset T(b, \forall c : T(c,p))$

(TX3) $T(a,p) \supset T(b, T(a,p))$.

Taken together, these results show that the T defined above is a suitable T-operator. In this way formulae of the T-calculus are mixed with well-formed

formulae from the original tense-logical system. Everything is included in one single language comprising the T-calculus as well as ordinary tense logic. This extended language obviously includes a logic of instant propositions. This way of seeing things is far from the 'main-stream' tradition within formal logic, where the axiomatics of the tense logic is kept strictly separated from semantics (in this case the T-calculus). But as Prior pointed out there is nothing semantically wrong with it, if the T-calculus is given an interpretation within tense logic. He also pointed out that such an interpretation could be 'metalogically useful', since in many cases $T(a, p)$ turns out to be easier to prove than the 'bare' tense-logical formula p itself [Prior, 1967b, p. 89].

Prior has thus shown how we can in fact interpret B-logic within A-logic, namely in a given modal context in which we can interpret instants as propositions and quantify over them. In this sense B-logical semantics is absorbed within an entirely A-logical axiomatics. In Prior's own words, this means "to treat the first order theory of the earlier-later relation as a mere by-product of tense logic" [Prior, 2003, p. 273].

He developed this view even further in his fourth grade, in which he suggested a tense logical definition of the necessity-operator such that the only primitive operators in the theory are the two tense logical ones: P and F. Prior himself favoured this fourth grade. It appears that his reasons for wanting to reduce modality to tenses were mainly metaphysical, since it has to do with his rejection of the concept of the (one) true (but still unknown) future. If one accepts the fourth grade of tense-logical involvement, it will turn out that something like the Peirce solution will be natural, and that we have to reject solutions which involve crucially the idea of a true or simple future — like the Ockhamistic theory.

In our opinion this idea of treating instants as some kind of world propositions was one of Prior's most interesting constructions. We believe that the full strength of this view has not yet been demonstrated. It is very interesting that all the basic ideas and ingredients of modern *hybrid logic* are in fact present in Prior's logic. Hybrid logic is currently being further developed and also applied to still new problem domains. A bit more needs to be said on the idea of hybrid logic. In 1977, a post-humous Prior-volume titled *Worlds, Times and Selves* appeared, edited and completed by Kit Fine. Herein Prior's ideas on hybrid logic (even though the term itself was not used) were elaborated in various ways. Then work on the subject apparently ceased. However, attention was drawn to Prior's third grade in 1988 by Peter Øhrstrøm in [1988] and in 1991 by Per Hasle in [1991]. It must be admitted, however, that the potential of Prior's third grade was still not fully realised. Although Prior was the first logician who developed the idea that formulas can be used as terms, it should be noted that the idea of hybrid languages was explored independently by the Sofia School in the mid-1980s (see http://www.hylo.net/). However, a good deal of the honour for the last decade's development of hybrid logic must be accorded to Patrick Blackburn, who in no small part gave a spark to its development in [1993] and [1994]. We ourselves further analysed Prior's ideas in [Øhrstrøm and Hasle, 1995], whereas Blackburn

and others worked independently on hybrid logic at the same time, and since. Efforts in the field seem now to converge.

7 CONCLUSION

Prior dealt with many problems within philosophical logic, and it was very important for him to view logic as strongly related to reality. He firmly rejected formalism, i.e. the theory that logic is just about symbols and not about things. He held that logic "is not primarily about language, but about the real world" [Copeland, 1996, p. 45]. In his opinion only the present exists [Prior, 1972]. In the same way as only one possible world is real ("the actual world"), Prior maintained that only one instant is real ("the present"). In this way, the tenses (past, present, and future) are essential for the understanding of reality. Prior stated:

> So far, then, as I have anything that you could call a philosophical creed, its first article is this: I believe in the reality of the distinction between past, present, and future. I believe that what we see as a progress of events is a progress of events, a coming to pass of one thing after another, and not just a timeless tapestry with everything stuck there for good and all [Copeland, 1996, p. 47].

Following this view, Prior stressed that "the tense of a statement must be taken seriously" [Copeland, 1996, p. 48]. He insisted that this idea should be taken into account in any attempt to understand reality. In fact, he held that tense logic is important not only in philosophy, but also in metaphysics and in physics (see [Øhrstrøm and Hasle, 1995, p. 197 ff.]). As is evident from *Past, Present and Future* and several of his other writings, Prior was very interested in the tense-logical formulation of relativistic physics. He argued that the physicist should understand that tense-logical questions ought to be taken into serious consideration in the development of relativistic physics and other parts of the natural sciences dealing with time. However, he never claimed that questions within physics can be answered only using tense logic, but he maintained that logic, in fact, can be applied to the study of nature. He said:

> The logician must be rather like a lawyer — not in Toulmin's sense, that of reasoning less rigorously than a mathematician — but in the sense that he is there to give the metaphysician, perhaps even the physicist, the tense logic that he wants, provided that it be consistent. He must tell his client what the consequences of a given choice will be ... and what alternatives are open to him; but I doubt whether he can, qua logician, do more [1967b, p. 59].

During the last years of his life Prior became very interesting in the logical aspect of the notion of the 'self' and in what he called 'Egocentric Logic'. In fact, a significant formal part of the book *Worlds, Times and Selves* (which Kit Fine

edited and completed in 1976) consists in developing the egocentric counterpart to ordinary tense or modal logic, whose crucial feature is the operator Q "that picks out those propositions that correspond to instants, worlds or selves, as the case may be" [Prior and Fine, 1977, p. 8].

Prior's most important achievement was his establishment of temporal logic as a research field within philosophical logic. He was indeed the founding father of modern temporal logic.

In Prior's view temporal logic should be conceived of as an important tool for anyone who wants to study the concept of time. In fact, the choice between the four grades is a choice between four different theories of time.

After a lecture which was in fact just one in a series of lectures on temporal logic, probably held somewhere in USA, Prior wrote the following addition to the paper which he was going to read at the next lecture in the series:

> A [a person present at the lecture] wants me to relativise my tenses to dates. It seems to me that behind this request there is a metaphysics. Behind this request there is the idea that the whole of time is absolutely there with all these dates, and all events and processes just are, located in various parts of this giant fixed frame. I do not believe this. I think this way is to treat all time as if it were already past. I don't believe this. I don't believe that events and processes are; rather events happen (and then come to have happened) and processes go on (and then come to have gone on), and even this is an abstraction — the basic reality is things acting. But even in this flux there is a pattern, and this pattern I try to trace with my tense-logic; and it is because this pattern exists that men have been able to construct their seemingly timeless frame of dates. Dates, like classes, are a wonderful and tremendously useful invention, but they are an invention; the reality is things acting [Prior, Unpublished a, p. 1].

Prior expressed his own theory in the following way:

> Time is not an object, but whatever is real exists and acts in time ... But this earlier-later calculus is only a convenient but indirect way of expressing truths that are not really about 'events' but about things ... [Copeland, 1996, p. 45].

He initiated a number of interesting studies within this new field and he clearly demonstrated that temporal logic can be understood as having fundamental relations to essential problems in physics, philosophy, and theology. He even seems to have realised that temporal logic could turn out to be very useful within computer science. In the chapter on modern temporal logic (in this volume) we shall discuss the further development in the field, which Prior founded in the 1950s and the 1960s.

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

[Barcan, 1946] Ruth C. Barcan. A functional calculus of first order based on strict implication. *Journal of Symbolic Logic*, xi:2, 1946.

[Becker, 1960] O. Becker. Zur Rekonstruktion des Kyrieuon Logos des Diodorus Kronos (mit besonderer Rücksicht auf die Arbeiten von A.N. Prior). 1960. In [Derbolav and Nicolin, 1960].

[Blackburn, 1993] Patrick Blackburn. Nominal tense logic. *Notre Dame Journal of Formal Logic*, 34(1):56–83, 1993.

[Blackburn, 1994] Patrick Blackburn. Tense, temporal reference, and tense logic. *Journal of Semantics*, 11:83–101, 1994.

[Borkowski, 1970] L. Borkowski, editor. *Jan Łukasiewicz: Selected Works*. North-Holland, Amsterdam, 1970.

[Brown and Koch, 1991] Charles Grant Brown and Gregers Koch, editors. *Natural Language Understanding and Logic Programming*. North-Holland, Amsterdam, 1991.

[Copeland, 1996] Jack Copeland, editor. *Logic and Reality: Essays in the Legacy of Arthur Prior*. Oxford University Press–Clarendon Press, Oxford, 1996.

[Copeland, 2002] Jack Copeland. The genesis of possible worlds semantics. *Journal of Philosophical Logic*, 31:99–137, 2002.

[Derbolav and Nicolin, 1960] J. Derbolav and F. Nicolin, editors. *Erkenntnis und Verantwortung: Festschrift für Theodor Litt*. Schwann, Düsseldorf, 1960.

[Edwards, 1967] Paul Edwards, editor. *The Encyclopedia of Philosophy*, volume 4. Collier-Macmillan, London, 1967.

[Faye et al., 1997] Jan Faye, Uwe Scheffler, and Max Urchs, editors. *Perspectives on Time*, volume 189 of *Boston Studies in the Philosophy of Science*. Kluwer Academic Publishers, 1997.

[Findlay, 1941] J.N. Findlay. Time: A treatment of some puzzles. *Australasian Journal of Psychology and Philosophy*, 19:216–35, 1941.

[Flew and MacIntyre, 1955] Antony Flew and Alasdair MacIntyre, editors. *New Essays in Philosophical Theology*. S.C.M. Press, London, 1955.

[Fraser et al., 1972] J.T. Fraser, F.C. Haber, and G.H. Müller, editors. *The Study of Time*, volume I. Springer-Verlag, 1972.

[Geach, 1970] P.T. Geach. Arthur Prior: A personal impression. *Theoria*, 3:186–188, 1970.

[Hasle, 1991] P. Hasle. Building a temporal logic for natural language understanding with the HOL-system. 1991. In [Brown and Koch, 1991, pp. 92–110].

[Hasle, 1999] P. Hasle. The problem of predestination — as a prelude to A.N. Prior's tense logic. 1999. In [Wegener, 1999, pp. 139–159].

[Hasle, 2003] P. Hasle. Life and work of Arthur N. Prior: An interview with Mary Prior. 2003. In [Prior, 2003, pp. 293–310].

[Hugly and Sayward, 1996] Philip Hugly and Charles Sayward. *Intensionality and Truth. An Essay on the Philosophy of A.N. Prior*. Kluwer Academic Publishers, 1996.

[Kenny, 1970] Anthony Kenny. Arthur Norman Prior (1914–1969). *Proceedings of the British Academy*, LVI:321–349, 1970.

[Lewis, 1943] C.S. Lewis. *The Abolition of Man*. Harper Collins Publishers, 1943.

[Łukasiewicz, 1920] Jan Łukasiewicz. On three-valued logic. 1920. Reprinted in [Borkowski, 1970].

[Łukasiewicz, 1930] Jan Łukasiewicz. Philosophical remarks on many-valued systems of propositional logic. 1930. Reprinted in [Borkowski, 1970].

[Mates, 1949] Benson Mates. Diodorean implication. *Philosophical Review*, 58:234–244, 1949.

[Mates, 1961] Benson Mates. *Stoic Logic*. University of California Press, 1961.

[McTaggart, 1908] J.M.E. McTaggart. The unreality of time. *Mind*, pages 457–474, 1908.

[Meredith and Prior, 1956] Carew Meredith and Arthur Prior. Interpretations of different modal logics in the 'property calculus', 1956. Originally published in mimeograph form by University of Canterbury. Published in [Copeland, 1996, pp. 133–134].

[Meredith and Prior, Unpublished] C.A. Meredith and A.N. Prior. Computations and speculations. MS, Bodleian Library, Box 8, Unpublished.

[Øhrstrøm and Hasle, 1995] P. Øhrstrøm and P. Hasle. *Temporal Logic — from Ancient Ideas to Artificial Intelligence*. Kluwer Academic Publishers, Dordrecht, 1995.

[Øhrstrøm, 1988] P. Øhrstrøm. *Nogle aspekter af tidsbegrebets rolle i de eksakte videnskaber med særligt henblik på logikken*. Aalborg University Press, Aalborg, 1988.

[Prior and Fine, 1977] A.N. Prior and Kit Fine. *Worlds, Times and Selves*. University of Massachusetts Press–Duckworth, Amherst, MA–London, 1977. Based on manuscipts by Prior. Edited by Kit Fine with a preface and a postscript by Kit Fine.

[Prior, 1940a] A.N. Prior. A Calvinist romantic. *Purpose*, 12(1):15–21, 1940.

[Prior, 1940b] A.N. Prior. Makers of modern thought (1): Kierkegaard. *The Student Movement*, pages 131–132, March 1940.

[Prior, 1940c] A.N. Prior. A Scot seeks God. *The Churchman*, pages 34–42, 1940.

[Prior, 1942] A.N. Prior. Can religion be discussed? *Australasian Journal of Psychology and Philosophy*, 20:141–151, 1942. Reprinted in [Flew and MacIntyre, 1955, pp. 1–11].

[Prior, 1946] A.N. Prior. The reformers reformed: Knox on predestination. *The Presbyter*, 4:19–23, 1946.

[Prior, 1947] A.N. Prior. Supralapsarianism. *The Presbyter*, 5:19–22, 1947.

[Prior, 1948a] A.N. Prior. Adam Gib and the philosophers. *Australasian Journal of Psychology and Philosophy*, 26:73–93, 1948.

[Prior, 1948b] A.N. Prior. Disruption. *Landfall*, 2:8–18, 1948.

[Prior, 1949] A.N. Prior. *Logic and the Basis of Ethics*. Oxford University Press, 1949.

[Prior, 1951] A.N. Prior. The ethical copula. *Australasian Journal of Philosophy*, 29:137–154, 1951. Republished in [Prior, 1976b, pp. 9–24].

[Prior, 1952] A.N. Prior. Łukasiewicz's symbolic logic. *Australasian Journal of Philosophy*, 30:121–130, 1952.

[Prior, 1953] A.N. Prior. Three-valued logic and future contingents. *The Philosophical Quarterly*, 3:317–326, 1953.

[Prior, 1955a] A.N. Prior. Diodoran modalities. *The Philosophical Quarterly*, 5:205–213, 1955.

[Prior, 1955b] A.N. Prior. *Formal Logic*. Clarendon Press, Oxford, 1955.

[Prior, 1955c] A.N. Prior. Speaking about God. *The Student*, (5):7–10, September 1955.

[Prior, 1957a] A.N. Prior. Symbolism and analogy. *The Listener*, April 25 1957.

[Prior, 1957b] A.N. Prior. *Time and Modality*. Oxford University Press, Oxford, 1957.

[Prior, 1958] A.N. Prior. 'The good life and religious faith' (East-West meeting at Canberra Dec. 1957). *Australasian Journal of Philosophy*, 36:1–13, 1958.

[Prior, 1959a] A.N. Prior. Creation in science and theology. *Southern Stars*, 18:82–89, 1959.

[Prior, 1959b] A.N. Prior. Thank goodness that's over. *Philosophy*, 34(128):11–17, January 1959.

[Prior, 1962a] A.N. Prior. The formalities of omniscience. *Philosophy*, 37:114–129, 1962. Republished in [Prior, 2003, pp. 39–58].

[Prior, 1962b] A.N. Prior. Logic in England today. Bodleian Library Box 5, 16 pages. English original of [Prior, 1962c], 1962.

[Prior, 1962c] A.N. Prior. 'Wspolczesca logica w Anglii' (Contemporary logic in England). *Ruch filozoficzny*, 21:251–256, 1962.

[Prior, 1965 1966] A.N. Prior. Time, existence and identity. *Proceedings of the Aristotelian Society*, 66:183–192, 1965–1966. Reprinted in [Prior, 2003, pp. 93–101].

[Prior, 1967a] A.N. Prior. 'Logic, history of':'Peirce. C.S.'. 1967. In [Edwards, 1967, pp. 546–549].

[Prior, 1967b] A.N. Prior. *Past, Present and Future*. Oxford University Press, Oxford, 1967.

[Prior, 1968] A.N. Prior. *Papers on Time and Tense*. Oxford University Press, Oxford, 1968.

[Prior, 1969] A.N. Prior. Recent advances in tense logic. *The Monist*, 53(3):325–339, 1969.

[Prior, 1972] A.N. Prior. The notion of the present. 1972. In [Fraser *et al.*, 1972, pp. 320–323].

[Prior, 1976a] A.N. Prior. *The Doctrine of Propositions and Terms*. Duckworth, London, 1976. Edited by P.T. Geach and A.J.P. Kenny.

[Prior, 1976b] A.N. Prior. *Papers in Logic and Ethics*. University of Massachusett Press, Amherst, MA, 1976. Edited by P.T. Geach and A.J.P. Kenny.

[Prior, 1976c] A.N. Prior. What is logic? 1976. In [Prior, 1976b, pp. 122–129].

[Prior, 1996a] A.N. Prior. Some free thinking about time, 1996. Published in P. Øhrstrøm's edition in [Copeland, 1996, pp. 45–46]. [SFTT].

[Prior, 1996b] A.N. Prior. A statement of temporal realism, 1996. Published in P. Øhrstrøm's edition in [Copeland, 1996, pp. 47–51]. [TR].

[Prior, 2003] A.N. Prior. *Papers on Time and Tense*. Oxford University Press, Oxford, second revised edition, 2003. Edited by Per Hasle, Peter Øhrstrøm, Torben Braüner, and Jack Copeland. First edition [Prior, 1968].

[Prior, Unpublished a] A.N. Prior. Determinism in philosophy and theology. MS, 4 p. Bodleian Library, Box 6, Unpublished.

[Prior, Unpublished b] A.N. Prior. The fable of the four preachers. Undated MS. Bodleian Library, Oxford, Box 6, Unpublished.

[Prior, Unpublished c] A.N. Prior. The logic of obligation and the obligation of the logician. Undated MS, Bodleian Library, Oxford, Box 6, Unpublished.

[Prior, Unpublished d] A.N. Prior. Logical criticisms of the theory identifying duty with self-interest. Undated MS. Bodleian Library, Oxford, Box 9, Unpublished.

[Prior, Unpublished e] A.N. Prior. A modernist stocktaking. MS, 6 p. Bodleian Library, Box 7, Unpublished.

[Prior, Unpublished f] A.N. Prior. On Frege and Kolakowski. Untitled MS, Bodleian Library, Box 6, Unpublished.

[Sedley, 1977] David Sedley. Diodorus Cronus and Hellenistic philosophy. *Proceedings of the Cambridge Philology Society*, 203:74–120, 1977.

[Wegener and Øhrstrøm, 1997] Mogens Wegener and Peter Øhrstrøm. A new tempo-modal logic for emerging truth. 1997. In [Faye *et al.*, 1997, pp. 417–441].

[Wegener, 1999] Mogens Wegener, editor. *Time, Creation and World-Order*. Aarhus University Press, Aarhus, 1999.

[Wright, 1951] Henrik von Wright. *An Essay in Modal Logic*. North-Holland, Amsterdam, 1951.

MODERN TEMPORAL LOGIC:
THE PHILOSOPHICAL BACKGROUND

Peter Øhrstrøm and Per Hasle

1 INTRODUCTION

Inspired by Kantian thinking, the Irish mathematician William Rowan Hamilton found that just as geometry can be understood as a pure mathematical study of space, a similar pure mathematical study of time ought to exist. The research programme emerging from this conviction can be described as an attempt to establish algebra as the 'science of pure time'. Hamilton encountered many difficulties in that endeavour. In fact, there are several indications that he actually gave up the fundamental idea himself [Øhrstrøm, 1985].

Another kind of algebraic approach to the study of time was carried out by George Boole (1815–1864). He was probably the first 19th century logician to include the concept of time explicitly in his theories of logic and reasoning (although only in a few passages). Some of his interesting considerations regarding the relation between time and logic can be found in the manuscript entitled *Sketch of a Theory and Method of Probabilities Founded upon the Calculus of Logic*, which Boole seems to have written between 1848 and 1854. Boole here used symbols x, y, z corresponding to elementary propositions such as 'The Thermometer falls' and 'It will rain'. In fact he regarded "the symbols as representing the times in which the elementary propositions to which they refer are true" [Boole, 1953, p. 146]. Boole obviously held that a proposition refers to one or more durations. If two propositions refer to the duration x and the duration y, then the conjunction between two propositions, xy, corresponds to the intersection between the two durations. Boole regarded the numerical constant 0 as "the representative of the nothing of time or never" and the constant 1 as representing "the Universal of time" [Boole, 1953, p. 146].

Charles Sanders Peirce (1839–1914) found the algebraic approach to time insufficient. In his *New Elements of Mathematics*, he specifically rejected Hamilton's programme making the following observation:

> Hamilton called algebra the Science of Time. But the most remarkable characteristic of time, namely that the passage from the past to the future is qualitatively different from the passage from the future to the past is not represented in algebra [Peirce, 1976, p. 9].

But how can we find an appropriate alternative to the algebraic approach? In particular: How can the temporal asymmetry between the past and the future be incorporated in a system of symbolic logic in a satisfactory manner? Peirce was certainly aware of the difficulties to which the incorporation of time within logic would give rise. He was, however, certain that such difficulties could in principle be overcome although the problem is difficult to solve. Peirce wrote:

> Time has usually been considered by logicians to be what is called 'extra-logical' matter. I have never shared this opinion. But I have thought that logic had not yet reached the state of development at which the introduction of temporal modifications of its forms would not result in great confusion; and I am much of that way of thinking yet [1931 1958, 4.523].

As it is described in another chapter in this volume, modern temporal logic was first shaped by A.N. Prior (1914–1969) as a detailed construction within philosophical and symbolic logic. Since Prior, logicians in general have considered 'temporal logic' to be a rather well established notion. However, we may as well realise from the beginning that the term 'temporal logic' is not so easily delimited. Firstly, 'temporal logic' is inseparable from a study of the modalities *possibility* and *necessity*, as was indeed signalled by the very title of the work which founded modern 'temporal logic', namely A.N. Prior's *Time and Modality* from [1957]. Moreover, the philosophical issues of free will and determinism versus indeterminism are obviously and inevitably related to questions concerning time and modality. Secondly, 'temporal logic' also can be seen as a position within the Philosophy of Logic. It was Prior's view that, properly understood, all of logic is really temporal, and that logical languages without some kind of temporal operators were really devoted just to a proper subset of logic. Thirdly, 'temporal logic' is studied as well as applied within other fields, especially within Computer Science and Logical Linguistics. Such studies also bring and have brought some results of direct importance in philosophical logic. With these caveats in mind, we shall now look at 'temporal logic' as a branch of philosophical logic.

Prior primarily wanted to clarify a number of conceptual relations regarding temporal notions and to contribute to the solution of some important philosophical problems concerning the nature of time (including some rather existential questions regarding human life). In so doing, Prior formulated a number of logical systems which were later studied in more detail and also further elaborated by several mathematicians and computer scientists who in many cases apparently did not know very much about the philosophical background of their enterprise. However, some writers in modern temporal logic obviously have been aware of the philosophical background of temporal logic, and they have in many cases contributed significantly to further clarifying the philosophical problems in question. The aim of this chapter is neither to describe the mathematical development of modern temporal logic, nor to describe the study of temporal logic in relation to computer science. For presentations of such technical issues we refer to [Gabbay *et*

al., 2000]. Rather, it is our intention in this chapter to discuss the continuous work in modern temporal logic focusing on the philosophical problems which originally inspired A.N. Prior in his pioneering work within temporal logic. This does not mean that the technical results are irrelevant for the philosophical investigation. On the contrary, many technical results turn out to be very important and indeed essential for the work with the philosophical problems which temporal logic was originally designed to treat. However, our focus will be on the philosophical motivation and the various conceptual aspects of the formalisms of modern temporal logic. Construed in this manner, temporal logic turns out to be a very nice illustration of how philosophy and mathematical logic can both benefit from a constructive symbiosis.

The strength of philosophical logic lies in its self-imposed obligation to take everyday language and common sense reasoning into serious consideration. For this reason it is natural that the first detailed theory of tenses developed in philosophical logic was based on a study of the grammatical tenses of natural language (see [Reichenbach, 1947]). However, neither everyday language nor common sense reasoning are unambiguous quantities. They certainly greatly depend on physical and metaphysical assumptions.

In his theory Hans Reichenbach (1891–1953) suggested a three-point structure for tenses [Reichenbach, 1947]. However, as we shall see in section 2, A.N. Prior clearly demonstrated that Reichenbach's theory fails to solve important problems concerning temporal notions in a satisfactory manner. On the other hand, it is also clear that Prior's work with Reichenbach's and other early contributions to the study of tenses was useful in the development of his own theory of time and tense.

Reichenbach's theory of a three-point structure for tenses was not the first contribution to analysis of time and tense in the 20th century. In his famous paper, *The Unreality of Time*, from [1908] J.M.E. McTaggart (1866–1925) offered an early discussion of time and tense. However, McTaggart's ideas had no significant role to play before Prior published his analysis of the paradox, and it would be misleading to see these early ideas as a proper theory of tenses. As we shall see in section 3, the analysis of McTaggart's so-called paradox became very important in the philosophical debate about time in the 1970s and later. In particular, the debate has turned out to be crucial when it comes to an understanding of the relations between time and tense.

In section 4 we shall discuss some of the tense logical systems which Prior and his followers suggested. Some of the systems will be presented as axiomatic systems and others will be introduced referring to semantical models. The relations between these two approaches will also be discussed.

In section 5 we are going to discuss the logic of future contingency. There can be little doubt that this theme is the most famous problem within the philosophical logic of time and tense.

During the 1960s, tense logic became a well established subject. However, some very interesting results obtained by Hans Kamp questioned the conceptual foundation of tense logic. According to Kamp's results the notions of 'since' and 'until' might be seen as even more fundamental than Prior's tense operators (past, present, and future). Prior was nevertheless able to defend his view. In section 6 we shall consider the some of the essential points in this debate.

Some of the critics of the Priorean approach to temporal logic maintained that the notion of instants conceived as durationless instants has to be rejected as far from reality and truth. On the contrary, they have held that temporal logic should be constructed as a durational logic according to which propositions are not true or false at instants but according to which propositions are true or false over various durations in time. In section 7 we shall present some basic ideas in such durational logics.

In section 8 we shall turn to the relation between temporal logic and physics. One very common criticism of Priorean tense-logic has been based on various interpretations of the special and the general theories of relativity. Several writers have argued that Priorean tense-logic contradicts the findings of these physical theories, and that the basic tense-logical position for this reason has to be rejected. Others have maintained that a real contradiction does not necessarily arise.

In the section 9 we are going to discuss various ideas regarding the incorporation of the notions of agency and time. We shall discuss some modern attempts at creating a theoretical integration of the notions of knowledge, obligation, and time.

2 AN EARLY THEORY OF TENSES

In his *Elements of Symbolic Logic* [1947], Hans Reichenbach suggested a description of tenses which was to have a significant impact. Reichenbach advocated the view that in order to understand how tenses work we must consider not only the time of utterance, and the time of the event in question, but also a 'point of reference'. It would be fair to say that this is the first detailed theory of tenses formulated in modern logic and philosophy. It should be added, however, that according to Reichenbach himself, [1947, p. 290] the idea of a three-point structure for tenses had already been suggested by the great Danish linguist Otto Jespersen (1860–1943). But Reichenbach certainly elaborated the idea in much detail, and as we have argued elsewhere there are significant differences between Jespersen's ideas and Reichenbach's detailed theory of tenses [Øhrstrøm and Hasle, 1995, pp. 158 ff.].

To understand the idea of this three-fold distinction, it is probably best first to consider the future perfect, as in 'I shall have seen John'. This sentence clearly speaks of a certain event, namely 'my seeing John'; but it is also clear that it directs us to a future time *different* from the time of the (expected) event — namely a time prior to which the event has already occurred. Thus, we must distinguish between the time of the event and the time to which the sentence refers. Reichenbach called

the former 'point of the event' and the latter 'point of reference', symbolised by E and R, respectively. Furthermore, both must of course be determined with respect to the time of utterance, the 'point of speech' S.

Armed with these distinctions Reichenbach could give the following diagram for the future perfect:

Future Perfect
I shall have seen John

```
  |   |   |
──┴───┴───┴──▶
  S   E   R
```
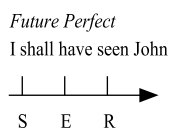
A quite similar analysis can be given for the past perfect 'I had seen John'. These two tenses, then — the past perfect and the future perfect — establish the *prima facie* case for distinguishing between E, S, and R in the description of tenses. However, if the difference between E and R is crucial in explaining the past perfect and the future perfect, it is precisely the *coincidence* between one or more of E, R, and S, which is crucial in explaining some of the other tenses. Indeed, what particularly impressed linguists was the elegant and concise account of the difference between the simple past and the present perfect which Reichenbach could give on the basis of the three-fold distinction.

In grammars of English, six tenses are standardly recognised; the diagrams for each of these can be seen in this figure (cf. [Reichenbach, 1947, p. 290]):

Future Perfect　　　　　*Simple Future*
I shall have seen John　　I shall see John

```
  |   |   |                 |       |
──┴───┴───┴──▶          ──┴───────┴──▶
  S   E   R                S, R    E
```

Past Perfect　　　　　　*Simple Past*
I had seen John　　　　　I saw John

```
  |   |   |                 |       |
──┴───┴───┴──▶          ──┴───────┴──▶
  E   R   S                R, E    S
```

Present Perfect　　　　*Present*
I have seen John　　　　I see John

```
  |       |                    |
──┴───────┴──▶          ──────┴──────▶
  E     S, R                 S, R, E
```
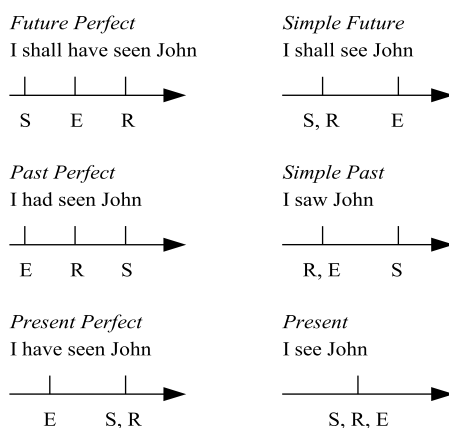
On this account, the crucial difference between the simple past and the present perfect is determined by the relative 'position' of the reference point. In the case of the simple past, the diagram clearly suggests that the point of reference coincides with the point of the event. Thus the sentence 'I saw John' clearly refers to the past, but it makes no discernible distinction between the time of the event — E — and the time from which this event is seen, i.e. the reference time R. In the case

of the present perfect, the event is also situated in the past, but here, the point of reference coincides with the point of speech.

Reichenbach's system makes a rather strong prediction about the notion of tenses, logically as well as grammatically. If tenses are in general to be construed as a three-point structure, the possible arrangements of this kind of structure must exhaust the set of possible tenses. In principle, Reichenbach's systematisation allows for 13 different tenses; he only regarded nine of these as significantly different (see [Reichenbach, 1947, p. 296]).

The fact that Reichenbach considered the relative positions of E and S as basically irrelevant explains a slight oddity about his diagram for the future perfect. The sentence 'I shall have seen John' would also seem to be true even if the speaker has in mind an event which has already occurred — that is, the structure would be E—S—R (this is perhaps a less natural reading, but quite possible). However, according to Reichenbach there is no important difference between E—S—R and S—E—R. Indeed, in summing up the possible tenses he explicitly aligns

$$S—E—R$$
$$S, E—R$$
$$E—S—R$$

under the common heading of 'future perfect'. A similar account is given for R—E—S, R—S—E, and R—S, E, which he collects under the heading 'posterior past'. None of the six traditional tenses corresponds to posterior past, but it can be stated by some transcription, as in 'I was to see John once more' or 'the letter was to cause her great anxiety'.

For all its intuitive elegance, it is clear that Reichenbach's formalism is very limited. It is certainly not a complete calculus, but at best it could be seen as a suggestion of some guidelines along which such a system could be constructed. However, even when measured on its own terms the system harbours severe difficulties.

Reichenbach makes a sharp distinction between 'point of reference' and 'point of event'. This is the fundamental idea on which the general viability of Reichenbach's systematisation rests — as well as its accounts of the individual tenses. One who clearly saw this was Prior, who in [1967] discussed the precursors of tense logic. Herein he gave Reichenbach some credit for his observations, but then went on to state that "Reichenbach's scheme, however, will not do as it stands; it is at once too simple and too complicated" [1967, p. 13]. The main target of Prior's attack was exactly the sharp distinction between 'point of reference' and 'point of event'. Consider a complicated future tense like this one:

'I shall have been going to see John'.

This sentence is perhaps not very natural, but it is grammatically correct, and it does express a tense-relation for which we must be able to account. It is not too hard to see that to describe this tense, we in fact need two points of reference. Prior's 'Reichenbachian' diagram for this case looks like this:
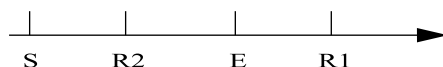
So, for such a tense the Reichenbachian framework would have to be extended to allow for two points of reference; and in general, an arbitrary number of 'reference points' might be needed. Prior could therefore observe that

> ... once this possibility is seen, it becomes unnecessary and misleading to make such a sharp distinction between the point or points of reference and the point of speech; the point of speech is just the *first* point of reference. (This, no doubt, destroys Reichenbach's way of distinguishing the simple past and the present perfect; but that distinction needs more subtle machinery in any case.) [1967, p. 13]

It is crucial for Reichenbach's system that three points of time should always be taken into consideration. But we have just seen that this may sometimes be too little; and, as the quotation also suggests, it is sometimes too much. For in the account of, say, the simple past — in terms of an R, E—S diagram, where R = E — why should we accept that there is really more than two temporal indicators involved? And even more so, why should we accept such a thing for the present S, R, E (where S=R=E)? Only cogent logico-linguistic reasons should make one accept that there are three temporal indicators at play in these cases. But referring to the fact that Reichenbach's account apparently explains the difference between the simple past and the present perfect is at best circumstantial evidence; for it *explains* this difference only if the distinctions are valid beforehand.

Incidentally, these observations also show that the Reichenbach framework really ought to distinguish between on one hand the *temporal indicators* — or *concepts* — of 'event', 'reference' and 'speech', and on the other hand the *points of time* which they 'indicate'. Thus for instance, if the event E occurs at t, we might say that $\tau(E) = t$. Only thus can a diagram like

$$\tau(R),\tau(E)—\tau(S)$$

make a meaningful distinction between more than two indicators. Here, R and E are co-extensive with respect to their time-parameter, but they must be assumed to be intensionally different (i.e. $\tau(R) = \tau(E)$, but E $\neq$R).

Reichenbach was a brilliant mind, and many of his results — also on the philosophy of time — have had lasting value. Fairness demands that this be acknowledged, and in the case of his 'three-point structure' it must at least be admitted that for its day it was an elegant and advanced proposal. But its real deficiencies together with its very success made it counter-productive — Prior considered Reichenbach's work in this respect to be an impediment rather than a help in the development of tense logic.

3 MCTAGGART'S PARADOX (A- AND B-SERIES)

The distinction between the logic of tenses and the logic of earlier and later (in terms of instants or in terms of durations) is essential for the understanding of modern temporal logic. This distinction was introduced by J.M.E. McTaggart in his famous paper, *The Unreality of Time* [1908]. In this paper McTaggart suggested the distinction between the so-called A- and B-series conceptions of time. According to the A-series conception, the tenses (past, present, and future) are the key notions for a proper understanding of time, whereas the earlier-later calculus is secondary. According to the B-series conception time is understood as a set of instants organized by the earlier-later relation, whereas the tenses are secondary.

As mentioned in our chapter on A.N. Prior's logic elsewhere in this volume, the founder of temporal logic became very interested in the writings of McTaggart, in particular when he realised that McTaggart had offered an argument to the effect that the B-series presupposes the A-series rather than vice versa. Prior was particularly concerned with McTaggart's argument against the reality of tenses.

McTaggart's A-series conception is based on the notions of past, present, and future, as opposed to a 'tapestry' view on time, as embodied by the B-series conception of time. He explicitly identified the dichotomy between the A-series and the B-series. He himself arrived at the conclusion that A-concepts are more fundamental than B-concepts. He did not, however, use this analysis as an argument in favour of A-theory. On the contrary, he used it for a refutation of the reality of time! He argued that A-concepts give rise to a contradiction — which has become known as 'McTaggart's Paradox'. Due to this putative contradiction within the fundamental conceptualisation of time, he went on to claim that time is not real.

The core of McTaggart's argument is that the notions of 'past', 'present' and 'future' are predicates applicable to events. The three predicates are supposed to be mutually exclusive — any concrete event happens just once (even though a *type* of event may be repeated). On the other hand, any of the three predicates can be applied to any event. In a book on history, it makes sense to speak of 'the death of Queen Anne' as a past event — call it e_1 — but in a document written in the lifetime of Queen Anne, it could well make sense to speak about her death as a future event. Apparently this gives rise to an inconsistency, since how can e_1 be both past and future — and present as well, by a similar argument? The answer must be that there is another event e_2, relative to which for instance e_1 has been present and future, and is going to be past. Now, the same kind of apparent inconsistency can be established with respect to e_2, and the problem can only be solved by introducing a new event e_3, for which a new apparent inconsistency will arise etc. — which seems to mean that we have to go ad infinitum in order to solve the inconsistency. The consequence appears to be that the inconsistency can never be resolved.

Prior, however, pointed out a basic flaw in McTaggart's argument. According to his view, the contradictions arise from an attempt at forcing the A-series notions

into a B-series framework [1967, p. 6]. Prior argued that events may be described in terms of instant-propositions, of which it also holds that they 'happen', i.e. are true, exactly once. Using a as an arbitrary instant proposition, the claim that the three tense-logical predicates are mutually exclusive can be formulated as:

$a \supset (\sim Pa \wedge \sim Fa)$
$Pa \supset (\sim a \wedge \sim Fa)$
$Fa \supset (\sim a \wedge \sim Pa)$.

Here Pa stands for 'it has been the case that a', whereas Fa stands for 'it will be the case that a'. The fact that any event can be past, present, and future, can be expressed in the following way, where the I-operator stands for 'the present':

$Ia \supset (PFa \wedge FPa)$
$Pa \supset (PIa \wedge PFa)$
$Fa \supset (FPa \wedge FIa)$.

But no contradiction follows from these 6 theses. It is thus revealed that McTaggart's paradox is in no way a cogent argument against the A-series notions, let alone the reality of time. Prior concluded that McTaggart's argument could not shake his fundamental belief in the ontological status of the tenses. Prior maintained that tense logic embodied a crucial ontological and epistemological point of view according to which "the tenses (it will be, it was the case) are primitive; only present objects exist" [Prior and Fine, 1977, p. 116]. To Prior, the present and the real were one and the same concept. Shortly before he died, he formulated his view in the following way:

> ... the present simply is the real considered in relation to two particular species of unreality, namely past and future [Prior, 1972, p. 320].

During the 20th century there has been much debate concerning the validity of McTaggart's argument and various reformulations of it. Some authors like David Mellor have maintained that there is a valid version of the argument, which should in fact force us to reject the tense-logical view of time, i.e. the A-series conception. According to Mellor, nothing in reality has tenses and "the A-series is disproved by a contradiction inherent in the idea that tenses change" [Mellor, 1981, p. 89]. Others have followed Prior in holding that all versions of McTaggart's argument are flawed. In his careful analysis of McTaggart's Paradox, William Lane Craig [2000, p. 169 ff.] has argued that no contradiction need be involved in a proper formalization of the A-series, and it may be concluded that McTaggart's argument is simply misdirected as a refutation of the tensed theory of time [Craig, 2000, p. 207].

As mentioned above, McTaggart's paradox can be solved if iterated tenses like PF and FP are introduced. It may be seen as part of McTaggart's argument that in this way we shall need still longer iterated tenses (like PPF, $FFPF$, $PPFFP$, ...) in order to solve the apparent contradiction, and we thereby have to deal with the problems of an infinite regress of this kind. It is, however, not obvious that any serious logical problem would follow from such an infinite regress. In addition,

as Richard Sylvan [1996, p. 122] has argued, the construction of iterated tenses in response to McTaggart's argument will probably not give rise to a proper infinite regress since "*expressivewise* the regress stops" as a consequence of the logical properties of the tense logic in question. The point is, that it is likely to be the case in all relevant systems that the number of non-equivalent (iterated) tenses is finite. This statement may be seen as an generalisation of Hamblin's famous fifteen-theorem for dense time (see [Øhrstrøm and Hasle, 1995, p. 176 ff.]).

It should be mentioned that the A-series versus B-series discussion has been somewhat "de-dramatised" within 'temporal logic' over the past few decades, probably because the development within 'temporal logic' and so-called hybrid logic (see [Blackburn *et al.*, 2001]) has made it perfectly possible for the two basic sets of notions to co-exist within one and the same language (as already suggested). Nevertheless, from a philosophical point of view this question concerning the "nature" of time is equally important today. Moreover, since we approach 'temporal logic' as a branch of *philosophical* logic, it is of paramount importance to identify the underlying assumptions, respectively the possible philosophical import of our formalisms — and these two ways of expressing temporal relations are what constitutes 'temporal logic'. In this context it is worth noting that the very development which has de-dramatised the difference between the two approaches — and in particular the development of hybrid logic — was in fact initiated and founded by A.N. Prior in [1968] and [2003, ch. XI] (and anticipated in [1967]) in order to show that a B-theory could be embedded within an A-theoretical language and hence that it was possible to maintain the primacy of A-theory, should one so wish (as Prior indeed did).

It seems clear from the extensive debate that a valid version of McTaggart's argument can only be established if some extra-philosophical assumptions are made. These additional assumption can all be questioned, but none of them represent *a priori* impossible positions. For this reason, it may be concluded that it is still logically possible to hold any of the two main positions. In fact, as Prior has argued, various relevant variations of the positions should be taken into consideration. As explained in the chapter on Prior's logic, he suggested a distinction between four possible grades of tense-logical involvement corresponding to four different views of how to relate the A-notions (past, present and future) to the B-notions ('earlier than', 'later than', 'simultaneous with'):

1. The B-notions are more fundamental than the A-notions. Therefore, in principle the A-notions have to be defined in terms of the B-notions.

2. The B-notions are just as fundamental as the A-notions. The A-notions cannot be defined in a satisfactory manner in terms of the B-notions (and vice versa). The two sets of notions have to be treated on a par.

3. The A-notions are more fundamental than the B-notions. There is also a primitive and fundamental notion of (temporal) possibility. In principle the B-notions have to be defined in terms of the A-notions and the primitive notion of temporal possibility.

    4. The A-notions are more fundamental than the B-notions. In principle the
       B-notions have to be defined in terms of the A-notions. Even the notion of
       temporal possibility can be defined on terms of the A-notions.

Understood in this way, it is obvious that Prior's four grades of tense-logical involvement (see [Øhrstrøm and Hasle, 1995, p. 176 ff.]) represent four different views of time and also four different foundations of temporal logic.

    The problem we address in the debate rooted in McTaggart's paradox and the relation between the A- and B-notions is clearly related to the problem of truth in temporal logic. What makes a statement like, 'It is now four o'clock', true or false? As Poidevin and MacBeath [1993, p. 2] have clearly described in their account of modern philosophy of time, this question can be answered in two different ways. The A-theorists say that the statement "It is now four o'clock" is true if and only if the time we have given the name "four o'clock", is in fact present. The B-theorists, on the other hand, claim that there are no tensed facts. According to their view the statement "It is now four o'clock" is true if and only if it is the case that the utterance is made at four o'clock. Similarly, the A-theorists claim that the statement "Julius Caesar was killed" is true because Julius Caesar was in fact killed, whereas the B-theorists say that this statement is true because the time of utterance is after the death of Julius Caesar. In this way the A-theorists hold that tensed statements have tensed truth-conditions, while the B-theorist find that tensed sentences are made true or false by tenseless truth-conditions. In their book, Poidevin and MacBeath [1993] have presented A.N. Prior and D.H. Mellor as prominent representatives of respectively the A- and the B-view.

    It may be useful to consider the formal aspects of the A- and B-notions a little closer. In order to do so we first of all have to deal with the general features of the tense-logical formulae which are essential for the formulation of the A-series conception. These formulae can be introduced inductively by the following rules of well formed formulae (wff):

   (i)   any propositional variable is a wff
  (ii) if ϕ is a wff, then $\sim\phi$ is also a wff
 (iii) if ϕ and φ are wffs, then $(\phi \wedge \varphi)$ is also a wff
 (iv) if ϕ is a wff, then $F\phi$ is also a wff
  (v) if ϕ is a wff, then $P\phi$ is also a wff
 (vi) nothing else is a wff.

From here an A-theorist would probably like to add a formalism of instant propositions. The B-theorists, on the other hand, would probably emphasise the need for truth-conditions established in terms of a model $M = (TIME, <, v)$, where $TIME$ is a set of temporal elements like instants or durations, $<$ is a binary relation on $TIME$ (corresponding to 'before'), and v is a valuation function from the cross product of $TIME$ and the set of propositional variables to $\{0, 1\}$. The expression $v(t, p)$ is said to be the truth-value of the propositional variable p at t. Given such a model the notion of truth for any tense-logical formula can be given by the following inductive definition:

      $M, t \models p$ if $v(t, p) = 1$

$M, t \models \sim\phi$ if not $M, t \models \phi$
$M, t \models (\phi \wedge \varphi)$ if $M, t \models \phi$ and $M, t \models \varphi$
$M, t \models F\phi$ if $M, t' \models \phi$ for some t' with $t < t'$
$M, t \models P\phi$ if $M, t' \models \phi$ for some t' with $t' < t$.

If $M, t \models \phi$ the proposition ϕ is true at t according to the model M. The B-theorist will emphasise that in this way truth of the tense-logical formulae of the object language is defined in terms of a tenseless metalanguage. For this reason, the B-theorist will point out that the A-language clearly depends on the B-language.

Obviously, the A-theorist has to follow another line of argumentation. It seems that at least two options are open for him. The first possibility was explicitly formulated by A.N. Prior, according to whom there is no sharp distinction between an object language and a metalanguage. Using what is now called a hybrid logic in which the instants are just a special kind of propositions Prior was able to define $T(t, \phi)$ (standing for 'ϕ is true at t') for any tenselogical formula, ϕ, in terms of the tenselogical language itself (see [Braüner, 2002a]). The second possibility for the A-theorist is the use of so-called homophonic theories of truth in which the constructions of the object language are interpreted in terms of analogous constructions of the metalanguage. Torben Braüner [2002a] has demonstrated that tense logics permit the existence of such a homophonic theory of truth, provided that they are stronger than the rather basic tense-logical system K_b (one of the systems with which we shall deal in next section). As pointed out by Torben Braüner [2002b], A.N. Prior himself was clearly aware of the possibility of a homophonic theory of truth as it is evident from the following quotation:

> The function of the operator F, in short, is that of forming a future-tense statement from the corresponding present-tense one, and the future-tense statement is not about the present tense one, but is about whatever the present-tense statement is about. ... But although the statement 'It will be the case that Professor Carnap is flying to the moon', that is, 'Professor Carnap will be flying to the moon', is not exactly a statement *about* the statement 'Professor Carnap is flying to the moon', we may say that the future-tense statement is true if and only if the present-tense statement *will be* true [Prior, 1957, pp. 8–9].

However, it seems that Prior never attempted to work out the details of a homophonic theory of truth.

In addition to the various logical approaches to McTaggart's argument, there may be alternative perspectives which may give rise to other kinds of considerations involving new ideas of time and reality. Thus Kit Fine [2005] has suggested a modernised approach to McTaggarts argument. According to Fine the argument can be reconstructed as being based on the following four assumptions:

Realism: Reality is constituted (at least, in part) by tensed facts.

Neutrality: No time is privileged, the tensed facts that constitute reality are not oriented towards one time as opposed to another.

Absolutism: The constitution of reality is an absolute matter, i.e. not relative to a time or other form of temporal standpoint.

Coherence: Reality is not contradictory, it is not constituted by facts with incompatible content.

The argument states that these assumptions, when taken together, lead to inconsistency. The B-theorist will of course have no problems in rejecting the above version of realism (i.e. the idea of reality constituted by tensed facts). The standard A-theorist will following Prior reject the neutrality assumption. In his paper Kit Fine [2005] has explored the possibilities of giving up either 'absolutism' or 'coherence'. In both cases we will be left with a rather complicated temporal logic. For this reason, it seems obvious to investigate Kit Fine's reasons for saying that the A-theorist should accept 'neutrality'. It appears that the main reason has to do with problems formulated on the basis of the special theory of relativity. We are going to deal with these problems in section 8.

4   THE LOGIC OF TIME AND TENSE

Many logicians and philosophers dealing with the concept of time have concentrated on the study of the features of time conceived as linear structure. They have understood time as an ordered set of instants, and, as we shall see in section 7, sometimes also as a corresponding set of partially ordered durations (or intervals). Conceived as a system of instants time is viewed as an ordered set $(TIME, =, <)$, where $TIME$ is a set of instants, and where $=$ and $<$ are binary relations on $TIME$ corresponding to identity and before/after. A number of interesting properties of this structure may be considered:

(Z1) $\forall x \in TIME : \sim(x < x)$ (irreflexivity)

(Z2) $\forall x, y \in TIME : x < y \supset \sim(y < x)$ (asymmetry)

(Z3) $\forall x, y, z \in TIME : (x < y \wedge y < z) \supset x < z$ (transitivity)

(Z4) $\forall x \in TIME : \exists y \in TIME : x < y$ (non-ending)

(Z5) $\forall x \in TIME : \exists y \in TIME : y < x$ (non-beginning)

(Z6) $\forall x, y \in TIME : \exists z \in TIME : x < y \supset (x < z \wedge z < y)$
                                                 (density)

(Z7) $\forall x, y, z \in TIME : (x < z \wedge y < z) \supset (x < y \vee y = x \vee y < x)$
                                    (backwards linearity)

(Z8) $\forall x, y, z \in TIME : (z < x \wedge z < y) \supset (x < y \vee y = x \vee y < x)$
                                    (forwards linearity)

(Z9) $\forall x, y \in TIME : (x < y \vee y = x \vee y < x)$ (connectedness).

It is possible to formulate many other possible properties of this kind (see [Burgess, 1984; Benthem, 1991; Rescher and Urquhart, 1971]). It is an open question which

of these many properties we should actually accept in our description of the structure of time. It is in fact very likely that this question cannot be answered definitively, since the answer may depend on the context and the purpose of the description.

Based on such a structure of instants $(TIME, =, <)$ we may introduce the idea of truth at an instant, $T(t, p)$ (read: p is true at t), where t is an instant in $TIME$, and p is a proposition from ordinary propositional logic.

(T1) $T(t, \sim p)$ iff not $T(t, p)$
(T2) $T(t, p \wedge q)$ iff $T(t, p)$ & $T(t, q)$

Here we have made a distinction between two kinds of conjunction i.e. '$\wedge$' in the object language and '&' in the meta-language. Similarly, there is a difference between two kinds of negations, '$\sim$' and 'not'. In the following we shall ignore this difference, since it will always be obvious how the formulae should be understood.

In order to introduce tenses, we use Prior's symbols and define Pq (i.e. 'it has been the case that q') and Fq (i.e. 'it will be the case that q') as

(T3) $T(t, Pq)$ iff $\exists s \in TIME : s < t \wedge T(s, q)$
(T4) $T(t, Fq)$ iff $\exists s \in TIME : t < s \wedge T(s, q)$.

Defining Gq (i.e. 'it will always be the case that q') as $\sim F \sim q$ and Hq (i.e. 'it has always been the case that q') as $\sim P \sim q$ we find

(T5) $T(t, Gq)$ iff $\forall s \in TIME : t < s \supset T(s, q)$
(T6) $T(t, Hq)$ iff $\forall s \in TIME : s < t \supset T(s, q)$.

With these definitions and a number of properties valid for the structure of instants we may study the logic of tenses as a by-product of the logic of 'truth at an instant'.

However, all this presentation is mainly a B-logical approach to temporal logic. A proper A-logical approach would start with the study of tenses. For this reason Prior in many of his writings concentrated on the study of tense-logical systems.

Any A-logic, i.e. tense logic, is based on the primitive tense-operators P and F; its axiomatisation is often formulated in terms of the derived operators H and G (as we have pointed out earlier, H and G are inter-definable with P and F, respectively, so either pair of operators can in fact be chosen as primitives). A very fundamental system has been named K_t (where the 'K' is probably in honour of Saul Kripke). This tense logic can be presented as an axiomatic system with the following axiom schemes [Prior, 1967, p. 176]; [McArthur, 1976, p. 17 ff.]:

(A1) p, where p is a tautology of the propositional calculus
(A2) $G(p \supset q) \supset (Gp \supset Gq)$
(A3) $H(p \supset q) \supset (Hp \supset Hq)$
(A4) $p \supset HFp$
(A5) $p \supset GPp$.

In (A2)–(A5), p and q are arbitrary, well-formed formulas. All axioms are said to be immediately provable, while other theses can be proved by inference. In K_t, *Modus Ponens* is the basic rule of inference:

(RMP) If $\vdash p$ and $\vdash p \supset q$, then $\vdash q$.

In addition we have two rules, which introduce tense-operators:

(RG) If $\vdash p$, then $\vdash Gp$.
(RH) If $\vdash p$, then $\vdash Hp$.

From K_t, other tense logical systems can be defined by adding more axioms to the above list, (A1–A5), as we shall see in the following.

It is easy to verify that the axioms (A1–A5) are true at any t for any model $M = (TIME, <, v)$, and that the same holds for all wff's which can be proved in K_t. This means that the system is sound. However, it can be demonstrated that the opposite (i.e. that the system is complete) also holds i.e. if a wff is true at any t for any model $M = (TIME, <, v)$, then it is also provable in K_t (see [Benthem, 1991, p. 165 ff.]. Soundness and completeness of the system K_t can be summarised in the following way:

$K_t \vdash \phi$ if and only if $M, t' \models \phi$ for any model $M = (TIME, <, v)$,
and any $t \in TIME$.

Several other tense-logical systems have been studied (see [Rescher and Urquhart, 1971], [McArthur, 1976], [Benthem, 1991], [Øhrstrøm and Hasle, 1995]). If we to K_t add the axioms

(A6) $FFp \supset Fp$
(A7) $FPp \supset (Pp \lor p \lor Fp)$

we obtain the system K_b. It is interesting that it possible to prove the 'mirror image' of (A6) within this system (see [McArthur, 1976, p. 26]) i.e.

$K_b \vdash PPp \supset Pp$.

It can be demonstrated to be sound and complete with respect to all models $(TIME, =, <, v)$ with transitivity (Z3) and backwards linearity (Z7) (see [Rescher and Urquhart, 1971, p. 74 ff.]). For this reason K_b is understood as a system of branching time i.e. the systems allows for alternative futures, but not for alternative pasts.

Normally, the axioms of K_b are presented with the following axiom instead of (A7) which is more directly than (A7) appealing to the idea of backwards linearity (i.e. no 'alternative pasts'):

(A7x) $(Pp \land Pq) \supset (P(p \land q) \lor P(p \land Pq) \lor P(Pp \land q))$.

It is easy to prove (A7) from (A1–A6)+(A7x). However, it is also possible to demonstrate (A7x) from (A1–A7). The proof can be found in [Øhrstrøm and Hasle, 1995, p. 207 ff.] and its essential ideas in fact can be traced back to A.N. Prior's pioneering work in tense logic.

The system, K_l which corresponds to linear time can be obtain by the addition of the following axiom to K_b:

(A8) $PFp \supset (Pp \lor p \lor Fp)$.

Similar to what is said about (A7) above, it should be mentioned that (A8) works just as well as a system with the following axiom in its place:

$$(A8x) \quad (Fp \wedge Fq) \supset (F(p \wedge q) \vee F(p \wedge Fq) \vee F(Fp \wedge q)).$$

In order to obtain a tense logical system, $K_{ld\infty}$, corresponding to non-beginning, non-ending, dense linear time, we need the following axioms:

$$(A9) \quad Gp \supset Fp$$
$$(A10) \quad Hp \supset Pp$$
$$(A11) \quad Fp \supset FFp.$$

Obviously, (A11) corresponds to the denseness, and it can be demonstrated that this property could just as well have been obtained by the axiom:

$$(A11x) \quad Pp \supset PPp.$$

Often logicians have wanted to extend the tense logical language introducing the metrical tenses, $P(x)$ and $F(x)$, which stand for expressions like 'it has been the case x time units ago that ...' and 'in x units it will be the case that ...', respectively. It is obvious that systems like K_b, K_l, and $K_{ld\infty}$ can be extended in this way.

Given the metric extension of tense logic we can express the basic problems related to the understanding of branching time in a very straightforward manner. For instance, we may consider the proposition $F(x)p \wedge F(x){\sim}p$, where p stands for a contingent statement. This conjunction turns out to be true in K_b for any contingent statement p. Such a result clearly fits badly with the idea of alternative futures in a branching time system. For this reason the metric extension of K_b obviously does not qualify as a satisfactory representation of the idea of alternative future possibilities related to the notion of future contingency. In order to deal with this problem we have to look for other kinds of tempo-modal systems.

5   THE LOGIC OF FUTURE CONTINGENCY

As we have seen in the chapter on Prior's logic, A.N. Prior found much inspiration for the development of modern temporal logic from his study of medieval philosophy and logic. Clearly, he did not see the medieval findings as interesting only from a historical point of view. He also held that the medieval writings on temporal logic may in fact be useful in the practical development of modern temporal logic provided that the medieval contributions are transformed and translated into the formal language used in modern logic. In particular this turns out to be the case with respect to the development of the logic of future contingency. After all, this problem was given high priority during several centuries of academic life of scholasticism. Realising the relevance of such studies, several writers interested in the philosophical perspectives of temporal logic have since Prior's death in 1969

continued his search in scholastic philosophy and logic looking for further conceptual clarification in the study of future contingency. (See for instance [Craig, 1988].)

During the Middle Ages logicians related their science to theology. Clearly they felt that they had something important to offer with regard to solving fundamental logical questions in theology. One of the most important questions of that kind was the problem of the contingent future, which may be stated in the following way: According to Christian tradition, divine foreknowledge is assumed to also comprise knowledge of the future choices to be made by men. But this apparently gives rise to a straightforward argument from divine foreknowledge to necessity of the future: if God already now knows the decision we will make tomorrow, then a now-unpreventable truth about our choices tomorrow is already given! Hence, there seems to be no basis for the claim that we have a free choice, a conclusion which violates the dogma of human freedom.

There exists a very extensive literature about the problem of the contingent future, and any attempt to produce a detailed exposition of the subject seems hopeless. However, an overview over the basic approaches to the problem within scholasticism can be found in the writings of the medieval logician and philosopher Richard Lavenham (c. 1380) in his treatise *De eventu futurorum*. (See [Øhrstrøm, 1983], [Tuggy, 1999].) Lavenham's central idea is quite clear: If two dogmas are seemingly contradictory, then one can solve the problem either by accepting or by rejecting the reality of the contradiction. If the contradiction is accepted then solving the problem will mean to deny at least one of the dogmas. If the contradiction is rejected it must be demonstrated that the contradiction is only apparent and not real.

Denial of the dogma of human freedom leads to fatalism (1st possibility). Denial of the dogma of God's perfect foreknowledge can either be based on the claim that God does not know the truth about the future (2nd possibility), or the assumption that no truth about the contingent future has yet been decided (3rd possibility). Rejection of the reality of the contradiction between the two dogmas must be based on the formulation of a system according to which the two dogmas, rightly understood, can be united in a consistent way (4th possibility).

Lavenham rejected the 1st and the 2nd possibility as contrary to the Christian faith. It should, however, be mentioned that the opinion of St. Thomas Aquinas can be read as a version of the 2nd possibility. Thomas claimed that the knowledge of God abstracts from the difference between past, present and future. According to this view it might be said that all events are 'always' present to God — in an atemporal sense of 'always'! For this reason one may say that God's knowledge is not a foreknowledge! The problem obviously bears on the theological task of clarifying questions such as 'In which way can God know the future?'

It seems that Lavenham, like Ockham, regarded the Aristotelian approach to propositions concerning the contingent future as being equivalent with the 3rd possibility. A number of Ockham's contemporaries favoured this possibility. Peter Aureole (c. 1280–1322), for instance, claimed that neither the statement 'Antichrist

will come' nor the statement 'Antichrist will not come' is true, whereas the disjunction of the two statements is actually true. From that point of view, one can naturally claim that the dogma of divine foreknowledge is still tenable, even if God does not know if Antichrist will come or not. God knows all the truths given, and cannot know if Antichrist will come due to the simple reason that no truth value for the statement 'Antichrist will come' yet exists.

Lavenham maintained that on Aristotle's account some propositions about future contingent facts are neither determinately true nor determinately false. It is, however, unclear whether he had in mind a third truth-value corresponding to "indeterminate", or simply held that no truth-value is defined for such future contingent propositions. Nevertheless Lavenham also rejected the 3rd possibility as contrary to the Christian faith.

Lavenham, like Ockham, preferred the 4th possibility. This solution was originally formulated by William of Ockham (d. 1349), although some of its elements can already be found in Anselm of Canterbury (d. 1109). It is also interesting that Leibniz (1646–1711) much later worked with a similar system as a part of his metaphysical considerations. (See [Øhrstrøm, 1984].)

The most characteristic feature of their theories is the concept of 'the true future'. The Christian faith says that God possesses certain knowledge not only of the necessary future, but also of the contingent future. This means that among the possible contingent futures there must be one which has a special status, simply because it corresponds to the actual course of events in the future. This line of thinking may be called 'the medieval solution', even though other approaches certainly existed. The justification for this is partly that the notion of 'the true future' is the specifically medieval contribution to this problem, and partly that leading medieval logicians regarded this solution as the best one. Lavenham himself called it '*opinio modernorum*', i.e. the opinion of the modern people.

Lavenham obviously knew that William of Ockham had discussed the problem of divine foreknowledge and human freedom in his work *Tractatus de praedestinatione et de futuris contingentibus*. Ockham asserted that God knows all future contingents, but he also maintained that human beings can choose between alternative possibilities. In his Tractatus he argued that the doctrines of divine foreknowledge and human freedom are compatible. Richard of Lavenham made a remarkable effort to capture and clearly present the logical features of Ockham's system as opposed to Aristotle's solution (i.e. the 3rd possibility).

Lavenham considered an argument from God's foreknowledge to the necessity of the future and the lack of human freedom. The main structure of this argument is very close to what is believed to have been the Master Argument of Diodorus Kronos (cf. the chapter on Prior's logic). It is clear from Lavenham's text that he had some knowledge of this old Stoic or Megaric argument through his reading of Cicero's *De Fato*. It is not necessary to view the problem as a theological problem. In fact, reformulated in a general philosophical setting, it has now come to be regarded as one of the most central problems in the logic of time and modality. Using a Priorean formalism the problem can easily be presented in terms of the

following operators from modal logic and metric tense logic:

$F(x)$ "in x time units it will be the case that ..."
$P(x)$ "x time units ago it was the case that ..."
$\Box$ "it is necessary that ...".

   The argument may be understood as based in the following five principles, where A and B represent arbitrary well-formed statements within the logic:

(P1) $F(y)A \supset P(x)F(x)F(y)A$
(P2) $\Box(P(x)F(x)A \supset A)$
(P3) $P(x)A \supset \Box P(x)A$
(P4) $(\Box(A \supset B) \wedge \Box A) \supset \Box B$
(P5) $F(x)A \vee F(x)\sim A.$

Using a deduction very similar to the one used in the chapter on Prior's logic, it can easily be demonstrated that (P1–P5) taken together lead to

(D) $\Box F(y)q \vee \Box F(y)\sim q.$

Here (D) is equivalent to a denial of the dogma of human freedom. Therefore, if one wants to save this dogma (and escape fatalism) at least one of the above principles (P1–P5) has to be rejected. Showing how that can be done Prior constructed the Peircean system (in which P1 and P5 are rejected) as well as the Ockhamistic system (in which P3 is rejected). It is well known that each of these systems provides a solution to the future contingency problem. Since Prior, several philosophers have discussed which one of Prior's systems should be accepted, or whether other, and more attractive, systems dealing with the problem may be constructed.

   As we have seen in the chapter on Prior's logic, Prior himself favoured the Peircean solution, which in fact corresponds to Lavenham's third solution. We may present this solution semantically in the following way: A Peircean model, $(TIME, <, =, C, Peirce)$, is a structure, where $(TIME, <, =)$ is a set of partially ordered instants, C is the set of all maximally ordered (i.e. linear) subsets in $(TIME, <, =)$ (i.e. the so-called 'histories' or 'chronicles'). The before/after relation is supposed to be irreflexive, asymmetric, transitive, and backwards linear. The valuation function, $Peirce(t, c, A)$, for any wff A at any time t and for any chronicle c with $t \in c$, can be defined recursively given a truth-value for any propositional constant at any moment in TIME:

(a) $Peirce(t, c, p \wedge q)$ iff both $Peirce(t, c, p)$ and $Peirce(t, c, q)$
(b) $Peirce(t, c, \sim p)$ iff not $Peirce(t, c, p)$
(c) $Peirce(t, c, Fp)$ iff $Peirce(t', c', p)$ for all c' with $t \in c'$ and some $t' \in c'$ with $t < t'$.
(d) $Peirce(t, c, Pp)$ iff $Peirce(t', c, p)$ for some $t' \in c$ with $t' < t$
(e) $Peirce(t, c, \Box p)$ iff $Peirce(t, c', p)$ for all c' with $t \in c'$.

$Peirce(t, c, A)$ can be read 'A is true at t in the chronicle c'. A formula A is said to be Peirce-valid if and only if $Peirce(t, c, A)$ for any t in any c in any branching time structure, $(TIME, <, =, C)$.

In the Peircean system we may also define another future operator f corresponding to the notion of 'possible future' i.e.

> (f) $Peirce(t, c, fp)$ iff $Peirce(t', c', p)$ for some c' with $t \in c'$
>   and some $t' \in c'$ with $t < t'$.

In addition, we may define G as $\sim f \sim$ and the operator g as $\sim F \sim$. In this way the Peircean system includes four different future-like operators.

If we want a metric version of the Peirce system we have to add a duration function, $dur(t_1, t_2, x)$, standing for the statement 't_2 is x time units after t_1'. Using this function (c) and (d) are replaced by:

> (c′) $Peirce(t, c, F(x)p)$ iff $Peirce(t', c', p)$ for all c' with $t \in c'$
>   and some $t' \in c'$ with $dur(t, t', x)$
> (d′) $Peirce(t, c, P(x)p)$ iff $Peirce(t', c, p)$ for some $t' \in c$ with
> $dur(t', t, x)$

According to this system it obviously follows that

$$F(x)q \supset \Box F(x)q$$

is a Peirce-valid formula. This means that a statement about the contingent future is only true in the Peircean sense if it is true in all possible futures i.e. if it has to be the case. It should also be noted that in the Peircean system the so-called 'determinateness of the future' is rejected. This means that the following expression is not a thesis in the system:

$$\sim F(x)q \supset F(x) \sim q.$$

According to the Peircean system the future should simply be identified with the necessary future. This position has many modern advocates. Although the denial of the determinateness of the future and the collapse of the future and the necessary future make the position rather counter-intuitive from a common sense point of view, A.N. Prior and many of his followers favoured this possibility.

The Peircean position also means that a statement like $F(z)p$ is true if p is true after z time units for any future development. According to the following model this means that $F(x)q$ is true at the event E2, whereas $F(x + y)q$ is false at E1:
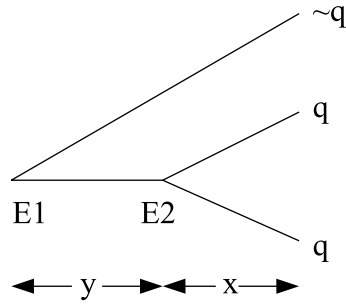
For this reason the Peircean system also includes the view that the expression
$$q \supset P(z)F(z)q$$
cannot hold in general i.e. it does not represent a thesis in the system.

Many writers have studied the formalities of the Peircean system. Axiomatizations of the non-metrical version of the system can be found in [Burgess, 1980] and in [Zanardo, 1990].

However, as argued in [Gabbay *et al.*, 2000, p. 65] the Peircean system has some obvious weaknesses which in fact make the system problematic as a satisfactory candidate to a theory for future contingency. First of all, the system fails to represent many everyday ways of reasoning when it comes to the notions of time. This is due to the fact that the idea of a plain future between possible future and necessary future is not taken into serious consideration. In addition, the handling of operators is very difficult in this Peircean system. For instance, we may notice the crucial feature of the system according to which the expressions $F(x){\sim}q$ and ${\sim}F(x)q$ are non-equivalent.

The Ockhamistic system, on the other hand, leads to the denial of (P3). It is a rather attractive system, although it is certainly also possible to criticise the Ockhamistic position in various respects — as we shall see in the following.

An Ockhamistic model, $(TIME, <, =, C, Ock)$, is a structure, where $(TIME, <, =)$ is a set of partially ordered instants, C is the set of all maximally ordered (i.e. linear) subsets in $(TIME, <, =)$ (i.e. the so-called 'histories' or 'chronicles'). The before/after relation is like in the Peircean case supposed to be irreflexive, asymmetric, transitive, and backwards linear. The valuation function, $Ock(t, c, A)$, for any wff A at any time t and for any chronicle c with $t \in c$, can be defined recursively given a truth-value for any propositional constant at any moment in $TIME$:

    (a) $Ock(t, c, p \wedge q)$ iff both $Ock(t, c, p)$ and $Ock(t, c, q)$
    (b) $Ock(t, c, {\sim}p)$ iff not $Ock(t, c, p)$
    (c) $Ock(t, c, Fp)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $t < t'$
    (d) $Ock(t, c, Pp)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $t' < t$
    (e) $Ock(t, c, \Box p)$ iff $Ock(t, c', p)$ for all c' with $t \in c'$.

$Ock(t, c, A)$ can be read 'A is true at t in the chronicle c'. A formula A is said to be Ockham-valid if and only if $Ock(t, c, A)$ for any t in any c in any branching time structure, $(TIME, <, =, C)$.

If we want a metric version of the Ockhamistic system we have to have a duration function, $dur(t_1, t_2, x)$, standing for the statement 't_2 is x time units after t_1'. Using this function (c) and (d) are replaced by:

(c$'$) $Ock(t, c, F(x)p)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $dur(t, t', x)$

(d$'$) $Ock(t, c, P(x)p)$ iff $Ock(t', c, p)$ for some $t' \in c$ with $dur(t', t, x)$

It is easy to verify that neither $P(x)A \supset \Box P(x)A$ nor $PA \supset \Box PA$ (i.e. (P3) in the above argument) are Ockham-valid for any A, although it will be if A does not contain any reference to the future.

For a long time the problem of axiomatising the non-metric version of the Ockhamist system was open. However, recently Mark Reynolds [2003] has presented a complete and finite axiomatization of this system. In addition to the system K_b for the tense operators, S5 for the modal operator, the ordinary rules of inference, the characteristic new elements in Reynolds' axiom system is the inference rule:

(IRR)   If $\vdash (p \wedge H \sim p) \supset \alpha$ then $\vdash \alpha$ (if p does not appear in α)

and the axioms

(Rey1) $P\alpha \supset \Box P \Diamond \alpha$

(Rey2) $G \bot \supset \Box G \bot$

where $\Diamond$ is defined as $\sim \Box \sim$, and where $\bot$ stands for the contradiction. It is interesting that (Rey1) is a weaker version of the principle (P3) mentioned above (i.e. a weaker version of the critical premises in the Diodoeran Master Argument), whereas the full (P3) is denied in the Ockhamist system. (Rey2) obviously corresponds to an intuition of the maximality of histories.

If the full (P3) is denied in general one may reject the inference from (2) to (3) in the above argument. According to Ockham (P3) should only be accepted for statements which are genuinely about the past i.e. which do not depend on the future. According to this view (P3) may be denied, precisely because the truth of statements like $P(x)F(x)F(y)q$ has not been settled yet since they depend on the future.

In this way, one can make a distinction between soft and hard facts regarding the past. Following the Ockhamistic position a statement like $P(x)q$ would correspond to a hard fact, whereas $P(x)F(x)F(y)q$ and $P(x)F(x)F(y)q$ would represent soft facts.

John Martin Fischer [1994] has questioned the Ockhamistic model. He has suggested a subdivision of the set of soft facts. He has introduced hard-core soft facts as well as hard-type soft facts and soft-type soft facts. His position is just that some soft facts are so hard that "they cannot be falsified without affecting some genuine feature of the past" [Fischer, 1994, p. 127]. We agree with William Lane Craig who has argued that this analysis of soft facts "has gone out of control" [Craig, 1989, p. 236–237]. Fischer's analysis does not provide a

strong argument against the view that the Ockham theory is a fairly accurate representation of our intuitions concerning valid temporal reasoning with regard to the future contingency problem.

It may be argued that Prior's Ockhamistic system does not fit the ideas formulated by William of Ockham completely. Although many of Ockham's original ideas are in fact satisfactorily modelled in Prior's Ockhamistic system, this system lacks a proper representation of the notion of 'the true future', which was in fact one of the most basic ideas in Ockham's world view. Ockham certainly believed that there is truth about the contingent future, which we as human beings cannot know, but which God knows. This assumption of a true future will in terms of modern logic mean that in a branching time model there is a privileged branch at any past, present or future branching point in the model. In consequence, $F(x)q$ is true at E2 and $F(x + y)q$ is true at E1.
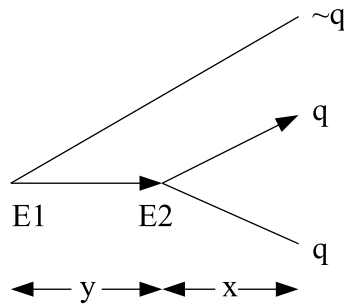
In an interesting paper Nuel Belnap and Mitchell Green [1994] have concentrated on another problem related to this Ockhamistic vision. They have argued that the model not only has to specify a preferred branch corresponding to the true history (past, present, and future). If we want to insist on a concept of future which is different from the possible future as well as the necessary future, it must be assumed that there is a preferred branch at every counterfactual moment. Belnap and Green have based their argument of the following statement:

> The coin will come up heads. It is possible, though that it will come up tails, and then later it will come up tails again (though at this moment it could come up heads), and then, inevitably, still later it will come up tails yet again [Belnap and Green, 1994, p. 379].

This statement may be represented in terms of tense logic

$$F_1 h \wedge \Diamond F_1(t \wedge \Diamond F_1 h \wedge F_1(t \wedge \Box F_1 t))$$

t: tails
h: heads
$F_1 \equiv_{def} F(1)$

and in terms of the following branching time structure

temporal metric
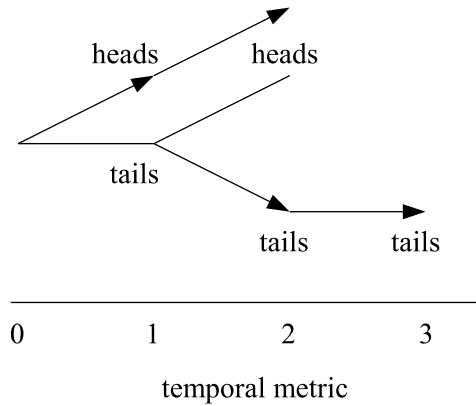
The example shows that if the model is taken seriously, then there must be a function TRL, which gives the true future (extended to a maximal set; Belnap and Green call it "the thin red line") corresponding to a given moment, m. But how can $TRL(m)$ be specified? Belnap and Green have argued that

(TRL1) $m \in TRL(m)$

should hold in general, and that in addition the following

(TRL2) $m_1 < m_2 \supset TRL(m_1) = TRL(m_2)$

may be considered. However, they argue that (TRL2) is inconsistent with the very idea of branching time. Therefore (TRL2) seems to be too strong a requirement. Rather than (TRL2), we propose the weaker condition:

(TRL2′) $(m_1 < m_2 \wedge m_2 \in TRL(m_1)) \supset TRL(m_1) = TRL(m_2)$

This seems to be much more natural in relation to the notion of an Ockhamistic branching time logic. Belnap has later accepted the relevance of (TRL2′) (see [Belnap *et al.*, 2001, p. 166]).

Belnap and Green have argued that any such TRL-function should give rise to a logic in which the following theorems hold:

(T1) $FFA \supset FA$
(T2) $PPA \supset PA$
(T3) $A \supset P(x)F(x)A$

No formal semantics is given by Belnap and Green; however, they seem to assume that the tense operators are interpreted only relative to an instant. This amounts to interpreting tenses using a two-place valuation operator:

$T(m, PA)$ iff $\exists m' : m' < m \wedge T(m', A)$
$T(m, FA)$ iff $\exists m' : m < m' \wedge m' \in TRL(m) \wedge T(m', A)$

Given such a semantics it is straightforward to check that (T1) is valid without (TRL2′). However, (T2) is not valid without (TRL2′), but it is if this assumption is made. The formula (T3) is not valid even if (TRL2′) is assumed. To see why this is the case, consider a situation with an instant t such that $m \notin TRL(m')$ for any $m' < m$. Assume that t is the only instant at which A is true. Then PFA, hence also $A \supset PFA$, will be false at m.

Even the formula (T3′)

(T3′) $A \supset P_1 F_1 A$

is false when evaluated with the following semantics:

$T(m, P(x)A)$ iff $\exists m' : before(m', m, x) \wedge T(m', A)$
$T(m, F(x)A)$ iff $\exists m' : before(m, m', x) \wedge m' \in TRL(m) \wedge T(m', A)$

m': ~A

m: A
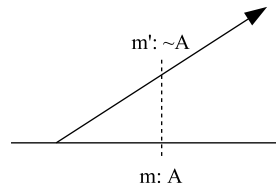
With this interpretation of the tenses (T3′) becomes invalid as illustrated above.

If somebody were to want to defend such a view, such a person would have to say something like: The counterfactual assumption of A does not invalidate the truth of the past prediction $P_1 F_1 \sim A$. If a person is writing now, it certainly was true yesterday that he was going to write after one day. That prediction was true (but of course not necessary) even if he now — while writing — imagines himself asleep. For this reason one may say, that the truth of $A \wedge P_1 F_1 \sim A$, where A stands for 'The person (the writer) is asleep', is in fact conceivable.

In our opinion, however, Prior's Ockhamist theory is indeed satisfactory for most cases. But as demonstrated by Hirokazu Nishimura [1979], there are some rare examples in which the Ockham theory is not sufficient. Such an example is given in [Gabbay *et al.*, 2000, p. 67]. We consider this example in a slightly modified form. The claim is that it is consistent to believe both of the following assumptions:

A1: There is life on earth now, but it will necessarily come to an end.

A2: It is necessarily always the case that if there is life on earth then it is possible for there to be life on earth in a succeeding period also.

A1 means that whatever anyone is doing, life on earth will some day come to an end. However, A2 implies that there is hope (and possibilities) for tomorrow

as long as there is life! Many people believe both A1 and A2. It is our claim that people with this belief should not be seen as inconsistent for this reason.

We think that a tense logical system corresponding to everyday reasoning should allow for the simultaneous belief in both assumptions, (A1–A2). It turns out, however, that this is not the case in an Ockhamistic structure, $(TIME, <, =, C, Ock)$, in which $TIME$ is discrete and each chronicle is isomorphic to the set of integers. We may think of the elements in $TIME$ as possible days, although the choice of time units is of course secondary. The discreteness of the model just means that the possible states in the (past and future) history of life on earth can be modelled using a discrete branching time structure. Given such a model, $(TIME, <, =, C, Ock)$, we shall in the following argue that (A1–A2) turns out to be an inconsistent pair of assumptions. Letting q stand for the statement 'there is life on earth', the two assumptions are represented by the following two expressions:

A_1: $Ock(t_0, c_0, q \wedge \Box F(q \wedge G\sim q))$, where t_0 stands for the 'now'
          and where c_0 stands for 'the true future' right now.
A_2: $Ock(t, c, q \supset \Diamond Fq)$ for any t and any c.

We shall show that the assumption that A_1 and A_2 can be true simultaneously is in conflict with the Ockhamistic theory. On the other hand, we shall argue that a person would not necessarily be inconsistent in holding the conjunction of A_1 and A_2 given another (and perhaps intuitively more satisfactory) theory. — Assume that A_1 and A_2 are both true now i.e. at $t_0 \in c_0$. It follows from A_1 that q is true at t_0 i.e. $Ock(t_0, c_0, q)$. Because of A_2, this means that $Ock(t_0, c_0, \Diamond Fq)$ i.e. there is a chronicle c_1 with $t_0 \in c_1$ and a $t_1 \in c_1$ with $t_0 < t_1$ such that $Ock(t_1, c_1, q)$. Using A_2 again in a similar way we find a chronicle c_2 with $t_1 \in c_2$ and a $t_2 \in c_2$ with $t_1 < t_2$ such that $Ock(t_2, c_2, q)$. This procedure can be carried out ad infinitum using A_2 repeatedly, and in this way we construct the time series: $t_0 < t_1 < t_2 < \ldots < t_i < t_{i+1} < \ldots$. For all t_i in this series we have $Ock(t_i, c_i, q)$. It should also be noted that the series $t_0 < t_1 < t_2 < \ldots < t_i < t_{i+1} < \ldots$ can be completed with segments of c_1, c_2, c_3, $\ldots$ since for each $i = 0,1,2,\ldots$ we have $t_i \in c_{i+1}$ as well as $t_{i+1} \in c_{i+1}$. Since the time structure is discrete, this means that the series $t_0 < t_1 < t_2 < \ldots < t_i < t_{i+1} < \ldots$ gives rise to a chronicle, c, within the Ockhamistic model. Obviously, $t_i \in c$ for $i = 0,1,2,\ldots$ Because of A_1 we have $Ock(t_0, c, F(q \wedge G\sim q))$ i.e. there is an instant t after which $Ock(t', c, \sim q)$ for all t'. However, this would contradict the fact that the series $t_0 < t_1 < t_2 < \ldots < t_i < t_{i+1} < \ldots$ is infinite and unlimited in c. This shows that the assumption of the conjunction of A_1 and A_2 in the context of an Ockhamistic model leads to a contradiction.

The construction procedure mentioned in the above argument can be illustrated by the following figure:
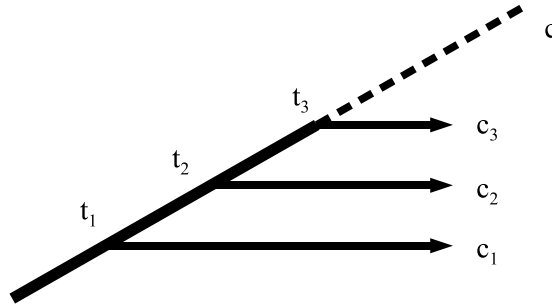
   If, on the other hand, we assume another tense logic different from the Ock-
hamistic system, in which the construction of the chronicle c from the series of
$t_0 < t_1 < t_2 < \ldots < t_i < t_{i+1} < \ldots$ and c_1, c_2, $\ldots$ is forbidden, then the
conjunction of A_1 and A_2 might be accepted without any inconsistency. In such
a tense logic the set of possible chronicles is not necessarily closed under the kind
of construction just mentioned. We must assume that not all linear subsets in
$(TIME, <, =)$ are possible chronicles. An Ockhamistic system revised in this
manner has an interesting affinity to Lebniz' philosophy. For this reason we shall
call such a modified Ockhamistic system a Leibniz system (see [Øhrstrøm and
Hasle, 1995, p. 270 ff.]). It may also be called a bundled logic (see [Gabbay *et al.*,
1994, p. 299 ff.], [Gabbay *et al.*, 2000, p. 67 ff.]). The idea is that when evaluating
a formula of the form $\Box A$ we should take certain (but not all) linear subsets in
$(TIME, <, =)$ passing through t into account i.e.

$$T(t, c, \Box A) \text{ iff } T(t, c', A) \text{ for all } c' \in C(t)$$

where $C(t)$ is set of chronicles which should be taken into account in the evaluation
of modal formulae.

   Following a similar idea, it is possible to obtain the validity of (T3) even if
we want to insist on the assumption of the 'thin red line'. In [Braüner *et al.*,
1998] we have proposed the following semantics of tenses: As usual we need a
set, $TIME$, equipped with a transitive and backwards linear relation, $<$, together
with a function T which assigns a truth value to each pair consisting of an instant
and a propositional letter. Furthermore, adopting Belnap and Green's idea, we
assume the presence of a function TRL which to each instant assigns a branch
such that the conditions (TRL1) and (TRL2$'$) are satisfied. A novel feature of the
semantics we give here is the notion of a (counterfactual) branch with the property
that at any future instant it coincides with the corresponding thin red line. Given
an instant t, the set $C(t)$ of such branches is defined as follows:

$$C(t) = \{c | t \in c \wedge TRL(t') = c \text{ for any } t' \in c \text{ with } t < t'\}$$

   Note that (TRL1) and (TRL2$'$) together say exactly that $TRL(t) \in C(t)$. Also

note that $C(t)$ may contain more branches that just $TRL(t)$. This allows for counterfactuality. In this semantical model truth is relative to an instant, t, as well as to a branch belonging to $C(t)$. By induction, we define the valuation operator T as follows:

$T(t,c,p)$ iff $T(t,p)$ where p is a propositional letter

$T(t,c,p \land q)$ iff $T(t,c,p)$ and $T(t,c,q)$

$T(t,c,\sim p)$ iff not $T(t,c,p)$

$T(t,c,FA)$ iff $T(t',c,A)$ for some $t' \in c$ with $t < t'$

$T(t,c,PA)$ iff $T(t',c,A)$ for some $t' \in c$ with $t' < t$

$T(t,c,\Box A)$ iff $T(t,c',A)$ for all $c' \in C(t)$

A formula A is said to be valid if and only if A is true in any structure $(TIME, <, T, TRL)$ for any instant t and branch c such that $c' \in C(t)$. The tense operators P and F are interpreted as usual in Ockhamistic semantics. It is straightforward to introduce metrical tense operators.

This semantics makes all of the formulas (T1), (T2), (T3), and (T3′) valid. On the other hand, the necessity operator is interpreted differently in the sense that fewer (counterfactual) branches are taken into account. This invalidates the formula

(T4) $F(x)\Diamond F(y)p \supset \Diamond F(x)F(y)p$

which is valid in the usual in Ockhamistic semantics. If the rejection has to be defended we have to accept something like the following: Tomorrow we may have some possibilities regarding the following day which today are not available as possibilities regarding the day after tomorrow. That is, new possibilities may show up. — However, this view can be questioned. Some would in fact insist that (T4) is valid.

Alternatively, one may define a moment as a pair, (t, c), consisting of an instant, t, and a branch $c \in C(t)$. It turns out that this semantics is equivalent to the Leibnizian semantics (see [Øhrstrøm and Hasle, 1995]) and it also fits with models defined in terms of so-called bundled trees (see [Zanardo, 2003a; 2003b; 2003c]). According to this model the branches in the branching time system have to be viewed as 'parallel lines' construed as sets on which a relation is defined corresponding to *indiscernibility up to a certain moment*. In such a model made up of 'parallel lines' the TRL-function will be trivial.
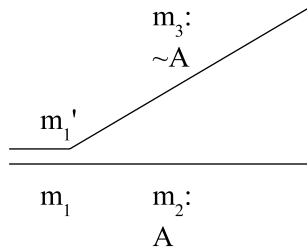
Evaluated on the basis of this Leibnizian view $A \supset P_1 F_1 A$, (T3$'$) clearly holds, whereas $A \supset P_1 \square F_1 A$ does not hold.

Let us return to Lavenham's indeterministic possibilities corresponding to the denials of (P1), (P5), and (P3), respectively. Of course, one may also consider denials of (P2) and (P4). But it is very hard to see how such denials may be defended. The same can be said about rejections of other ingredients of the apparatus involved in Lavenham's argument.

It is obviously possible to deny more than one of the principles (P1–P5). But accepting the basic nature of the above principles, Lavenham's analysis leaves us at the basic level with exactly 4 possible positions relative to the underlying argument. In our opinion, this analysis is rather convincing and the result is certainly interesting, also in a modern context. Moreover, we have argued that if one wants to defend the idea of 'the true future', then this idea should be understood in terms of a Leibnizian model. We agree with the point made in [Gabbay *et al.*, 2000], where it is argued that the Peircean system is unsatisfactory in some respects, and that we, for this reason, should turn to the Ockhamistic logic or perhaps even to the Leibnizian (or bundled) logic.

6   THE CONCEPTUAL BASIS OF TENSE LOGIC QUESTIONED

As a young graduate student Hans Kamp attended A.N. Prior's lectures on tense logic at University of California (UCLA) from September 1965 to January 1966. He became deeply interested in the field and Prior's lectures very much inspired him in his further PhD-studies (see [Copeland, 1996, p. 24]). In the following years Hans Kamp found some very influential results, which in certain respects may be seen as challenging the basic ideas of tense logic. Until his death in 1969 Prior often referred to Kamp's results trying to solve the problems to which they gave rise.

In Kamp's PhD thesis, *On Tense Logic and the Theory of Order*, [1968], which he wrote under the supervision of Richard Montague, Hans Kamp discussed the two-place operators Spq, "q since p", and Upq, "q until p", which semantically can be introduced in the following way:

$$T(t, Spq) \equiv \exists t\,' : (t\,' < t \wedge T(t\,', p) \wedge \forall t\,'' : (t\,' < t\,'' < t \supset T(q, t\,'')))$$
$$T(t, Upq) \equiv \exists t\,' : (t < t\,' \wedge T(t\,', p) \wedge \forall t\,'' : (t < t\,'' < t\,' \supset T(q, t\,'')))$$

Spq may be read "it has been the case that p, and between then and now it has been the case that q". Upq may be read "it will be the case that p, and between now and then it will be the case that q".

Kamp was able to demonstrate that given a linear, dense and infinite temporal order, S and U cannot be defined of truth-functions and 1-place tenses like P and F. For this reason the language to which the Priorean tenses can give rise will not be rich enough to describe the full temporal language. Kamp was, however, also able to show that given the temporal structure satisfying so-called Dedekind continuity, any temporal relation can in fact be defined in terms of the two-place operators S and U; (see [Benthem, 1991, p. 152]). In consequence, it might seem

that we should view S and U (and not P and F) as the cornerstones of tense logic. In this way, Kamp's results may be understood as a strong argument against Priorean tense logic conceived as the logic based on the two operators P and F.

Kamp communicated early versions of his work to Prior, who in his *Past, Present and Future* [Prior, 1967, p. 107 ff.] responded in defence of a tense logic based on P and F. Firstly, Prior pointed out that 'since' and 'until' can in fact be defined in terms of the metric operators, $P(n)$ and $F(n)$. Secondly, he explained that 'since' and 'until' may be defined in terms of P and F provided that we allow the use of propositional quantifiers and accept the idea of instant propositions. The latter solution is obviously very relevant in the modern context of hybrid logic (see the chapter on Prior's logic in this volume).

Kamp also located another problem related to tense logic pointing out that some sentences involving a reference to 'now' cannot be expressed in the Priorean language. One example could be the sentence 'A child was born which will be king' (see [Benthem, 1991, p. 130]). Here three instants (including 'the point of speech') are obviously needed for a proper understanding of the meaning of the sentence. Examples like this may suggest that Reichenbach's threefold distinction (mentioned in section 2 above) is needed after all. Kamp himself, however, suggested another solution involving a Now-operator. In fact he even managed to convince Prior of the usability of the extra tense-logical operator. In a paper published in 1968 Prior wrote: "... until recently I would have ... said that the formalist not only *can* do without the idiomatic 'now' but *must* do without it — that our ordinary use of 'now' has a certain disorderliness about it which makes it unamenable to formalisation ... Recently, however, I have been convinced to the contrary by Hans Kamp ..., and have now myself produced an extension of tense logic with a symbol corresponding fairly closely to the idiomatic 'now'" [Prior, 2003, p. 174]. In addition, Prior has pointed out that an alternative approach may be based on the incorporation of a propositional constant, n, standing for 'the world' or 'everything that is the case'. Given this constant, we may define "it is now that case that p" as $\Box(n \supset p)$. Again this solution turns out to be based on what is now called hybrid logic.

Whereas Kamp's criticism of the Priorean approach is not really a criticism of the basic idea of tense logic but more an identification of some problems which have to be dealt with in tense logic. Other writers, however, have argued that the tense-logical approach as such has to be rejected either because it is based on some unacceptable internal conceptual weaknesses or because it contradicts fundamental features of the external reality.

Robin Le Poidevin [Poidevin, 1996, p. 472] is one of the writers who have argued that the tense-logical approach is unacceptable because it is conceptually inconsistent. The reason is that the A-theorist has to explain the meaning of the B-concept, and for that purpose he will need propositional connectives such as 'and'. However, according to Poidevin it turns out that he would have to say that 'and' means 'and simultaneously', which means that he would have to refer to a B-concept (simultaneity) as something basic. Such arguments are, however, normally

very weak, since the very existence of a well established tense-logical system is a
very strong argument against internal inconsistencies in the conceptual framework
of tense logic. It is very unlikely that there should be such internal and unnoticed
inconsistencies without fatal consequences for the systems as such.

7   THE LOGIC OF DURATIONS

Early in the history modern tense logic it was argued that the idea of a proposition
being true at an instant may be rather problematic, at least if the instant is
understood as durationless. Some have even argued that very few things can
be said to be at an instant without duration. What is, for instance, a tone at
an instant? This kind of criticism of early versions of tense logic and temporal
logic based on durationless instants has led to the development of various kinds
of durational logic. In this section we are going to deal with some of these logical
systems.

Given a structure of instants we can easily construct a set D of durations (some-
times also called periods or intervals). A duration is defined as a pair of instants
(x, y) with $x < y$. Obviously, the order of the instants gives rise to an ordering
relation, '$\angle$' on the set of durations using the following definition:
(Def. $\angle$) $(x, y)\angle(u, v)$ iff $y < u \vee y = u$.

It is easy to deduce various properties of '$\angle$' from the properties of '$<$'. For
instance, if '$<$' is irreflexive and transitive, then it follows that '$\angle$' will also be
irreflexive and transitive.

Several authors have found the relation between $(T, <)$ and $(D, \angle)$ interesting.
It has been pointed out that $(D, \angle)$ does not have to be seen as derived from
$(T, <)$. In fact, it has been demonstrated that the opposite derivation would also
be possible i.e. given $(D, \angle)$ with certain basic properties we may define instants
and construct the structure $(T, <)$.

In his book *Our Knowledge of the External World* [1914] Bertrand Russell pre-
sented a way of constructing instants from durations (events). He further elabo-
rated this idea in the paper 'On order in time' [1936]. According to Russell an
instant can be defined as "a group of events having the following two properties:

    (1) Any two members of the group overlap in time, i.e. neither
          is wholly before the other.

    (2) No event outside the group overlaps with all of them".

A decade later A.G. Walker [1947] suggested a similar and more elaborated
construction. He considered a structure, which we may term $(D, \angle)$, where D
is a non-empty set of periods. This set is ordered by a partial ordering relation
'$\angle$', analogous to the before-after-relation among instants. Two interesting and
related aspects of this model should be mentioned right away: first, it does not
seem counterintuitive to call one period 'earlier' than another one, even if they
'overlap'. Thus 'Mary opened the door before John rushed in' seems quite right,
even if John begins his rushing in before Mary concludes her opening the door.

Nevertheless, the '$a\angle b$'-relation is to be considered as 'strict' in the sense that no overlap between a and b is permitted. Second, since the ordering relation is only partial, and since the notion of overlap has already made itself manifest, it is interesting to consider also the latter relation, defined as

$$a|b \equiv_{def} \sim(a\angle b \vee b\angle a).$$

This obviously corresponds to the idea of two periods a and b overlapping each other. — Walker formulated an axiomatic system using the following two axioms:

(W1) $a|a$
(W2) $(a\angle b \wedge b|c \wedge c\angle d) \supset a\angle d$

In relation to these axioms Walker was able to construct a set-theoretic structure of triplets (A, B, C), where A, B, and C are all sets of durations such that

1. A and B are non-empty

2. the union of A, B and C is the set of all durations

3. every element in A is before every element in B

4. every element in C is overlapping some element in A as well as some element in B.

Walker demonstrated that the structure of these triplets has all the algebraic properties which we would intuitively expect the structure of temporal instants to have. For this reason it may be reasonable to view a temporal instant as such a 'secondary' construct from the logic of durations.

Given the right conditions on the partially ordered structure of events, $(D, \angle)$, we may show that the structure of instants has the mathematical properties which we intuitively expect it to have.

S.K. Thomason [1984] has compared Walker's construction with Russell's and he has argued that Walker's construction should be preferred. In his opinion Walker's theory — much better than Russell's — offers a plausible explanation of time as a continuum.

More than two decades later than A.G. Walker, C.L. Hamblin [1972] also put forth a theory of the logic of durations. Hamblin was not aware of Walker's work when he developed his theory [Hamblin, 1972, p. 331], but he achieved some similar results using a different technique. Hamblin also considered a fundamental structure consisting of a set of durations with a partial ordering relation $(D, \angle)$. In addition he defined the following relations for arbitrary durations, where $(a \bowtie b)$ may be read 'b follows immediately after a', and $a \subseteq b$ may be read 'a is contained in b':

$$a \bowtie b \equiv_{def} (a\angle b \wedge \sim(\exists c : a\angle c \wedge c\angle b))$$
$$a \subseteq b \equiv_{def} \forall c : (c|a \supset c|b)$$

Using the definition of $a \bowtie b$, Hamblin could also offer a derived notion of an instant:

> Any pair of durations (a, b) uniquely defines an instant if and only if $a \bowtie b$.

We shall use expressions like $a \bowtie b \bowtie c$ for the conjunction of $a \bowtie b$ and $b \bowtie c$. Hamblin's axioms can be formulated in the following way using our notation (and omitting external universal quantification):

(Hamblin 1): $\sim (a \angle a)$
(Hamblin 2): $(a \angle b \land c \angle d) \supset (a \angle d \lor c \angle b)$
(Hamblin 3): $a \angle b \supset (a \bowtie b \lor \exists c : a \bowtie c \bowtie b)$
(Hamblin 4): $(a \bowtie c \land a \bowtie d \land b \bowtie c) \supset b \bowtie d$
(Hamblin 5): $(a \bowtie b \bowtie d \land a \bowtie c \bowtie d) \supset b = c$
(Hamblin 6): $\exists b : a \angle b$
(Hamblin 7): $\exists b : b \angle a$
(Hamblin 8): $\exists b : (b \subseteq a \land \sim (b = a))$
(Hamblin 9): $b \subseteq a \supset (T(a, p) \supset T(b, p))$
(Hamblin 10): $\forall b : (b \subseteq a \supset (\exists c : c \subseteq b \land T(c, p))) \supset T(a, p)$

(Hamblin 9) states a kind of dissectiveness: if some proposition p 'is true with respect to' some interval a, and b is contained in a, then p is true also with respect to b. We might also say that this expresses 'downwards inheritance'. In a dual manner, (Hamblin 10) expresses a sort of cumulativity. However, it is well known, at least from later literature on durations, that not all 'properties' of durations behave like this: thus for instance, an 'accomplishment' like 'Mary baked a cake' (say, from 1 p.m. to 4 p.m.) does not entail that Mary baked a cake during the sub-periods, say, from 2 p.m. to 3 p.m. (Note that even though it may be tempting to say that Mary was 'engaged in the process' also during all sub-periods, she certainly did not accomplish it during any of those). It is therefore clear that Hamblin's theory is confined to certain subsets of (properties of) durations.

During the last decade various kinds of durational logic have been studied and applied within artificial intelligence research and natural language understanding (usually under the heading 'interval semantics', which seems more popular in this scientific community). Two researchers in this field who have contributed significantly to the development of durational logic are James Allen and Patrick J. Hayes [1985; 1989]. Like Walker and Hamblin, Allen and Hayes have taken as their starting point the study of the structure of the partially ordered set of durations. They have suggested an axiomatic system, which we reformulate in the following way, where '$\underline{\vee}$' stands for the exclusive disjunction:

(AH1) $(a \bowtie c \land a \bowtie d \land b \bowtie c) \supset b \bowtie d$
(AH2) $(a \bowtie b \land c \bowtie d) \supset (a \bowtie d \underline{\vee} \exists e : a \bowtie e \bowtie d \underline{\vee} \exists f : c \bowtie f \bowtie b)$
(AH3) $\exists b, c : b \bowtie a \bowtie c$
(AH4) $(a \bowtie b \bowtie d \land a \bowtie c \bowtie d) \supset b = c$
(AH5) $a \bowtie b \supset \exists e \forall c, d : (c \bowtie a \bowtie b \bowtie d \supset c \bowtie e \bowtie d)$

This axiomatic system obviously takes the $\bowtie$-relation as the primitive. However, this does not constitute any essential step away from Hamblin's system, in which

the opposite implication of (Hamblin 3) can easily be proved. We therefore have as a theorem

$$(\text{Hamblin3}'): \quad a \angle b \equiv (a \bowtie b \vee \exists c : a \bowtie c \bowtie b)$$

which may obviously be used as a definition of the $\angle$-relation in the AH-system. With this definition (Hamblin 2) is provable in the AH-system. (AH1) and (AH4) are just (Hamblin 4) and (Hamblin 5), and (Hamblin 6–7) are immediate consequences of (AH3). Because of the exclusive disjunctions in (AH2), we can derive $\sim(a \bowtie a)$, i.e. (Hamblin 1). So it seems that (Hamblin 8) is the only difference between the systems (if we disregard Hamblin's special requirements of cumulativity and dissectiveness, cf. Hamblin 9–10).

It follows from (AH4) that the e in (AH5) is uniquely determined by the durations a and b. Following Allen and Hayes, we shall call this resulting duration the *sum* of a and b, i.e. $e = a + b$. However, we point out that this sum-operator is *not* commutative and is in effect a kind of concatenation rather than a 'usual' sum-operator.

Allen and Hayes have shown that two arbitrary durations can be related in exactly 13 different ways, which can all be expressed solely in terms of the $\bowtie$-relation and equality:

a meets $b \equiv_{def} a \bowtie b$
a is met by $b \equiv_{def} b \bowtie a$
a is before $b \equiv_{def} \exists c : a \bowtie c \bowtie b$
a is after $b \equiv_{def} \exists c : b \bowtie c \bowtie a$
a starts $b \equiv_{def} \exists c : b = a + c$
a is started by $b \equiv_{def} \exists c : a = b + c$
a finishes $b \equiv_{def} \exists c : b = c + a$
a is finished by $b \equiv_{def} \exists c : a = c + b$
a overlaps $b \equiv_{def} \exists c, d, e : (a = c + d \wedge b = d + e)$
a is overlapped by $b \equiv_{def} \exists c, d, e : (b = c + d \wedge a = d + e)$
a during $b \equiv_{def} \exists c, d : b = c + a + d$
a contains $b \equiv_{def} \exists c, d : a = c + b + d$
a equals $b \equiv_{def} a = b$

It is very illuminating to study various combinations among these 13 relations. Using Allen's and Hayes' axiomatisation, it is possible to implement a reasoning system, by means of which statements like

If a overlaps b and b is started by c, then a overlaps c;
If a finishes b and b starts c, then a during c;
If a during b and b overlaps c, then a is not met by c;

can be proved. This kind of reasoning will be important in any system, which should be able to perform or simulate common-sense reasoning involving time periods.

We have already pointed out that for some durations, or perhaps rather, certain types of events, there are no sub-parts; for instance, if 'John opened the door' during some period a, it will not be true to say that John opened the door during

any sub-interval b contained in a. In this case, dissectiveness does not obtain (cf. Hamblin 9). When reasoning about durations we often come across durations without parts corresponding to for example opening a door. Allen's and Hayes' reason for excluding in general the axiom (Hamblin 8) is precisely that they want to study these so-called 'moments', which can be understood as durations without any internal structure (not to be confused with 'instants'). It appears that nothing is contained in a moment, and that two moments cannot overlap each other.

Hamblin's (as well as Allen's and Hayes') durational logic is based on a conception of durations as something similar to real intervals. A number of interesting theorems can be proved from Hamblin's axioms, but the system is not sufficient to establish that linear intuition about time on which it is obviously based. The reason for this is that there is nothing in (Hamblin 1–8) to exclude a genuine branching time model. On the other hand, if time should in fact be conceived as branching, then the 'containment'-relation $\subseteq$ in the above axioms will yield some very strange results, and will be rather far from the inclusion relation that Hamblin probably had in mind.

Peter Röper [1980] has developed a more fine-grained logic from very much the same intuition as Hamblin's. Röper starts from a non-empty set S of durations and a relation $\subseteq$ defined on S, which should express the 'inclusion' relation among durations. Röper defines a P-frame as a structure $(S, \subseteq, \angle)$ satisfying:

(A1)   If $x \angle y$, $x' \subseteq x$ and $y' \subseteq y$, then $x' \angle y'$.
(A2)   If for every $x' \subseteq x$ and $y' \subseteq y$ there are $x'' \subseteq x'$ and $y'' \subseteq y'$ such that $x'' \angle y''$, then $x \angle y$.
(A3)   If $x \angle y$ and $y \angle z$, then $x \angle z$.
(A4.1) For any x, there exists $x' \subseteq x$ and y such that $x' \angle y$.
(A4.2) For any x, there exists $x' \subseteq x$ and y such that $y \angle x'$.
(A5.1) For any x, y and z, if $x \angle y$ and $x \angle z$, then there exists $y' \subseteq y$ and $z' \subseteq z$ such that $z' \angle y'$ or $y' \angle z'$.
(A5.2) For any x, y and z, if $y \angle x$ and $z \angle x$, then there exists $y' \subseteq y$ and $z' \subseteq z$ such that $z' \angle y'$ or $y' \angle z'$.

Obviously, (A5.1) corresponds to forwards linearity, whereas (A5.2) ensures backwards linearity. On the other hand, there is nothing in Röper's system to ensure the irreflexivity of the ordering relation.

Some of the further details of Röper's system are mainly concerned with that distinction between dissective and non-dissective 'events', which we have already suggested. We shall recapitulate the main problem by considering the following two propositions:

p: 'Percival drinks a pint of bitter'
q: 'Araminta is in Oxford'.

Let us assume that both propositions are true for a duration a, and let b be an arbitrary sub-duration, i.e. $b \subseteq a$. Then a proposition such as q will also be true for the duration b. Following Röper, we shall say that q is *persistent* (i.e. dissective). This can be symbolically expressed as:

$(T(a, q) \wedge b \subseteq a) \supset T(b, q)$.

A persistent proposition denotes 'a property' in Allen and Hayes' terminology. On the other hand, a proposition such as p may be false for some or all sub-durations. That is, it is in general conceivable that for some sub-duration b, the following formula holds:

$T(a, p) \wedge b \subseteq a \wedge \sim T(b, p)$.

Without doubt, this is true for our present example. Suppose that Percival drank one pint of bitter, beginning at 11:30 a.m. and finishing at 11:40 a.m. Then it is false that he drank one pint of bitter during the subinterval from 11:35 to 11:36. — Allen and Hayes reserve the term 'an event' for propositions of this type. The distinction between these two types of propositions is central for any attempt at establishing an adequate durational logic.

It is evident that Hamblin's theory (cf. Hamblin 9–10) is about what Allen and Hayes have called properties, that is, persistent propositions. Röper, however, makes a distinction between the logic of what he has called *homogeneous* sentences and the logic of 'other sentences'. According to Röper a sentence p is homogeneous if and only if it is 1) persistent (dissective) and 2) cumulative (i.e. for any a, if p is true for all sub-durations of a, then p is true for a).

Röper's way of assigning truth-values to homogeneous sentences closely follows the intuitions embodied by (Hamblin 9–10). A semantical model for the logic of non-homogeneous sentences has to be constructed slightly differently (see [Øhrstrøm and Hasle, 1995, pp. 313 ff.]).

8 TENSE LOGIC AND RELATIVITY

A very common criticism of Priorean tense-logic has been based on various interpretations of the special theory of relativity (STR) and sometimes also interpretations of the general theory of relativity. Prior himself became early aware of the potential conflict between tense logic and STR. In fact, Saul Kripke mentioned the problem in a letter to Prior as early as 1958 (see [Hasle and Øhrstrøm, 1998]). According to Prior many philosophers and scientists who accept what he called the tapestry view of time (i.e. the A-theory of time) have claimed that "they have on their side a very august scientific theory, the theory of relativity, and of course it wouldn't do for mere philosophers to question august scientific theories" [Prior, 1996, p. 49]. Several writers have argued that Priorean tense-logic (i.e. the A-theory of time) contradicts the findings of STR and related physical theories, and that the basic tense-logical position for this reason has to be rejected. Others have maintained that there is not necessarily any contradiction here. Prior himself was aware of the fact that there is a problem here, which should be discussed. He described the conflict in a very clear way:

> The trouble arises when we come to compare another's experiences, when, for example, I want to know whether I saw a certain flash of light before you did, or you saw it before I did. ... It could happen

> that if I assumed myself to be stationary and you moving, I'd get one result — say that I saw the flash first — and if you assumed that you were stationary and I moving, you'd get a different result ... And the conclusion drawn in the theory of relativity is that this question — the question as to which of us is right, which of us really saw it first — is a meaningless question ... Now I don't want to be disrespectful to people whose researches lie in other fields than my own, but I feel compelled to say that this just won't do [Prior, 1996, p. 49].

It is easy to understand what Prior means. Suppose that two observers, A and B, are moving with velocities v and $-v$, from an emitter E, both leaving E when the E-clock reads $t = 0$.
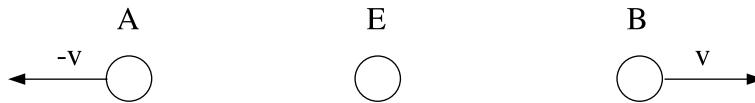
Now, consider an event with the co-ordinates (t_E, x_E) measured from E's inertial system. According to STR we may calculate the time co-ordinates relative to A and B using the following transformations:

$$t_A = L(t_E + \frac{vx_E}{c^2})$$
$$t_B = L(t_E - \frac{vx_E}{c^2})$$

where c is the speed of light, and where $L = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$. A flash is emitted from E and received simultaneously by A and B, yielding same readings, t_E, on the E-clocks. The time co-ordinates for seeing the flash on A $(x_E = -vt_E)$ and B $(x_E = vt_E)$ can be calculated in A's system in the following way:

$$t_{A,A} = L(1 - \frac{v^2}{c^2})t_E$$
$$t_{A,B} = L(1 + \frac{v^2}{c^2})t_E$$

Clearly according to this A is the first to see the flash. The arrivals of the light signals can also be calculated in the B-system:

$$t_{B,A} = L(1 + \frac{v^2}{c^2})t_E$$
$$t_{B,B} = L(1 - \frac{v^2}{c^2})t_E$$

According to this calculation B sees the flash before A. For this reason some physicists would say that the question as to which of the two observers really saw a certain flash first can only make sense if an inertial frame is specified relative to which the calculation should be carried out.

However, Prior thought that the question as to which of the two observers really saw a certain flash first is indeed a meaningful one. He stated that what it means is simply this: "When I was seeing the flash, had you already seen it, or had you not?" [Prior, 1996, p. 50]. Of course, it might be doubted that a physicist committed to the ordinary interpretation of STR would be convinced by that definition. He would probably say that this is begging the question. As a precondition for accepting the question as a meaningful one he would probably

instead demand some experimental procedure, by means of which the question can be settled.

Prior insisted that there is a basic ontological difference between past, present and future. He admitted, however, that we cannot in all cases know whether a given event is present or not, i.e. whether it is really taking place 'now' or not, but he maintained that this epistemological question is very different from the corresponding ontological question. He wanted to make it clear that all what physics could show would be that "in some cases we can never know, we can never *physically find out* [our italics], whether something is actually happening or merely has happened or will happen" [Prior, 1972, p. 323]. Nevertheless, many modern physicists want to go even further, and claim with Albert Einstein:

> There is no irreversibility in the basic laws of physics. You have to accept the idea that subjective time with its emphasis on the now has no objective meaning [Prigogine, 1980, p. 203, Letter to Michele Besso.].

On the other hand, Prior could also note — without doubt with some pleasure — that not even Einstein was quite content with this view. Einstein once said to Carnap that the problem of the Now worried him seriously, explaining that "the experience of the Now means something special for men, something different from the past and the future, but that this important difference does not and cannot occur within physics" [Prior, 2003, pp. 136–137]. Following this kind of reasoning, Prior maintained that questions concerning the human Now make sense, even though we cannot be sure that such questions can ever be decided by physical means. On logical and philosophical grounds Prior maintained that when an event X is happening, another event Y either has happened or has not happened. He strongly rejected the idea of treating 'having happened' as a property that can attach to an event from one point of view whilst not from some other point of view:

> So it seems to me that there's a strong case for just digging our heels in here and saying that, relativity or no relativity, if I say I saw a certain flash before you, and you say you saw it first, one of us is just wrong — is misled it may be, by the effect of speed on his instruments — even if there is just no physical means whatever of deciding which of us it is [Prior, 1996, p. 50].

David Mellor [1998, p. 56 f.] has put forward an interesting argument against an A-theoretical position like Prior's. Mellor's argument has been carefully analysed by Thomas Müller [2000, p. 175]. Following Müller's analysis the crucial conclusion follows from four premises, which can be presented in the following way:

(1) According to the A-theoretical position there is an ontological difference between past, present, and future.

(2) An absolute Now cannot be defined in any reasonable manner in terms of the STR concepts of being past, present or future relative to an inertial system. The point is that on the premises of STR no inertial system is preferred.

(3) The A-theoretical ontological difference between past, present and future must be defined in terms of the STR concepts of being past, present or future relative to an inertial system.

(4) The A-theoretical ontological difference between past, present and future must include the notion of an absolute Now (i.e. a transitive simultaneity relation which does not differ from system to system) as an essential feature.

Given these 4 premises it obviously follows that the A-theoretical position contradicts STR. Premise (1) is a matter of definition, and premise (2) is a crucial and well established belief, which we have no reason to doubt. This means that if the above argument is to be questioned we have to concentrate on the premises (3) and (4). However, as Thomas Müller has convincingly argued, we may in fact reject either of these two premises. This gives rise to at least two different ways of solving the apparent conflict between tense logic and STR without in any way denying the empirical (or measurable) consequences of STR. After the death of A.N. Prior in 1969, Prior's former student, W.H. Newton-Smith, seems to have supported the possibility of rejecting premise (4). Arguing about the possible tension between tense logic and relativity theory he concluded as follows:

> If there is such a tension (between the STR and our ordinary conceptions of past, present and future) I would argue that it is to be resolved through a modification of our ordinary conceptions of past, present and future [Newton-Smith, 1980, p. 187].

This solution involves the idea that ontological judgements may depend on the perspective or the point of view. However, it turns out to be possible to formulate a relativistic tense logic. In fact, Prior himself had pointed out that there is a logic of such functors as 'It appears from a certain point of view that —'. Hence, it is possible to make good sense out of talk about an infinity of different 'apparent' time-series. Prior suspected that the infinity of 'local proper times', which figure in relativistic physics, amounts simply to what appears from various points of view, or what appears to be the course of events in various 'frames of reference'. If the physicist wants to obtain a more general picture, he can "indicate what features of the course of events (what temporal orderings of those events) will be common to all points of view, and one can work out a tense logic for that too" [Prior, 2003, p. 136]. Prior himself made some contributions to the development of such a relativistic tense logic [Prior, 1967, p. 203 ff.] even though he felt that the project of a relativistic tense logic was on the whole a bit strange. In his analysis Thomas Müller [2000, p. 200 ff.] has demonstrated in more details what it would mean to extend traditional tense logic with a logic of perspectives.

Although he accepted that a tense logic without an absolute Now may be formulated, Prior clearly preferred solving the apparent conflict between STR and tense logic in a way which involves a denial of premise 3). If the difference between past, present, and future does not have to be defined in terms of the STR concepts of being past, present or future relative to an inertial system, then premise 2) cannot be used to rule out the possibility of an absolute Now. Having an absolute Now was essential to Prior, since he wanted the difference between past, present, and future to be ontological and independent of the actual choice of perspective. In this way he found it easier to support his general belief that only the present is real. However, Prior himself did not do much to analyse what it would mean in physics if his views were to be accepted. But after Prior's death in 1969 several writers have discussed the problem. In [Øhrstrøm, 1988] and [Øhrstrøm, 2000] a number of conceptual possibilities for upholding at the same time the assumptions of STR and Prior's equating reality with the present are analysed. One of the most obvious ways presupposes the selection of a privileged inertial system, to whose time-coordinates special meanings are attributed. If such a selection is not to be made ad hoc, then it must be possible to list the reasons (preferably cosmological ones) for it. It should be pointed out that the principle of relativity does not exclude a cosmological time (that is, a 'natural' inertial system, which distinguishes itself through the distribution and movement of matter in the universe). Following the British tradition of relativistic cosmology a notion of cosmic time has in many cases been established as an essential component of the models (see [Wegener, 2004]). However, even on the assumption of a homogeneous universe it can be doubted that cosmic time can actually be viewed as an ontological feature of the universe; Whitrow, sharing the assumption of a homogeneous universe, stated:

> It is doubtful whether there exists a precise definition which has so great merits that there would be sufficient reason to consider the time thus obtained as the true one [Whitrow, 1980, p. 304].

This point of view is not shared by all researchers. As Mogens Wegener has pointed out [Wegener, 1999b] some scientists think that the cosmological evidence supports the existence of a universal substratum relative to which a cosmic and absolute simultaneity can be introduced. As pointed out by Thomas Müller [2000, p. 186–187] most cosmological models do in fact at least allow for the definition of an absolute cosmic time. In fact, many writers have like S.J. Prokhovnik [1985, chs. 4–6] and W.L. Craig [2001; 2002] have argued that the very idea of an expanding universe in a very natural manner gives rise to the idea of a cosmic time. Craig has formulated this point in the following way:

> ... the universe contains a privileged class of fundamental observers whose individual planes of simultaneity mutually combine to align with the hypersurface which demarcates the cosmic time. These hypothetical observers are conceived to be moving along with the cosmological fluid so that, although space is expanding and they are therefore mu-

tually receding from each other, each is in fact at rest with respect to space itself [Craig, 2002, p. 117].

At least, it is clear that it is possible to hold Prior's very strong tense-logical position without violating any of the empirical consequences of special relativity, as long as we conceive the tenses as relative to one privileged observer. Arguing from a theological point of view, J.R. Lucas [1989, p. 220] has come to the same conclusion. Lucas points out that "the canon of simultaneity implicit in the instantaneous acquisition of knowledge by an omniscient being" is not incompatible with the STR, since there may be "a divinely preferred frame of reference". As Lucas [1999, p. 104] later argued STR has no bearings on the ontological status of the tenses. He maintained that although STR itself certainly does not include the idea of a preferred inertial system, the theory cannot rule out that a certain inertial system should be preferred for other reasons. In a similar way W.L. Craig sees no strong arguments based on current physics against the idea of a cosmic time as an essential component of his A-theoretical and theological world view:

> In God's temporal experience, there is a moment, which is present in metaphysical time, wholly independently of physical clock times. Thus God would know, without any dependence on clock synchronization procedures, or on any physical operations at all, which events were simultaneously present in metaphysical time — and He would know this simply in virtue of His knowing at every such moment the unique set of present-tense propositions true at that moment, without any need of a sensorium or any physical observation of the universe [Craig, 2002, p. 109–110].

If there is some privileged frame of reference, then the temporal co-ordinates relative other inertial systems as they appear in the equations of STR do not strictly speaking represent proper time. For this reason Prior claimed:

> we may say that the theory of relativity isn't about real space and time . . .   the time which enters into the so-called space-time of relativity theory . . .   is just part of an artificial framework which the scientists have constructed to link together observed facts in the simplest way possible. . .   [Prior, 1996, p. 50–51].

Prior did not mind playing that parlour game, too. He realised that the non-linear structure of space-time points, ordered with absolute before-after relations, possibly of a causal nature, constitutes an interesting object of study for the tense logician. The structure branches both forwards and backwards, so it is not immediately clear how the corresponding tense logic is to be axiomatised. He argued [Prior, 1967, p. 203 ff.] that the characteristic axioms for relativistic space-time are:

$$FGq \supset GFq$$
$$PHq \supset HPq.$$

The following diagram is an illustration of the first of these two theorems:
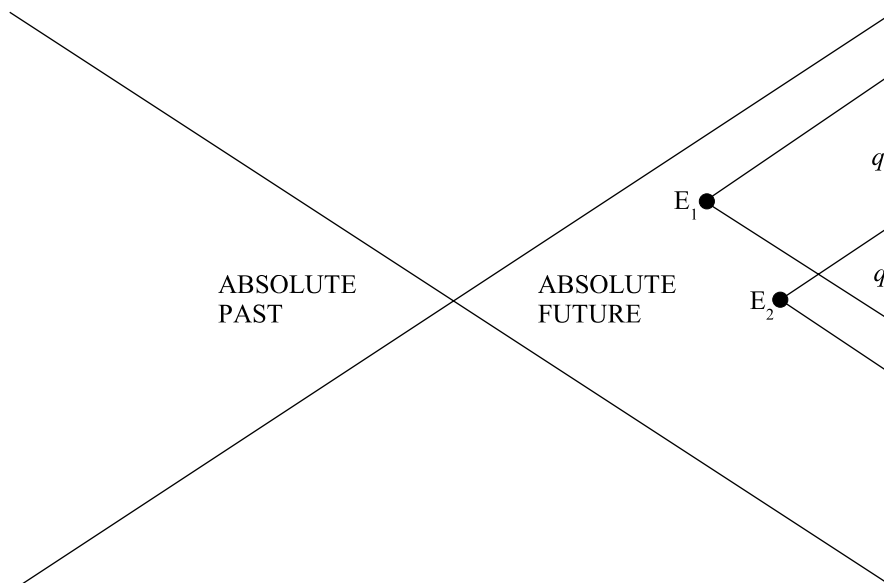
The antecedent of this theorem, FGq, means that there is an event E_1 in the absolute future, at which Gq holds. Given that this is the case, the diagram illustrates that Fq will be the case at any future event, E_2, i.e. that GFq is also the case now. — The other theorem, $PHq \supset HPq$, can be illustrated in a similar way.

Prior's argumentation was thorough and detailed, although a more systematic investigation of the relation between special relativity and tense logic was not carried out until 1980 (see [Goldblatt, 1980]). A decade earlier on, Gerald Massey had directed a frontal attack on tense logic as a new discipline. He had specifically referred to results from the STR, accusing Prior of promoting "bad physics and indefensible metaphysics" [Massey, 1969]. However, in the light of the analysis above and the later results like Goldblatt's, Massey's attack turned out to be misconceived.

Although some results regarding relativistic tense logic have been obtained by Prior and his followers, J.P. Burgess [1984] in his overview of tense logic had to observe that a tense logic for special relativity had not yet been worked out fully. In our opinion this is still the case, although some important results within the field have been produced over the last two decades. One of the most remarkable works in the period is Nuel Belnap's work on branching space-time [1992]. In this work the ambition is to develop an indeterministic tense logic based on the traditional relativistic view that events in general should be conceived as local rather than

global. Where Prior's notion of branching time may be seen as basicly Newtonian, Belnap's branching space-time is Einsteinian. Whereas Priorean branching time may be seen as a system of the histories (or chronicles), Belnap's branching space-time should be conceived as a system of four-dimensional space-time units. Belnap's work on branching space-time has later been continued by others (see e.g. [Rakić, 1997], [Xu, 1997], [Müller, 2002]).

We have seen that Priorean tense logic is not consistent with the consequences of STR. As argued by Müller two different formulations of tense logic consistent with STR were in fact suggested by Prior himself, although he did not work out these theories in details. The first of these may when combined with the idea of branching time be seen as corresponding to Belnap's branching space-time, whereas the other approach to relativistic tense logic presupposes the additional idea of a preferred inertial system.

It may be concluded that although some interpretations of STR may seem to be in conflict with the Priorean view of the tense logic (i.e. the A-theory) there is basicly no contradiction between the A-theoretical view of the tense logic and the empirical results of modern physics. On the contrary, as pointed out by Lucas [1999, p. 105] "it looks as if a tensed view of time is in fact required by physics...". It should also be mentioned that Storrs McCall in his 'A Model of the Universe' [McCall, 1994] has convincingly demonstrated how the study of physics and temporal logic can be integrated in a very fruitful and useful manner.

9 AGENCY AND TEMPORAL LOGIC

In recent works, Vincent F. Hendricks [2003a; 2003b] has argued that since knowledge is in principle acquired over time, a theory of knowledge should be based on a temporal logic. This is required not only for a proper treatment of knowledge, but also with reference to other notions which presuppose the involvement of agents, for instance obligation and belief. For this reason, the development of a temporal logic taking agency specifically into account will be worthwhile. In fact agency is an implicit background assumption of branching time itself, since the very idea of branching is related to free choices of agents — or at the very least some kind of indeterministic behaviour.

In response to the standard claim that modern academic society is divided into a scientist and a humanist culture, Nuel Belnap [1996, p. 241] has suggested that branching time with agents and choices should be seen as "a high-level, broadly empirical theory of our world that counts equally as proto-physical and proto-humanist". This means that the crucial features of our world may be integrated on the basis of a temporal logic incorporating essential notions related to a proper understanding of agency.

Prior himself was well aware of the importance of the notion of agency and its relations to 'decision' and 'contemplation' (see e.g. [Prior, 2003, p. 59 ff.]). He was particularly interested in the relations between 'knowledge' and 'free action'. Although Prior did not seek to establish any essential definition of knowledge,

he did presuppose that if a statement can be known now, it must be true now. At the same time, Prior maintained that no free action, and no decision in the precise sense of the word, can be known beforehand by anybody. In his view the freedom of choice presupposes some incompleteness of knowledge regarding the future. When it comes to free choices there is "nothing to be known beforehand" [Prior, 2003, p. 62]. As Prior saw it, a statement like "A is going to perform the act X tomorrow" cannot be true now, if the act in question is to be free in the proper sense of the word, that is, if A has a genuine choice between doing X and not doing X tomorrow. Prior's view is closely related to his Peircean approach to the semantics for future tense statements. Belnap [1996, p. 265 ff] has followed this Priorean line. He has stated the position as being based on a trilemma, which can be paraphrased in the following way:

(K1)   Knowledge entails truth.
(K2)   A future tense statement can only be true now if it is necessarily true now.
(K3)   Knowledge of free future actions cannot be ruled out.

It is easy to see that the conjunction of (K1-3) gives rise to a contradiction: Let us following (K3) assume that the person A knows that the person B is going to perform the action X freely. According to (K1) the proposition 'A knows that B is going to perform the action X' entails that 'B is going to perform the action X' is true now. According to (K2) this means that it is necessary that B is going to do X. But then B will have no alternative to performing the action X, and therefore the action X will not be free, which is contrary to the assumption.

As we have seen, Prior and Belnap have maintained that this problem should be solved by denying (K3), i.e. by claiming that nobody (not even God) can know beforehand what anybody is going to do *freely*. This analysis is obviously closely related to the problems regarding future contingency, and already in section 5 it was made clear that such a Peircean understanding of future tense statements can indeed be denied with reference to the Ockhamistic position. In other words, logical models exist which allow for the simultaneous truth of (K1), (K2) and (K3).

The works already mentioned by Vincent F. Hendricks [2003a; 2003b] in fact suggest an analysis of knowledge based on Ockhamistic branching time semantics. Following some interesting ideas proposed by K. Kelly [1996], Hendricks has studied a model in which possible worlds are represented as pairs of the form (ϵ, n), where n is a natural number, and where $\epsilon = (a_0, a_1, \ldots, a_n, \ldots)$ is a so-called evidence stream (i.e. an ω-sequence of natural numbers). The model also includes so-called 'handles' and 'fans':

$$\epsilon = (a_0, a_1, \ldots, a_n, \ldots) \qquad \text{(evidence stream)}$$
$$(\epsilon, n) \qquad \text{(possible world)}$$
$$\epsilon|n = a_0, a_1, \ldots, a_{n-1} \qquad \text{(handle)}$$
$$[\epsilon|n] = \{(\tau, k)|k \in \omega \text{ and } \tau|n = \epsilon|n\} \qquad \text{(fan)}$$
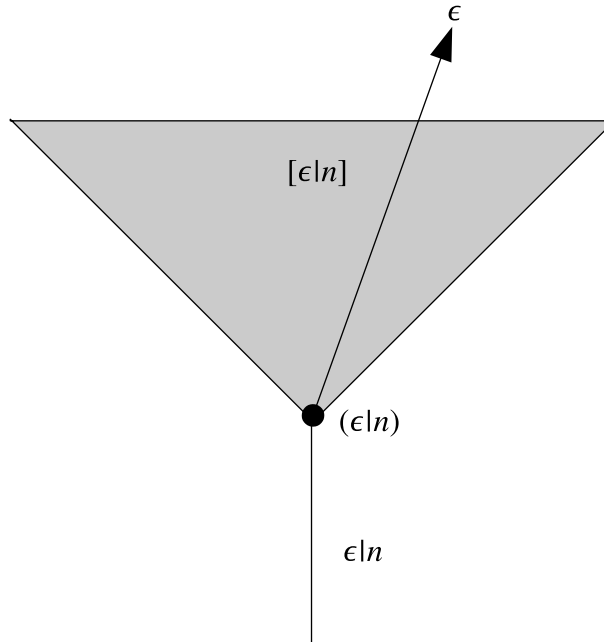
Given this formalism we may speak of the set M of all triples of the form (ϵ, n, a_n). Obviously, all truths about the model follow from the information included in M. On top of the model Hendricks has introduced a formal language that includes epistemic modalities. The key notion is a so-called discovery method, δ, which is a function taking a handle $\tau|n$ as input and producing as output a hypothesis, construed as a set of possible worlds. In fact $\delta(\tau|n)$ can be read as "the hypothesis (i.e. the suggested knowledge) obtained by the method δ on the basis of the evidence $\tau|n$". Hendricks has defined knowledge as limiting convergence such that δ is said to know the hypothesis h at (ϵ, n), if h corresponds with M, and if after a certain time the hypothesis produced by δ will remain unchanged as h. Thus, a person adhering to this discovery method may thereby acquire knowledge. In short we shall say that the method *knows* something! On the basis of this definition, we can introduce an epistemic operator K_δ corresponding to the discovery method δ. Whether a contingent hypothesis h can be known by a method δ will obviously depend on the properties of δ, but it is not in principle ruled out. It might be that δ at (ϵ, n) knows that h is the case at (ϵ, n'), where $n' \geq n$, although there is some (τ, n') in $[\epsilon|n]$ such that (τ, n') does not belong to h.

With the semantics sketched here it is possible to establish a tense-logical system extended with the epistemic operator K_δ. In this system it turns out that for instance the implication $K_\delta h \supset GK_\delta h$ is a valid theorem. Theorems like this one, which involve temporal as well as epistemic operators, nicely illustrate the interesting formal features of the kind of modal operator epistemology suggested

by Hendricks.

Hendricks' framework for dealing with agency and epistemic logic in the context of a temporal logic thus has some most interesting features, but it is by no means the only attempt in this direction. As indicated above, crucial aspects of the discussion can be traced back to the works of A.N. Prior. Moreover, a proper theory of agency has to incorporate several aspects in addition to knowledge. It also has to deal with notions such a belief, desire, and obligation.

As for the notion of obligation, Georg Henrik von Wright [1951] was the first philosopher to study the formalities of what is known as deontic operators, which are introduced in order to formulate the basic logic of obligation. As can be seen in the chapter on Prior's Logic in this volume, Prior was already in the 1950s quite preoccupied with the potentials of deontic logic as a new branch of logic. He also realized already then that the logic of obligation should be conceived in a broader context, namely as an integrated part within a tempo-modal framework. He never got around to working out the details of such a system, though. After the death of Prior one of the important milestones in this respect was the work of Richmond H. Thomason [1981a; 1981b], who demonstrated how a logic of obligation (i.e. a deontic logic) can be constructed within the framework of a branching time logic.

Since the beginning of the 1990s, considerable progress has been made with respect to integrating theories of agency, obligation and temporal logic. This work, commonly known as 'stit-theory', has pivoted around expressions of the form 'α sees to it that Q', which is formally represented as [α stit : Q]. Much important work on these 'stit-theories' has been carried out (see e.g. [Belnap and Perloff, 1988], [Perloff, 1991], [Perloff, 1995], [Horty, 2001], [Horty and Belnap, 1995], and [Belnap *et al.*, 2001]). Normally a distinction is made between *achievement stits*, represented as 'astit', and *deliberative stits*, represented as 'dstit' (see [Belnap *et al.*, 2001, p. 29]). For our purposes here the logic of "dstit" will be sufficient.

It is not obvious how the stit-grammar should be extended into a deontic logic. John F. Horty has developed a theory, which he has described as "a deontic logic designed to represent what agents ought to do within a framework that allows, also, for the formulation of a particular variant of act utilitarianism, the dominance theory" [Horty, 2001, p. 78]. However, Horty's approach is by no means the only possible way to establish a logic of obligation on the basis of the fundamental stit-grammar.

While progress has been made it has also become evident that there are considerable complications involved with the formulation of a satisfactory model for the syntax and logic of the various forms of stit-expressions. Let us give an example of the kind of problems encountered when studying the logic of agency and time. Consider as an example assuming the following scenario, which is an elaborated version of the 'Good Samaritan example' given in [Belnap *et al.*, 2001, p. 309 ff.]. (In the following, *to be obliged to* means the same as *being under an obligation to*.)

> (1)   Arthur is not obliged to kill Joe a week from now (i.e. Arthur
> is not obliged to do the act K), although he is in fact going
> to do so.

(2) Joe is wounded now, and Arthur is obligated to help him surviving (i.e. Arthur is obligated to do the act H).

(3) Arthur's doing H now, entails his doing the act K.

(4) If "α performs X entails that α performs Y", then "α is obligated to do X entails that α is obligated to do Y".

Since it may in fact be the case that Arthur is going to kill Joe a week from now, these assumptions appear to constitute a perfectly consistent scenario. However, it turns out that (1–4) taken together may easily lead to a contradiction. — At least it appears obvious that (3) and (4) imply:

(5) The fact that Arthur is obliged to do H entails that he is obliged to do the act K.

Because of (2) this implies:

(6) Arthur is obliged to do the act K, i.e. Arthur is obliged to kill Joe a week from now.

But (6) evidently contradicts (1).

One may try to find a way out of this paradox through a careful use of the stit-formalism. The above assumptions may be represented in the following way, where $Oblg$ stands for an operator corresponding to 'it is obligatory that':

(S1) $\sim Oblg : [\text{Arthur dstit} : K]$

(S2) $Oblg : [\text{Arthur dstit} : H]$

(S3) $[\text{Arthur dstit} : H] \supset [\text{Arthur dstit} : K]$

(S4) $([\alpha \text{ dstit} : X] \supset [\alpha \text{ dstit} : Y]) \supset (Oblg : [\alpha \text{ dstit} : X] \supset Oblg : [\alpha \text{ dstit} : Y])$

When represented in this way it becomes obvious why we are seemingly led into a contradiction. The catch is, however, that (S4) is invalid even though it seems intuitively correct. The consequent only follows if the antecedent is settled (i.e. holds universally). This means that we have:

(S4′) $(Sett : ([\alpha \text{ dstit} : X] \supset [\alpha \text{ dstit} : Y])) \supset (Oblg : [\alpha \text{ dstit} : X] \supset Oblg : [\alpha \text{ dstit} : Y])$

Here the operator '$Sett$' corresponds to the necessity operator, $\Box$, which we have used in earlier sections. However, if we assume that Arthur for some reason must kill Joe a week from now, given that he helps him surviving to-day, we have:

(S3′) $Sett : ([\text{Arthur dstit} : H] \supset [\text{Arthur dstit} : K])$

But rewritten in this way, the contradiction will occur again. As pointed out in [Belnap *et al.*, 2001, p. 309 ff.], we should however take the temporal aspect into account in the representation, since the killing-act described in this example is in fact something to be carried out in the future.

$(S1')$ $\sim Oblg : Will : [\text{Arthur dstit} : K]$

$(S2)$ $Oblg : [\text{Arthur dstit} : H]$

$(S3'')$ $Sett : ([\text{Arthur dstit} : H] \supset Will : [\text{Arthur dstit} : K])$

$(S4')$ $Sett : ([\alpha \text{ dstit} : X] \supset [\alpha \text{ dstit} : Y]) \supset (Oblg : [\alpha \text{ dstit} : X] \supset Oblg : [\alpha \text{ dstit} : Y])$

This set of assumptions does not lead to any contradiction, but a contradiction will appear again if (S4′) is replaced by:

$(S4'')$ $Sett : ([\alpha \text{ dstit} : X] \supset Will : [\alpha \text{ dstit} : Y]) \supset (Oblg : [\alpha \text{ dstit} : X] \supset Oblg : Will : [\alpha \text{ dstit} : Y])$

Here "$Will$" is in fact a tense operator. In this way the analysis nicely illustrates the benefits of a proper integration of tense logic and the logic of obligation.

It may be argued that (S4″) is reasonable. However, according to [Belnap *et al.*, 2001, p. 309 ff.] the answer to this new version of the argument is that the general validity of (S3″) should be rejected. It is, however, evident that a denial of (S3″) being true in some possible cases may also be seen as somewhat problematic. Given that Arthur is acting freely, (S3″) is obviously false if "$Will$" is interpreted in the Peircean manner. This is however not the case if "$Will$" is interpreted in the Ockhamistic manner. For this reason the solution given by [Belnap *et al.*, 2001] is far from self-evident. In our opinion further discussion is needed.

The above example clearly illustrates that a number of rather complicated problems appear when we try to incorporate problems from real life in the context of an integrated theory of time and agency. The researchers who are working with the stit-theory have actually located a number of other problems concerning agency in the context of a temporal logic. Among other things it turns out that the problems regarding quantification and the grammar of nested stit-expressions with deontic modalities are rather challenging, as pointed out in [Belnap *et al.*, 2001, p. 318 ff.].

10 TOWARDS A NEW TEMPO-MODAL FRAMEWORK

Temporal logic is a huge field, philosophically as well as technically. Only the core of its philosophical background has been covered here, according to the evident selection of those issues and definitions which the authors of this chapter have deemed to be the vital ones.

The considerations in section 9 have taken us to the boundaries of temporal logic, and possibly beyond. But they do illustrate, partly how temporal issues come to touch on other issues such as obligation and agency, and partly how complicated it can be to obtain a conceptually acceptable representation, when one tries to incorporate these notions into a tempo-modal framework. The solutions certainly do not appear to be straightforward. Although the works of Belnap, Perloff, Horty, Xu, and Hendricks undoubtedly represent significant steps forward in theory development, and although these researchers have in fact solved some

important problems, there is obviously still a lot to be done in order to establish a satisfactory theory dealing with time and agency. On the other hand, we have no reason to doubt that it will be possible to formulate a general logical theory for time and agency, and that it can be done in a Priorean spirit. As argued by Nuel Belnap [1996, p. 241], such a theory may qualify as a fundamental theory on the basis of which a number of important scientific and humanistic studies can be carried out. In fact, this new tempo-modal framework presupposes a new world-view, a new cosmology, very far from the so-called "block-universe" which is an attempt to represent reality as a multi-dimensional and timelessly existing unity. As argued by Mogens Wegener, "nothing less than the full acceptance of a temporal flow will do" [2000, p. 258].

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

[Allen and Hayes, 1985] J.F. Allen and J.P. Hayes. A common-sense theory of time. In Aravind Joshi, editor, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence: IJCAI-85, August 18–23*, pages 528–531. University of California and American Association for Artificial Intelligence, Los Angeles, CA, 1985.

[Allen and Hayes, 1989] J.F. Allen and J.P. Hayes. Moments and points in an interval-based temporal logic. *Computer Intelligence*, 5:225–238, 1989.

[Balbiani *et al.*, 2003] P. Balbiani, N-Y. Suzuki, F. Wolter, and M. Zakharyaschev, editors. *Advances in Modal Logic*, volume 4. King's College Publications, London, 2003.

[Belnap and Green, 1994] Nuel Belnap and M.S. Green. Indeterminism and the thin red line. In J. Tomberlin, editor, *Logic and Language*, volume 8 of *Philosophical Perspectives*, pages 365–388. Ridgeview Press, Atascadero, CA, 1994.

[Belnap and Perloff, 1988] Nuel Belnap and M. Perloff. Seeing to it that: A canonical form for agentives. *Theoria*, 54:175–99, 1988.

[Belnap *et al.*, 2001] Nuel Belnap, M. Perloff, and M. Xu. *Facing the Future. Agents and Choices in Our Indeterminist World*. Oxford University Press, 2001.

[Belnap, 1992] Nuel Belnap. Branching space-time. *Synthese*, 92:385–434, 1992.

[Belnap, 1996] Nuel Belnap. Agents in branching time. 1996. In [Copeland, 1996, pp. 239–271].

[Benthem, 1991] Johan van Benthem. *The Logic of Time*, volume 156 of *Synthese Library*. Kluwer Academic Publishers, London, second edition, 1991.

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. Extended modal logic. In *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*, pages 413–484. Cambridge University Press, Cambridge, 2001.

[Boole, 1953] George Boole. *Studies in Logic and Probability*. Watts & Co, London, 1953.

[Braüner *et al.*, 1998] Torben Braüner, Per Hasle, and Peter Øhrstrøm. Ockhamistic logics and true futures of counterfactual moments. In L. Khatib and R. Morris, editors, *Proceedings of Fifth International Workshop on Temporal Representation and Reasoning*, pages 132–139. IEEE Computer Science Society Press, Long Beach, CA, 1998. Affiliated with *11th Annual Florida Artificial Intelligence Research Symposium*.

[Braüner *et al.*, 2005] Torben Braüner, Per Hasle, and Peter Øhrstrøm, editors. *Proceedings of Conference on The Logic of Time and Modality*. Kluwer, 2005. Special issue of the journal *Synthese*. In preparation.

[Braüner, 2002a] Torben Braüner. Homophonic theory of truth for tense logic. In Frank Wolter, Heinrich Wansing, Maarten de Rijke, and Michael Zakharyaschev, editors, *Advances in Modal Logic*, volume 3, pages 59–72. World Scientific Publishing Company, New Jersey, London, Singapore, Hong Kong, 2002.

[Braüner, 2002b] Torben Braüner. Modal logic, truth, and the master modality. *Journal of Philosophical Logic*, 31:359–386, 2002.

[Burgess, 1980] John P. Burgess. Decidability for branching time. *Studia Logica*, 39:203–218, 1980.

[Burgess, 1984] John P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Extensions of Classical Logic*, volume II of *Handbook of Philosophical Logic*, pages 89–133. Reidel, Dordrecht, 1984.

[Copeland, 1996] Jack Copeland, editor. *Logic and Reality. Essays on the Legacy of Arthur Prior*. Clarendon Press, 1996.

[Craig, 1988] W.L Craig. *The Problem of Divine Foreknowledge and Future Contingents from Aristotle to Suarez*. E.J. Brill, 1988.

[Craig, 1989] W.L. Craig. Nice soft facts: Fischer on foreknowledge. *Religious Studies*, 25:235–246, 1989.

[Craig, 2000] W.L. Craig. *The Tensed Theory of Time. A Critical Examination*, volume 293 of *Synthese Library*. Kluwer Academic Publishers, 2000.

[Craig, 2001] W.L. Craig. *God, Time and Eternity*. Kluwer Academic Publishers, 2001.

[Craig, 2002] W.L. Craig. Relativity and the 'elimination' of absolute time. 2002. In [Øhrstrøm, 2002, pp. 91–128].

[Duffy and Wegener, 2000] M.C. Duffy and Mogens Wegener, editors. *Formal Interpretations*, volume 1 of *Recent Advances in Relativity Theory*. Hadronic Press, 2000.

[Fine, 2005] Kit Fine. The reality of tense, 2005. To be published in [Braüner *et al.*, 2005].

[Fischer, 1994] J. M. Fischer. *The Metaphysics of Free Will*. Blackwell, Oxford, 1994.

[Fraser *et al.*, 1972] J.T. Fraser, F.C. Haber, and G.H. Müller, editors. *The Study of Time*. Springer-Verlag, Berlin, 1972.

[Gabbay *et al.*, 1994] D.M. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal logic. Mathematical foundations and computational aspects*, volume 1. Clarendon Press, Oxford, 1994.

[Gabbay *et al.*, 2000] D.M. Gabbay, M.A. Reynolds, and M. Finger. *Temporal logic. Mathematical foundations and computational aspects*, volume 2. Clarendon Press, Oxford, 2000.

[Goldblatt, 1980] R. Goldblatt. Diodorean modality in Minkowski spacetime. *Studia Logica*, 39(2–3):220–236, 1980.

[Hamblin, 1972] C.L. Hamblin. Instants and intervals. 1972. In [Fraser *et al.*, 1972, pp. 324–331].

[Hasle and Øhrstrøm, 1998] Per Hasle and Peter Øhrstrøm. Foundations of temporal logic — the WWW-site for Prior-studies. `http://www.kommunikation.aau.dk/prior`, 1998. Based on the collection in Bodleian Library, Oxford.

[Hendricks, 2003a] Vincent F. Hendricks. Active agents. *Journal of Logic, Language and Information*, 12:469–95, 2003.

[Hendricks, 2003b] Vincent F. Hendricks. *Forcing Epistemology*, 2003. Preprint, Roskilde University.

[Hilpinen, 1981] Risto Hilpinen, editor. *New Studies in Deontic Logic, Norms, Actions, and Foundations of Ethics*. Synthese Library. Reidel, Dordrecht, 1981.

[Horty and Belnap, 1995] John F. Horty and Nuel Belnap. The deliberative stit: A study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.

[Horty, 2001] John F. Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.

[Kamp, 1968] H. Kamp. *On Tense Logic and the Theory of Order*. PhD thesis, University of California in Los Angeles, 1968.

[Kelly, 1996] K. Kelly. *The Logic of Reliably Inquiry*. Oxford University Press, New York, 1996.

[Lucas, 1989] J.R. Lucas. *The Future. An Essay on God, Temporality and Truth*. Basil Blackwell, Oxford, 1989.

[Lucas, 1999] J.R. Lucas. A note on time and tense. 1999. In [Wegener, 1999a, pp. 100–106].

[Marsh, 1956] R.C. Marsh, editor. *Logic and Knowledge*. Allen and Unwin, London, 1956.

[Massey, 1969] G. Massey. Tense logic. Why bother? *NOUS*, 3:17–32, 1969.

[McArthur, 1976] R.P. McArthur. *Tense Logic*. Reidel, 1976.

[McCall, 1994] Storrs McCall. *A Model of the Universe. Space-Time, Probability, and Decision*. Clarendon Press, Oxford, 1994.

[McTaggart, 1908] J.M.E. McTaggart. The unreality of time. *Mind*, 17:456–473, 1908.

[Mellor, 1981] David Mellor. *Real Time.* Cambridge University Press, 1981.

[Mellor, 1998] David Mellor. *Real Time II.* Routledge, London, 1998.

[Müller, 2000] Thomas Müller. *Arthur Priors Zeitlogik. Eine Problomorientierte Darstellung.* PhD thesis, Albert-Ludwigs-Universität, Freiburg, 2000.

[Müller, 2002] Thomas Müller. Branching space-time, modal logic, and the counterfactual conditional. 2002. In [Placek and Butterfield, 2002, pp. 273–291].

[Newton-Smith, 1980] W.H. Newton-Smith. *The Structure of Time.* Routledge & Kegan Paul, London, 1980.

[Nishimura, 1979] Hirokazu Nishimura. Is the semantics of branching structures adequate for non-metric Ockhamist tense logics. *Journal of Philosohical Logic*, 8:477–478, 1979.

[Øhrstrøm and Hasle, 1995] Peter Øhrstrøm and Per Hasle. *Temporal Logic — from Ancient Ideas to Artificial Intelligence*, volume 57 of *Studies in Linguistics and Philosophy.* Kluwer Academic Publishers, Dordrecht, 1995.

[Øhrstrøm, 1983] Peter Øhrstrøm. Richard Lavenham on future contingents. *Cahiers de l'Institut du Moyen-Âge Grec et Latin*, 44:180–186, 1983.

[Øhrstrøm, 1984] Peter Øhrstrøm. Anselm, Ockham and Leibniz on divine foreknowledge and human freedom. *Erkenntnis*, 21:209–222, 1984.

[Øhrstrøm, 1985] Peter Øhrstrøm. Hamilton's view of algebra as the science of pure time and his revision of this view. *Historia Mathematica*, 12:45–55, 1985.

[Øhrstrøm, 1988] Peter Øhrstrøm. *Nogle aspekter af tidsbegrebets rolle i de eksakte videnskaber med særligt henblik på logikken — en videnskabshistorisk og en videnskabsteoretisk undersøgelse.* Aalborg Universitetsforlag, 1988.

[Øhrstrøm, 2000] Peter Øhrstrøm. Tense logic and the special theory of relativity. pages 140–148. 2000. In [Duffy and Wegener, 2000].

[Øhrstrøm, 2002] Peter Øhrstrøm, editor. *Time, Reality and Transcendence in Rational Perspective.* Aalborg University Press, Aalborg, 2002.

[Peirce, 1931 1958] Charles Sanders Peirce. *Collected Papers of Charles Sanders Peirce*, volume I–VIII. Harvard University Press, 1931–1958. Edited by C. Hartshorne, P. Weiss, and A. Burke.

[Peirce, 1976] Charles Sanders Peirce. *New Elements of Mathematics*, volume II. Humanities Press, 1976. Edited by Carolyn Eisele.

[Perloff, 1991] M. Perloff. Stit and the language of agency. *Synthese*, 86:379–408, 1991.

[Perloff, 1995] M. Perloff. Stit and the imperative. *American Philosophical Quarterly*, 32:71–81, 1995.

[Placek and Butterfield, 2002] T. Placek and J. Butterfield, editors. *Non-locality and Modality*, volume 64 of *NATO Science Series II: Mathematics, Physics and Chemistry.* Kluwer Academic Publishers, Dordrecht, 2002.

[Poidevin and MacBeath, 1993] Robin Le Poidevin and Murray MacBeath, editors. *The Philosophy of Time.* Oxford University Press, 1993.

[Poidevin, 1996] Robin Le Poidevin. Time, tense and topology. *Philosophical Quarterly*, 46:467–481, 1996.

[Prigogine, 1980] Ilya Prigogine. *From Being to Becoming, Time and Complexity in the Physical Sciences.* W.H. Freeman & Co., San Francisco, CA, 1980.

[Prior and Fine, 1977] A.N. Prior and Kit Fine. *Worlds, Times and Selves.* University of Massachusetts Press/Duckworth, Amherst, MA/London, 1977. Based on manuscipts by Prior. Edited by, and with a preface and postscript by Kit Fine.

[Prior, 1957] A.N. Prior. *Time and Modality.* Oxford University Press, Oxford, 1957.

[Prior, 1967] A.N. Prior. *Past, Present and Future.* Clarendon Press, Oxford, 1967.

[Prior, 1968] A.N. Prior. *Papers on Time and Tense.* Oxford University Press, Oxford, 1968.

[Prior, 1972] A.N. Prior. The notion of the present. 1972. In [Fraser *et al.*, 1972, pp. 320–323].

[Prior, 1996] A.N. Prior. Some free thinking about time. 1996. In [Copeland, 1996, p. 47–51]. Box 7, Bodleian Library.

[Prior, 2003] A.N. Prior. *Papers on Time and Tense.* Oxford University Press, second edition, 2003. Edited by Per Hasle, Peter Øhrstrøm, Torben Braüner, and Jack Copeland.

[Prokhovnik, 1985] S.J. Prokhovnik. *Light in Einstein's Universe.* Reidel, 1985.

[Rakić, 1997] Nataša Rakić. *Common Sense Time and Special Relativity.* PhD thesis, University of Amsterdam, 1997.

[Reichenbach, 1947] H. Reichenbach. *Elements of Symbolic Logic.* Macmillan, New York, 1947.

[Rescher and Urquhart, 1971] N. Rescher and A. Urquhart. *Temporal Logic*. Springer, 1971.

[Reynolds, 2003] M. Reynolds. An axiomatization of Prior's Ockhamist logic of historical necessity. 2003. In [Balbiani *et al.*, 2003, pp. 355–370].

[Röper, 1980] Peter Röper. Intervals and tenses. *Journal of Philosophical Logic*, 9:451–469, 1980.

[Russell, 1914] Bertrand Russell. *Our Knowledge of the External World*. Allen and Unwin, London, 1914.

[Russell, 1936] Bertrand Russell. On order in time. 1936. In [Marsh, 1956, pp. 347–363].

[Sylvan, 1996] Richard Sylvan. Other withered stumps of time. 1996. In [Copeland, 1996, pp. 111–130].

[Thomason, 1981a] Richmond H. Thomason. Deontic logic and the rôle of freedom in moral deliberations. 1981. In [Hilpinen, 1981, pp. 177–186].

[Thomason, 1981b] Richmond H. Thomason. Deontic logic as founded on tense logic. 1981. In [Hilpinen, 1981, pp. 165–176].

[Thomason, 1984] S.K. Thomason. On constructing instants from events. *Journal of Philosophical Logic*, 13:85–96, 1984.

[Tuggy, 1999] Dale Tuggy. A short text of Lavenham. 1999. In [Wegener, 1999a, 260–264]. See also *A Tract on Future Events*. Translation and notes by Dale Tuggy, Brown University, http://www.fredonia.edu/department/Philosophy/Lavenham.pdf.

[Walker, 1947] A.G. Walker. Durées et instants. *La Revue Scientifique*, (3266):131 ff., 1947.

[Wegener, 1999a] Mogens Wegener, editor. *Time, Creation and World-Order*. Aarhus University Press, Aarhus, 1999.

[Wegener, 1999b] Mogens Wegener. Towards a new metaphysics of time and creation. 1999. In [Wegener, 1999a, pp. 237–258].

[Wegener, 2000] Mogens Wegener. Ideas of cosmology, a philosopher's synthesis. 2000. In [Duffy and Wegener, 2000, pp. 255–274].

[Wegener, 2004] Mogens Wegener. The idea of a cosmic time. *Foundations of Physics*, 34(11):1777–1799, November 2004.

[Whitrow, 1980] G.J. Whitrow. *The Natural Philosophy of Time*. Clarendon Press, Oxford, 2nd edition, 1980.

[Wright, 1951] Georg Henrik von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[Xu, 1997] Ming Xu. Causation in branching time (1): Transitions, events and causes. *Synthese*, 112:137–192, 1997.

[Zanardo, 1990] Alberto Zanardo. Axiomatization of 'Peircean' branching-time logic. *Studia Logica*, 49:183–195, 1990.

[Zanardo, 2003a] Alberto Zanardo. Branching-time as a relative closeness relation among histories, 2003. To appear in the *Proceedings of The Fifth Tbilisi Symposium on Language, Logic and Computation*, Tbilisi (Georgia), 6–10 October, 2003.

[Zanardo, 2003b] Alberto Zanardo. First-order and second-order aspects of branching-time semantics, 2003. To appear in the *Proceedings of HPLMC-02: Second International Workshop on the History and Philosophy of Logic, Mathematics and Computation*, Donostia – SanSebastian, 7–9 November, 2002. Preprint Dipartimento di Matematica Universit di Padova, N. 03, 17 Febbraio 2003.

[Zanardo, 2003c] Alberto Zanardo. Moment/history duality in Prior's logics of branching-time. 2003. To appear in [Braüner *et al.*, 2005].

THE GAMUT OF DYNAMIC LOGICS

Jan van Eijck and Martin Stokhof

1 INTRODUCTION

Notions involving change often have a dual character, an interplay between process and product. While travelling from one place to another, one can either focus on the process of 'being on the road' or on the result of this process, 'being somewhere else'. Intellectual activities also have this dual nature: *scientific discovery* denotes a process of reaching for new insights but also the resulting insights, *judgement* denotes both the process of reaching a rational decision and the decision that results from that process, *computation* involves a process of stepwise changes, and the outcome of such a process, and so on.

The logical study of the interplay between process and product is called dynamic logic. This paper gives an overview of various systems of dynamic logic, with illustrations drawn from various application areas: programming, communicative action and interaction, cognitive processing, natural language understanding. It is aimed at researchers who have an interest in the formal analysis of computational and communicative processes. A more extended textbook introduction to dynamic logic that is explicitly geared to computer science is the informative [Harel *et al.*, 2000]. An earlier overview is [Harel, 1984]. Cf. also [van Benthem, 1996] for an introduction that focuses on cognitive applications.

Dynamic logic can be viewed as dealing with the logic of action and the result of action, and it can be used to model various kinds of actions and their results. A rough classification might be the following. First of all there are *computations*, i.e. actions performed on computers. Examples are computing the factorial function, computing square roots, etc. Such actions typically involve changing the memory state of a machine. Another type of action is that of *communicative actions*, such as reading an English sentence and updating one's state of knowledge accordingly, engaging in a conversation, sending an email with cc's, telling one's husband a secret. These actions typically change the cognitive states of the agents involved. And then there are *actions in the world*, such as building churches, destroying bridges, spilling milk. Such actions change the state of the world. Of course there are connections between these categories and actions of a mixed nature: a communicative action will usually involve some computation involving memory, and the utterance of an imperative is a communicative action that aims at an action in the world.

For a researcher who is interested in the formal analysis of actions of various kinds dynamic logic can be viewed as a tool box: it provides concepts and methods for description of actions and means to characterise the properties of the resulting systems. Using these tools the researcher can then develop specialised, tailored systems for dealing with specific kinds of actions: logics of computation, logics of communication, logics of action. Inasmuch as they are geared toward specific applications such systems may differ quite widely, but in many cases their core can nevertheless be characterised formally in a uniform way: many of these logics can be related to some variety of modal logic, taken in a suitably broad sense, viz., as the logic of 'labelled transition systems'.

A labelled transition system (or LTS, or multi-modal Kripke model) over signature $\langle P, A \rangle$, with P a set of propositions and A a set of actions, is a triple $\langle S, V, R \rangle$ where S is a set of states, $V : S \to \mathcal{P}(P)$ is a valuation function, and $R = \{ \overset{a}{\to} \subseteq S \times S \mid a \in A \}$ is a set of labelled transitions, i.e. binary relations on S, one for each label a. Let us illustrate the idea of an LTS by a few simple examples.

If one interprets the labelled transitions as the changes in the memory state of a computer, LTSs model computations, for example the simple assignment $x := y$:
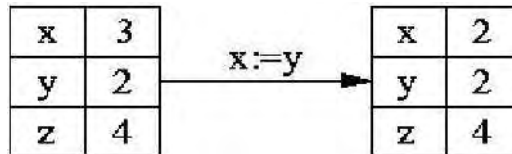
The command to put the value of register y in register x makes the contents of registers x and y equal. Pioneer papers in the logic of computation are [Floyd, 1967; Hoare, 1969].

If one interprets the labelled transitions as accessibility relations on the cognitive state space of a group of agents, LTSs can be used to model the information that such agents have about the world, about each other's information about the world, each other's information about each other's information about the world, and so on. And it can be used to describe changes in such information states:
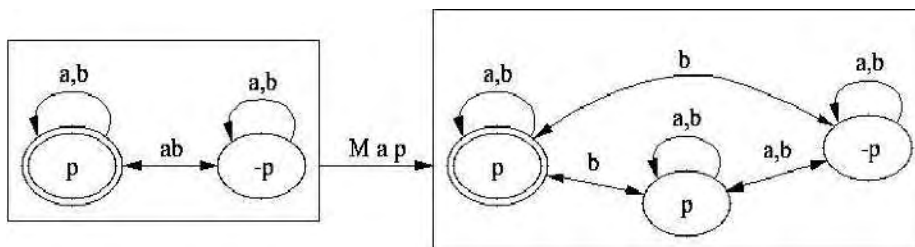
On the left is an epistemic situation where p is in fact the case (indicated by a double circle), but a and b cannot distinguish between p and $\neg p$. If in such a situation a receives the message that p is the case, while b is not informed of this, the epistemic situation changes to what is pictured on the right. In the new situation, a knows that p, and a also is aware of the fact that b does not know, while b still does not not know, and b still assumes that a does not know. See [Hintikka, 1962] for one of the earliest treatments of epistemic logic along these lines. An overview of the development of epistemic logic is given in [Gochet and Gribomont, 2005]. Cf., also [van Benthem, 1996].

Communicative actions may provide more detailed information about the world than the information that a certain state of affairs is realised. In a discourse (text, conversation), information is (often) conveyed piecemeal, and languages contains various means for keeping track of what has been said about what. Anaphoric pronouns are a case in point. Their role can be modelled by interpreting states as consisting of discourse items to which information is added in an incremental fashion. The following illustrates the action on such a state that is triggered by the use of an anaphoric pronoun:
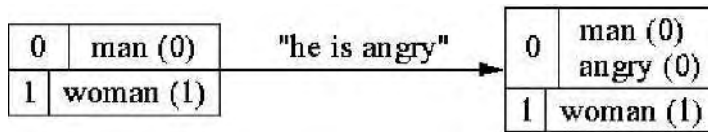
In a discourse where a man and a woman have been mentioned recently, an utterance of 'He is angry' receives a natural interpretation by linking the pronoun to the most salient appropriate discourse item, viz., the man that was just mentioned. Early work in this area is in [Karttunen, 1976; Heim, 1982; Kamp, 1981]. See [Gochet, 2002] for an overview.

Yet another illustration of how LTSs can be used to model action is when one interprets labelled transitions as actions on the state of the world. In that case LTSs model changes in the world itself:
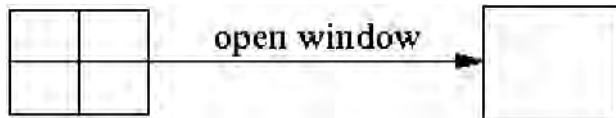
The action of window-opening changes a state in which the window is closed into one in which it is open. More complex actions call for more complex models, of course, in particular when we are interested in a more fine grained analysis of the causality involved in bringing about changes. An early overview of the logic of action is in [Wright, 1983]. For a more recent survey, cf., [Segerberg, 1992]. A different approach is the *stit*-logic of Belnap, cf. [Belnap *et al.*, 2001].

These examples illustrate that it is possible to approach a wide variety of kinds of actions from a unified perspective. What follows is intended to show that this is not only possible, but also fruitful. Note that the diversity of applications of dynamic logic also indicates that it is difficult to trace the various systems and application to a single historic root. In fact, some of what appears uniform now, as a matter of historical fact had quite diverse origins. For this reason we have opted for a mainly systematic treatment, with occasional historical side remarks where relevant.

*

The larger part of the survey of dynamic logic that follows is devoted to an exposition of two core systems of dynamic logic, viz., *propositional dynamic logic* and *quantificational dynamic logic*, and three illustrative areas of application, viz., programming, communicative action and dynamic semantics of natural language.

One of the seminal papers in computer science is Hoare's [Hoare, 1969]. where the following notation is introduced for specifying what an imperative program does:

$$\{P\} \quad C \quad \{Q\}.$$

Here C is a program from a formally defined programming language for imperative programming, and P and Q are conditions on the programming variables used in C. Statement $\{P\} C \{Q\}$ is true if whenever C is executed in a state satisfying P and if the execution of C terminates, then the state in which execution of C terminates satisfies Q. The 'Hoare-triple' $\{P\} C \{Q\}$ is called a partial correctness specification; P is called its precondition and Q its postcondition. Floyd-Hoare logic, as the logic of reasoning with such correctness specifications is called, is the precursor of all the dynamic logics known today. We will demonstrate Floyd-Hoare logic in Section 2.4, for the toy language specified in Section 2.1. The specification of a toy programming language has the additional benefit that it will allow us to demonstrate various approaches to the semantics of programming. We will present example programs, formulate questions about their behaviour, and show how some of these questions are answered with Floyd-Hoare logic. After that, we turn to dynamic logic proper as a more general means of tackling such questions.

In section 3 we present what is perhaps the most basic system of dynamic logic, propositional dynamic logic (PDL), a logic in which basic actions are primitives. This feature makes PDL applicable in a wide variety of cases. For example, if one interprets the basic actions as communicative actions that affect cognitive states of sets of interacting agents, then dynamic logic takes the shape of dynamic epistemic logic. This important area of application is treated in detail in section 4.

When one takes memory change as the basic action, one gets quantified dynamic logic (QDL), the system that is introduced and discussed in section 5. QDL has its origin in correctness reasoning based on annotating programs with pre- and postconditions. These historical connections are briefly traced. It is possible to

interpret QDL programs also in a different way, viz., as changing the cognitive state of a language user. This potential relevance of QDL for an understanding of natural language was actualised in what has been called the 'dynamic turn' in natural language semantics. In section 6 we focus on dynamic predicate logic (DPL) as a subsystem of QDL. A more detailed treatment of the application of dynamic concepts in natural language semantics is given in section 7.

2 DESCRIBING CHANGE AND REASONING ABOUT CHANGE

Consider the following problem concerning the outcome of a pebble drawing action.

> A vase contains 35 white pebbles and 35 black pebbles. Proceed as follows to draw pebbles from the vase, as long as this is possible. Every round, draw two pebbles from the vase. If they have the same colour, then put a black pebble back into the vase, if they have different colours, then put the white pebble back. You may assume that there are enough additional black pebbles. In every round one pebble is removed from the vase, so after 69 rounds there is a single pebble left. What is the colour of this pebble?

Here is an implementation of this procedure, where the vase is represented as a list of integers, the white pebbles are the occurrences of 0, and the black pebbles the occurrences of 1. The `draw` function is coded in the programming language Haskell [Jones, 2003]:

```
draw :: [Integer] -> [Integer]
draw [x] = [x]
draw (0:0:xs) = draw (1:xs)
draw (1:1:xs) = draw (1:xs)
draw (0:1:xs) = draw (0:xs)
draw (1:0:xs) = draw (0:xs)
```

The question: if this function is called with a list of thirty-five 0's and thirty-five 1's, in unknown order, will the outcome of the function be [0] or [1]?

The key to the solution is finding an *invariant* of the procedure, i.e. finding a condition that does not change when a single pebble is removed from the vase. It is not hard to see that when a pebble is drawn, the number of white pebbles always remains odd. It follows that the last pebble is white. So the `draw` function will return [0] on any permutation of the list of thirty-five 0's and thirty-five 1's.

With this piece of reasoning we are in the realm of dynamic logic. Rather than encode examples in an existing programming language like *Haskell* or *Java*, it will turn out to be useful to introduce our own toy language for illustrations. As dynamic logic describes the interplay between actions and resulting states, the action description language is part and parcel of the dynamic logic language.

2.1   The WHILE Language

In what follows we define a simple programming language for programming over the data type of the natural numbers, i.e. the set $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, with functions $+$ for addition, $*$ for product, and $\dot{-}$ for cut-off subtraction.

First, we distinguish between numbers and their names. Numbers are objects in the mathematical realm, names are syntactic objects. A numeral is a name for a natural number. E.g., '5' is a name for the natural number 5. Assume N is a set of numerals. Assume V is a set of variables. The sets N and V may have further internal structure, but we will not bother to spell this out. Given sets N, V, arithmetic expressions can be defined by means of $+, *, \dot{-}$, as follows (assume n ranges over the numerals and v over the variables):

$$a \quad ::= \quad n \mid v \mid a_1 + a_2 \mid a_1 * a_2 \mid a_1 \dot{-} a_2.$$

This says that $345 * (67 + 8)$ and $(345 * 67) + 8$ are arithmetic expressions. (The brackets indicate the manner of construction).

In terms of these arithmetic expressions we will now fix a small programming language for programming with the natural numbers. We assume two further primitive relation symbols '$=$' for 'equal', and '$\leq$' for 'less than or equal'. This allows us to define Boolean expressions (named after [Boole, 1854]), as follows:

$$B \quad ::= \quad \top \mid a_1 = a_2 \mid a_1 \leq a_2 \mid \neg B \mid B_1 \vee B_2$$

Note that instead of listing equalities $a_1 = a_2$ explicitly, we might have introduced them by way of abbreviation, as shorthand for $a_1 \leq a_2 \wedge a_2 \leq a_1$. Arithmetic expressions and Boolean expressions figure in programming commands, as follows:

$$C \quad ::= \quad \text{SKIP} \mid v := a \mid C_1 \; ; \; C_2 \mid \text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2 \mid \text{WHILE } B \text{ DO } C.$$

The basic programming constructs of the WHILE language are SKIP for the program that does nothing, and $v := a$ for the program that assigns the value of a to the variable v. Programs or commands can be composed by means of sequencing, by means of conditionalisation, and by means of guarded repetition. Further programming constructs can now be defined, e.g., *REPEAT*:

$$\text{REPEAT } C \text{ UNTIL } B \quad := \quad C \; ; \; \text{WHILE } \neg B \text{ DO } C.$$

The WHILE language looks deceptively simple, but it is extremely expressive. In fact, this little language is Turing complete, i.e. one can specify the behaviour of any Turing machine in it ([Turing, 1936]). This means that anything that can be computed on the natural numbers can (in principle) be computed by means of a WHILE program.

2.2   Semantics

To specify the semantics, we take the natural numbers $\mathbb{N}$ with the operations $+, *, \dot{-}$ and the relation $\leq$ as given. We also assume that every numeral n in N

has an interpretation $I(n) \in \mathbb{N}$. Let g be a mapping from V to $\mathbb{N}$ (an assignment of natural numbers to the variables). The arithmetic expressions of the language are now interpreted relative to assignment g, as follows:

$$
\begin{aligned}
[\![n]\!]_g &:= I(n) \\
[\![v]\!]_g &:= g(v) \\
[\![a_1 + a_2]\!]_g &:= [\![a_1]\!]_g + [\![a_2]\!]_g \\
[\![a_1 * a_2]\!]_g &:= [\![a_1]\!]_g * [\![a_2]\!]_g \\
[\![a_1 \dot{-} a_2]\!]_g &:= [\![a_1]\!]_g \dot{-} [\![a_2]\!]_g
\end{aligned}
$$

The semantics of the Boolean expressions (or 'Booleans') of the language is defined as follows:

$$
\begin{aligned}
[\![\top]\!]_g &:= T \\
[\![a_1 = a_2]\!]_g &:= \begin{cases} T & \text{if } [\![a_1]\!]_g = [\![a_2]\!]_g \\ F & \text{otherwise} \end{cases} \\
[\![a_1 \leq a_2]\!]_g &:= \begin{cases} T & \text{if } [\![a_1]\!]_g \leq [\![a_2]\!]_g \\ F & \text{otherwise} \end{cases} \\
[\![\neg B]\!]_g &:= \begin{cases} T & \text{if } [\![B]\!]_g = F \\ F & \text{otherwise} \end{cases} \\
[\![B_1 \vee B_2]\!]_g &:= \begin{cases} T & \text{if } [\![B_1]\!]_g = T \text{ or } [\![B_2]\!]_g = T \\ F & \text{otherwise} \end{cases}
\end{aligned}
$$

Natural Semantics for Commands

The semantics of the commands can be given in various styles. First we give the so-called natural semantics, in the form of a specification of a transition system.

For any valuation g, any variable v and any natural number d, let $g[v \mapsto d]$ be the valuation g' that differs from g at most in the fact that $g'(v) = d$. This notion is familiar from the semantics of first order logic. Then the transition for assignment commands is given by:

$$
g \xrightarrow{\;v := a\;} g[v \mapsto [\![a]\!]_g]
$$

The SKIP command does nothing:

$$
g \xrightarrow{\;\text{SKIP}\;} g
$$

Sequential composition combines two transition arrows:

$$
\frac{g \xrightarrow{\;C_1\;} g' \quad g' \xrightarrow{\;C_2\;} g''}{g \xrightarrow{\;C_1 \,;\, C_2\;} g''}
$$

Conditional action makes a choice from two transition relations, depending on the evaluation of the condition.

$$\frac{g \xrightarrow{C_1} g'}{g \xrightarrow{\text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2} g'} \ [\![B]\!]_g = T$$

$$\frac{g \xrightarrow{C_2} g'}{g \xrightarrow{\text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2} g'} \ [\![B]\!]_g = F$$

Guarded iteration does nothing if the guard fails to hold:

$$\frac{}{g \xrightarrow{\text{WHILE } B \text{ DO } C} g} \ [\![B]\!]_g = F$$

Otherwise the guarded action is performed and the WHILE command is executed again in the result state.

$$\frac{g \xrightarrow{C} g' \quad g' \xrightarrow{\text{WHILE } B \text{ DO } C} g''}{g \xrightarrow{\text{WHILE } B \text{ DO } C} g''} \ [\![B]\!]_g = T$$

These rules define a transition relation $\xrightarrow{C}$ on the set of all valuations, for every command C. In order to derive a transition $g \xrightarrow{C} g'$, construct a finite derivation tree with $g \xrightarrow{C} g'$ at the root, with axioms at the leaves and each internal nodes licensed by a transition rule. Here is an example, for the command $z := x \ ; \ x := y \ ; \ y := z$, executed in the state $g = \{x \mapsto 3, y \mapsto 2, z \mapsto 5\}$. We use g_1 as shorthand for $\{x \mapsto 3, y \mapsto 2, z \mapsto 3\}$, g_2 as shorthand for $\{x \mapsto 2, y \mapsto 2, z \mapsto 3\}$, g_3 as shorthand for $\{x \mapsto 2, y \mapsto 3, z \mapsto 3\}$.

$$\frac{g \xrightarrow{z:=x} g_1 \quad \dfrac{g_1 \xrightarrow{x:=y} g_2 \quad g_2 \xrightarrow{y:=z} g_3}{g_1 \xrightarrow{x:=y \ ; \ y:=z} g_3}}{g \xrightarrow{z:=x \ ; \ x:=y \ ; \ y:=z} g_3}$$

This command computes the remainder upon division of x by y in x:

$$\text{WHILE } y \leq x \text{ DO } x := x \dot{-} y.$$

The following variant computes the result of the division of x by y in z, and the remainder in x:

$$z := 0 \ ; \ \text{WHILE } y \leq x \text{ DO } (x := x \dot{-} y \ ; \ z := z + 1).$$

Abbreviate $\neg a_1 = a_2$ as $a_1 \neq a_2$, $\neg a_1 \leq a_2$ as $a_1 > a_2$ and $\neg a_1 \geq a_2$ as $a_1 < a_2$. Euclid's well known Greatest Common Divisor algorithm is now readily expressed as a WHILE command. The following program computes the GCD of x and y in x (and in y).

(1)   WHILE $x \neq y$ DO IF $x > y$ THEN $x := x \dot{-} y$ ELSE $y := y \dot{-} x$.

For state $g = \{x \mapsto 24, y \mapsto 9\}$, program (1) leads to the following execution:

$$\{x \mapsto 24, y \mapsto 9\} \quad x := x \dot{-} y \quad \{x \mapsto 15, y \mapsto 9\}$$
$$x := x \dot{-} y \quad \{x \mapsto 6, y \mapsto 9\}$$
$$y := y \dot{-} x \quad \{x \mapsto 6, y \mapsto 3\}$$
$$x := x \dot{-} y \quad \{x \mapsto 3, y \mapsto 3\}.$$

Consider the following command:

(2) $y := 1$;   WHILE $x \neq 1$ DO $(y := y * x$; $x := x \dot{-} 1)$.

Let g be a valuation with $g(x) = 3$. Then one can use the transition rules to show:

$$g \xrightarrow{\;y:=1\;;\;\;\text{WHILE } x \neq 1 \text{ DO } (y:=y*x\;;\;x:=x\dot{-}1)\;} g[x \mapsto 1, y \mapsto 6].$$

When executed in a state g, command (2) computes the factorial of $g(x)$ in y.

We say that a command C *terminates* in state g if there is a state g' with $g \xrightarrow{C} g'$, and that C *loops* in state g if C does not terminate in state g. It can be shown by induction that it holds for all C that if $g \xrightarrow{C} g'$ and $g \xrightarrow{C} g''$ then $g' = g''$ (WHILE programs are *deterministic*).

In simple cases it is easy to say whether a command terminates in a given state. For example, the factorial command terminates for all states g, and the command

$$\text{WHILE } x > 0 \text{ DO } x := x + 1$$

loops for all states g with $g(x) \neq 0$. In general, however, termination of WHILE programs for infinite state sets is undecidable. As an example of a difficult decision problem about program termination, take the question whether the following program terminates for all states with positive x:

$$\text{WHILE } x \neq 1 \text{ DO   IF   even } (x) \text{ THEN } x := x/2 \text{ ELSE } x := (3 * x) + 1$$

Note that this example uses an operator / for integer division and a predicate for evenness, but this is not crucial, for these extensions are definable in the WHILE language. Here is an example run of the program:

$$
\begin{aligned}
&& x_0 &= 7 \\
x := (3 * x) + 1 &\ \to\ & x_1 &= 22 \\
x := x/2 &\ \to\ & x_2 &= 11 \\
x := (3 * x) + 1 &\ \to\ & x_3 &= 34 \\
x := x/2 &\ \to\ & x_4 &= 17 \\
x := (3 * x) + 1 &\ \to\ & x_5 &= 52 \\
x := x/2 &\ \to\ & x_6 &= 26 \\
x := x/2 &\ \to\ & x_7 &= 13 \\
x := (3 * x) + 1 &\ \to\ & x_8 &= 40 \\
x := x/2 &\ \to\ & x_9 &= 20 \\
x := x/2 &\ \to\ & x_{10} &= 10 \\
x := x/2 &\ \to\ & x_{11} &= 5 \\
x := (3 * x) + 1 &\ \to\ & x_{12} &= 16 \\
x := x/2 &\ \to\ & x_{13} &= 8 \\
x := x/2 &\ \to\ & x_{14} &= 4 \\
x := x/2 &\ \to\ & x_{15} &= 2 \\
x := x/2 &\ \to\ & x_{16} &= 1
\end{aligned}
$$

Counterexamples against termination have never been found, but a proof of termination has not been found either. This termination problem was posed by Lothar Collatz in 1937, and it is still open [Guy, 1981, Problem E 16].

Structural Operational Semantics for Commands

An alternative fashion of specifying the semantics of an imperative programming language, due to Plotkin [Plotkin, 1981], specifies the transition system for a program in a slightly different way, focusing on the smallest steps that a computation can take. Here are the rules of what is called 'structural operational semantics', or 'small step semantics'. The transitions are now from pairs of a state and a command to a state (such a transition expresses that the command finishes in a single step), and from pairs of a state and a command to a new state and a new command (such a transition expresses that the first step of the command causes a shift to the new state, where the remainder of the command is left to be executed).

Assignment commands finish in one step:

$$(g, v := a) \ \implies\ g[v \mapsto [\![a]\!]_g].$$

The SKIP command also finishes in a single step, and it does not change the state.

$$(g, \text{SKIP}) \ \implies\ g.$$

If the first command of a command sequence finishes in a single step, then the second command of the sequence is all that is left:

$$\frac{(g, C_1) \implies g'}{(g, C_1 \; ; \; C_2) \implies (g', C_2)}$$

If the first command of a command sequence does not finish in a single step, we get:

$$\frac{(g, C_1) \implies (g', C_1')}{(g, C_1 \; ; \; C_2) \implies (g', C_1' \; ; \; C_2)}$$

Rules for conditional action: the action depends on the outcome of the test.

$$\frac{}{(g, \text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2) \implies (g, C_1)} \; [\![B]\!]_g = T$$

$$\frac{}{(g, \text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2) \implies (g, C_2)} \; [\![B]\!]_g = F$$

Finally, the guarded iteration command. If the guard is not satisfied, the command finishes in a single step, and it does not change the state:

$$\frac{}{(g, \text{WHILE } B \text{ DO } C) \implies g} \; [\![B]\!]_g = F$$

Otherwise the first step of the guarded action is performed, and in the result state the remainder of the action plus the conditional iteration command are put on the to-do list:

$$\frac{(g, C) \implies (g', C')}{(g, \text{ WHILE } B \text{ DO } C) \implies (g', C'; \text{WHILE } B \text{ DO } C)} \; [\![B]\!]_g = T$$

To see how this works, consider the command $z := x \; ; \; x := y \; ; \; y := z$, executed in the state $g = \{x \mapsto 3, y \mapsto 2, z \mapsto 5\}$. The structural operational semantics rules yield the following:

$$
\begin{aligned}
& (\{x \mapsto 3, y \mapsto 2, z \mapsto 5\}, z := x \; ; \; x := y \; ; \; y := z) \\
\implies & (\{x \mapsto 3, y \mapsto 2, z \mapsto 3\}, x := y \; ; \; y := z) \\
\implies & (\{x \mapsto 2, y \mapsto 2, z \mapsto 3\}, y := z) \\
\implies & \{x \mapsto 2, y \mapsto 3, z \mapsto 3\}
\end{aligned}
$$

It can now be proved by induction that these rules define the same 'extensional' behaviour as the original rules, in the sense that $g \xrightarrow{C} g'$ iff $(g, C) \implies^* g'$.

The difference between natural semantics (large step semantics) and structural operational semantics (small step semantics) shows up as soon as we add a construct for error abortion to the language. Suppose ABORT is a program that in any state g stops execution without yielding a new output state. Then the difference between SKIP and ABORT is that we have $(g, \text{SKIP}) \implies g$ and $g \xrightarrow{\text{SKIP}} g$, while from (g, ABORT) there are no $\implies$ arrows, and there are no

states g' with $g \xrightarrow{\text{ABORT}} g'$. It turns out that in natural semantics there is no way to distinguish between abnormal termination and looping behaviour, while in structural operational semantics there is. In natural semantics, ABORT and WHILE $\top$ DO SKIP are equivalent, but in structural operational semantics they are not, for the first has no derivation sequence at all, while the second has an infinite one:

$$
\begin{aligned}
(g, \text{WHILE } \top \text{ DO SKIP}) &\implies (g, \text{WHILE } \top \text{ DO SKIP}) \\
&\implies (g, \text{WHILE } \top \text{ DO SKIP}) \\
&\implies \ldots
\end{aligned}
$$

The natural semantics can be made more expressive by adding a special error state $\bullet$ different from all the regular states, and adding the transition rules $g \xrightarrow{\text{ABORT}} \bullet$, and $\bullet \xrightarrow{C} \bullet$ for all commands C. Under this modification ABORT and WHILE $\top$ DO SKIP become distinguishable again in natural semantics, for the first has a transition to $\bullet$ from anywhere, and the second has no transitions from anywhere.

Interpreted versus Uninterpreted Semantics

The WHILE language over $\mathbb{N}$ is an example of an interpreted language. We can also choose to interpret WHILE over different data structures. To see that this makes a difference, consider the following program:

$$
\text{WHILE } x \neq 0 \text{ DO } x := p(x)
$$

If p is interpreted as predecessor, this program will always terminate when executed on $\mathbb{N}$, but it will only terminate for states with a non-negative value for x when executed on $\mathbb{Z}$ (the domain of integers). As another example, let $\mathcal{T}$ be the infinite binary tree given by:

$$
\mathcal{T} \quad ::= \quad \langle\rangle \mid \mathcal{T}0 \mid \mathcal{T}1
$$

with a unary function $\uparrow :: \mathcal{T} \to \mathcal{T}$ defined by means of

$$
\uparrow\langle\rangle = \langle\rangle, \uparrow\mathcal{T}0 = \uparrow\mathcal{T}1 = \mathcal{T}.
$$

This specifies the following infinite binary tree:
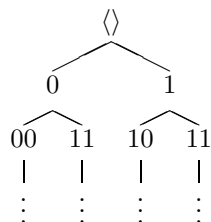
Then the following WHILE program over $\mathcal{T}$

$$\text{WHILE } x \neq \langle\rangle \wedge y \neq \langle\rangle \text{ DO } (x := \uparrow x \; ; \; y := \uparrow y)$$

will always terminate in a state where $x = \langle\rangle$ or $y = \langle\rangle$, depending on which of x, y is closer to the root $\langle\rangle$ in the initial state.

WHILE programs can also be studied under the aspect of uninterpreted computation. Given a first order signature σ, we may be interested in equivalence of WHILE programs for arbitrary σ models. E.g., the commands

$$\text{IF } B \text{ THEN } C_1 \text{ ELSE } C_2$$

and

$$\text{IF } \neg B \text{ THEN } C_2 \text{ ELSE } C_1$$

are equivalent for any choice of B, C_1, C_2 and any model $\mathcal{M}$ for the predicate and function symbols that occur in B, C_1, C_2. Uninterpreted reasoning is the right level for comparing expressive power of programming language constructs, for on the fixed domain $\mathbb{N}$ with zero, successor, addition and multiplication all reasonable programming language have the same expressive power: they all compute exactly the partial recursive functions. At the uninterpreted level, extending the WHILE language with a construct for non-deterministic choice C_1 OR C_2 strictly increases expressive power.

2.3 Non-determinism

Non-deterministic WHILE is the extension of WHILE with a construct for choice C_1 OR C_2, with semantics given by the following transition rules:

$$\frac{g \xrightarrow{\;C_1\;} g'}{g \xrightarrow{\;C_1 \text{ OR } C_2\;} g'}$$

$$\frac{g \xrightarrow{\;C_2\;} g'}{g \xrightarrow{\;C_1 \text{ OR } C_2\;} g'}$$

What this says is that a program like $x := x + 1$ OR $x := x + 2$, when executed in a state $\{x \mapsto 3\}$ will produce two output states $\{x \mapsto 4\}$ and $\{x \mapsto 5\}$.

The structural operational semantics rules for choice are as follows:

$$\frac{}{(g, C_1 \text{ OR } C_2) \implies (g, C_1)}$$

$$\frac{}{(g, C_1 \text{ OR } C_2) \implies (g, C_2)}$$

Now consider program (3).

(3) (WHILE $\top$ DO SKIP) OR $x := x + 2$.

According to the natural semantics, for no input state g is there an output state g' with $g \xrightarrow{\text{WHILE } \top \text{ DO SKIP}} g'$. Therefore, program (3) will only get one derivation tree, namely that for:

$$g \xrightarrow{\text{(WHILE } \top \text{ DO SKIP) OR } x:=x+2} g\{x \mapsto x + 2\}.$$

According to the structural operational semantics, we get two derivation sequences, one infinite

$$(g, \text{ (WHILE } \top \text{ DO SKIP) OR } x := x + 2)$$
$$\Longrightarrow (g, \text{ (WHILE } \top \text{ DO SKIP)})$$
$$\Longrightarrow (g, \text{ (WHILE } \top \text{ DO SKIP)})$$
$$\Longrightarrow \ldots$$

and the other finite

$$(g, \text{ (WHILE } \top \text{ DO SKIP) OR } x := x + 2)$$
$$\Longrightarrow (g, x := x + 2)$$
$$\Longrightarrow g\{x \mapsto x + 2\}.$$

This illustrates that the structural operational semantics is more 'fine-grained' than the natural semantics. It also shows that the presence of non-determinism may make looping behaviour more difficult to detect.

Programming language semantics in various styles for WHILE and its extensions are discussed in [Nielson and Nielson, 1992]. Classics on denotational semantics for programming are [Stoy, 1977] and [Schmidt, 1986].

2.4   Floyd-Hoare Logic

One way of reasoning about WHILE commands (or about imperative programs in general) is by using first order predicate logic for making assertions about command execution. Floyd [Floyd, 1967] and Hoare [Hoare, 1969] proposed to use *correctness statements* of the following form:

$$\{\varphi\} \, C \, \{\psi\}$$

This expresses that command C takes us from a precondition φ, true at the state where the command gets executed (the input state), to a postcondition ψ, true immediately after execution of the command. Since we are programming over the natural numbers, we interpret the pre- and postconditions in $\mathbb{N}$. This gives the following formal interpretation of Floyd-Hoare correctness triples:

$$\mathbb{N} \models \{\varphi\} \, C \, \{\psi\} \qquad \text{iff}$$
$$\text{for all } g, h, \quad \text{if } \mathbb{N} \models_g \varphi \text{ and } g \xrightarrow{C} h, \text{ then } \mathbb{N} \models_h \psi.$$

An example of a true correctness statement is the following:

$$\{x! = Z\} \, y := 1 \; ; \; \text{WHILE } x \neq 1 \text{ DO } (y := y * x \; ; \; x := x \dot{-} 1) \, \{y = Z\}$$

Figure 1. Floyd-Hoare Calculus for WHILE

assignment	$$\overline{\{\varphi_a^v\}\ v := a\ \{\varphi\}}$$
skip	$$\overline{\{\varphi\}\ \mathrm{SKIP}\ \{\varphi\}}$$
sequence	$$\frac{\{\varphi\}\ C_1\ \{\psi\} \qquad \{\psi\}\ C_2\ \{\chi\}}{\{\varphi\}\ C_1\ ;\ C_2\ \{\chi\}}$$
conditional choice	$$\frac{\{\varphi \wedge B\}\ C_1\ \{\psi\} \qquad \{\varphi \wedge \neg B\}\ C_2\ \{\psi\}}{\{\varphi\}\ \text{if } B \text{ then } C_1 \text{ else } C_2\ \{\psi\}}$$
guarded iteration	$$\frac{\{\varphi \wedge B\}\ C\ \{\varphi\}}{\{\varphi\}\ \text{while } B \text{ do } C\ \{\varphi \wedge \neg B\}}$$
precondition strengthening	$$\frac{\mathbb{N} \models \varphi' \rightarrow \varphi \qquad \{\varphi\}\ C\ \{\psi\}}{\{\varphi'\}\ C\ \{\psi\}}$$
postcondition weakening	$$\frac{\{\varphi\}\ C\ \{\psi\} \qquad \mathbb{N} \models \psi \rightarrow \psi'}{\{\varphi\}\ C\ \{\psi'\}}$$

In connection with Floyd-Hoare style correctness assertions, the notions of *strongest postcondition* and *weakest liberal precondition* arise in a natural way.

The strongest postcondition $\mathrm{SP}(\varphi, C)$ of a predicate logical formula φ and a command C is the condition that holds in a state g if there is a state h satisfying φ that has a C transition to g. Formally:

$$\mathbb{N} \models_g \mathrm{SP}(\varphi, C) \text{ iff there is an } h \text{ with } \mathbb{N} \models_h \varphi \text{ and } h \xrightarrow{C} g.$$

The weakest liberal precondition $\mathrm{WLP}(C, \varphi)$ of a predicate logical formula φ and a command C has the following interpretation:

$$\mathbb{N} \models_g \mathrm{WLP}(C, \varphi) \text{ iff there is an } h \text{ with } \mathbb{N} \models_h \varphi \text{ and } g \xrightarrow{C} h.$$

The connection with Floyd-Hoare correctness statements is as follows:

$$\mathbb{N} \models \{\varphi\} \ C \ \{\mathrm{SP}(\varphi, C)\},$$

$$\text{if } \mathbb{N} \models \{\varphi\} \ C \ \{\psi\} \text{ then } \mathbb{N} \models \mathrm{SP}(\varphi, C) \to \psi,$$

$$\mathbb{N} \models \{\mathrm{WLP}(C, \varphi)\} \ C \ \{\varphi\},$$

$$\text{if } \mathbb{N} \models \{\varphi\} \ C \ \{\psi\} \text{ then } \mathbb{N} \models \varphi \to \mathrm{WLP}(C, \psi).$$

This illustrates the view of WHILE programs as predicate transformers, mapping weakest precondition predicates on the natural numbers into strongest postcondition predicates on the natural numbers.

A Floyd-Hoare calculus

for WHILE programs is given in Figure 1. In the rule for assignment, φ_a^v denotes the result of substitution of a for v in φ. At first sight, one might think that the assignment axiom should run $\{\varphi\} \ v := a \ \{\varphi_a^v\}$ instead of $\{\varphi_a^v\} \ v := a \ \{\varphi\}$. This would be a mistake, for consider the example where φ equals the statement $v = 0$, and a equals $v + 1$. Then the rule $\{\varphi\} \ v := a \ \{\varphi_a^v\}$ yields the *incorrect* statement $\{v = 0\} \ v := v + 1 \ \{v + 1 = 0\}$, while the correct rule $\{\varphi_a^v\} \ v := a \ \{\varphi\}$ yields the correct statement $\{v + 1 = 0\} \ v := v + 1 \ \{v = 0\}$.

Note that the rules of precondition strengthening and postcondition weakening in $\mathbb{N}$ are a kind of oracle rules, for implications $\psi \to \psi'$ on the natural numbers may be undecidable.

Illustration   To illustrate the use of the calculus, consider the factorial program (2) again. Here are the correctness statements that prove the fact that this program actually computes the factorial function:

1. $\{x! = Z\} \ y := 1 \ \{y * x! = Z\}$

2. $\{y * x! = Z \wedge x \neq 0\} \ y := y * x \ \{y * x! = Z * x\}$

3. $\{y * x! = Z * x \wedge x \neq 0\} \ x := x \dot{-} 1 \ \{y * x! = Z\}$

4. $\{y * x! = Z \wedge x \neq 0\} \ y := y * x \ ; \ x := x \dot{-} 1 \ \{y * x! = Z\}$

5. $\{y * x! = Z\}$ WHILE $x \neq 0$ DO $(y := y * x \ ; \ x := x \dot{-} 1) \ \{y * x! = Z \wedge x = 0\}$.

6. $\{x! = Z\}$
   $y := 1 \ ; \ $ WHILE $x \neq 0$ DO $(y := y * x \ ; \ x := x \dot{-} 1)$
   $\{y * x! = Z \wedge x = 0\}$.

7. $\{x! = Z\}$
   $y := 1 \ ; \ $ WHILE $x \neq 0$ DO $(y := y * x \ ; \ x := x \dot{-} 1)$
   $\{y = Z\}$.

Properties

The Floyd-Hoare calculus for WHILE programs is sound, in the following sense: if $\{\varphi\} \ C \ \{\psi\}$ is derivable, using the rules for precondition strengthening and postcondition weakening in $\mathbb{N}$, then $\mathbb{N} \models \{\varphi\} \ C \ \{\psi\}$. Soundness is easily shown by induction on the length of Floyd-Hoare derivations.

The presence of the precondition strengthening and postcondition weakening introduce an element of *model checking* into the Floyd-Hoare calculus, making it into a hybrid tool for deduction and evaluation in $\mathbb{N}$.

Since arithmetical truth is not effectively axiomatisable, the true correctness statements for WHILE programs over $\mathbb{N}$ are not effectively axiomatisable either. Indeed, we have, for every arithmetical formula φ:

$$\mathbb{N} \models \varphi \text{ iff } \mathbb{N} \models \{\top\} \text{ SKIP } \{\varphi\}.$$

However, because strongest postconditions can be expressed in the language of $\mathbb{N}$ by means of encoding, we can get around this by allowing members of $\text{Th}(\mathbb{N})$ (the set of all predicate logical statements that are true on the natural numbers) in correctness proofs [Cook, 1978]:

THEOREM 1 (Cook, Relative Completeness). $\mathbb{N} \models \{\varphi\} \ C \ \{\psi\}$ *implies that* $\{\varphi\} \ C \ \{\psi\}$ *is derivable using Floyd-Hoare rules together with* $\text{Th}(\mathbb{N})$.

Proof. An induction on the structure of programs works. We just give the case of guarded iterations. Let $\mathbb{N} \models \{\varphi\}$ WHILE B DO C $\{\psi\}$. Now use the fact that strongest postconditions are encodable in $\mathbb{N}$ to define

$$\chi = \exists y_1 \cdots y_n (\text{SP}(\varphi, \text{ WHILE } B \wedge (x_1 \neq y_1 \vee \cdots \vee x_n \neq y_n) \text{ DO } C))$$

where $x_1, \ldots, x_n$ are all the variables occurring in C, and $y_1, \ldots, y_n$ are new. Then χ defines the states that can be reached from a φ state by means of a finite number of C transitions through B states. Thus, $\mathbb{N} \models \{\chi \wedge B\} \ C \ \{\chi\}$. This formula is derivable by the induction hypothesis. By the Floyd-Hoare rule for guarded iteration, it follows from this that

$$\{\chi\} \text{ WHILE } B \text{ DO } C \ \{\chi \wedge \neg B\}$$

is derivable too. Since $\varphi \to \chi$ and $\chi \wedge \neg B \to \psi$ are both true in $\mathbb{N}$ (the latter because $\chi \wedge \neg B$ is equivalent to $SP(\varphi, \text{ WHILE } B \text{ DO } C)$), by the rules for precondition strengthening and postcondition weakening we get that

$$\{\varphi\} \text{ WHILE } B \text{ DO } C \{\psi\}$$

must be derivable too.   ■

It is important to note that Floyd-Hoare correctness statements if this simple form are not expressive enough to reason about termination. The following correctness statement is true:

$\{x \geq 1\}$
WHILE $x \neq 1$ DO IF even (x) THEN $x := x/2$ ELSE $x := (3 * x) + 1$
$\{x = 1\}$

This expresses that *if* the command is executed in a state where x has a positive value, after termination x will have value 1. It does *not* express that the command will terminate for all states with x positive. This is the reason that Floyd-Hoare correctness statements are sometimes called partial correctness statements.

To remedy this, calculi have been proposed with a stronger interpretation, for reasoning about Floyd-Hoare triples expressing total correctness:

$$\{\varphi\} C \{\Downarrow \psi\}$$

Such a total correctness statement expresses that if precondition φ is fulfilled then C is guaranteed to terminate in a state satisfying ψ. To make this work, the rule for guarded iteration has to be reformulated in terms of a decreasing measure function M on the natural numbers, as follows (it is assumed that $\mathbb{N} \models (\varphi \wedge M = i+1) \to B$ and $\mathbb{N} \models (\varphi \wedge M = 0) \to \neg B$):

$$\frac{\{\varphi \wedge M = i + 1\} C \{\Downarrow \varphi \wedge M = i\}}{\{\exists i(\varphi \wedge M = i)\} \text{ WHILE } B \text{ DO } C \{\Downarrow \varphi \wedge M = 0\}}$$

An overview of the development of Floyd-Hoare reasoning can be found in [Apt, 1981]. Floyd-Hoare reasoning is still a dominant tradition in program verification; pre- and postcondition annotations can be used as formal specifications with respect to which a program can be verified, where the verification process can be partially automated [Gordon, 1988; Huth and Ryan, 2000].

Floyd-Hoare reasoning, the original flavour of dynamic logic for the analysis of programming, is applicable to sequential transformational programs. Sequential programs run on a single processor without involving concurrency. Transformational programs are programs that are expected to terminate with an output after a finite number of steps. Sequential transformational programs are in the realm of dynamic logic in the sense of the present paper.

Reactive systems are systems that are expected to 'run forever'; examples are text editors, operating systems. Concurrent reactive systems also involve interaction between processes; examples can be found in hardware systems, and embedded systems like the software that controls ignition and fuel injection of cars. The analysis and verification of (concurrent) reactive systems calls for model checking methods using temporal computation tree logics such as CTL, LTL and CTL* [Pnueli, 1981; Clarke and Emerson, 1982; Clarke et $al.$, 1993], and is outside the scope of our survey (but see Section 3.6 below).

3 PROPOSITIONAL DYNAMIC LOGIC

The language of propositional dynamic logic was defined by Pratt in [Pratt, 1976; 1980] as a generic language for reasoning about computation. Axiomatisations were given independently by Segerberg [Segerberg, 1982], Fisher/Ladner [Fischer and Ladner, 1979], and Parikh [Parikh, 1978]. These axiomatisations make the connection between propositional dynamic logic and modal logic very clear.

3.1 Language

Propositional dynamic logic can be viewed as a basic logic of change. Propositional dynamic logic abstracts over the set of basic actions, in the sense that basic actions are atoms. This means that its range of applicability is vast. In the WHILE language, the basic actions are definite assignments $v := a$ and the trivial action SKIP. Now the basic actions can be anything. The only thing that matters about a basic action a is that it is interpreted by some binary relation on a state set.

Dynamic logics have two basic syntactic categories: formulae and programs. Formulae are used for talking about states, programs for classifying transitions between states.

The same distinction can be found in all imperative programming languages, by the way. Imperative programming languages have programs (often called 'statements') versus formulae (often called 'Boolean expressions'). In the case of the WHILE language, the booleans appeared as conditions in conditional statements and as guards in guarded iterations.

Propositional dynamic logic is an extension of propositional logic with programs, just like basic modal logic is an extension of propositional logic with modalities. Let a set of basic propositions P be given. Appropriate states will contain valuations for these propositions. Assume a set of basic actions A. Every basic action corresponds to a binary relation on the state set.

Let p range over the set of basic propositions P, and let a range over a set of basic actions A. Then the formulae φ and programs α of propositional dynamic logic are given by:

$$\begin{aligned} \varphi &::= \ \top \mid p \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \langle\alpha\rangle\varphi \\ \alpha &::= \ a \mid ?\varphi \mid \alpha_1 \ ; \ \alpha_2 \mid \alpha_1 \cup \alpha_2 \mid \alpha^* \end{aligned}$$

We employ the usual abbreviations: $\bot$ is shorthand for $\neg\top$, $\varphi_1 \wedge \varphi_2$ is shorthand for $\neg(\neg\varphi_1 \vee \neg\varphi_2)$, $\varphi_1 \rightarrow \varphi_2$ is shorthand for $\neg\varphi_1 \vee \varphi_2$, $\varphi_1 \leftrightarrow \varphi_2$ is shorthand for $(\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_2 \rightarrow \varphi_1)$, and $[\alpha]\varphi$ is shorthand for $\neg\langle\alpha\rangle\neg\varphi$. Also, we will use α^n for the program consisting of a sequence of n copies of α, i.e. we define α^n by means of $\alpha^0 := ?\top, \alpha^{n+1} := \alpha \; ; \; \alpha^n$.

Taking the basic actions to be computations, we can use PDL to talk about programming: for any program α, $\langle\alpha\rangle\top$ expresses that the program has at least one successful computation, and $[\alpha]\bot$ expresses that the program fails (does not produce any output). If the basic actions are communicative actions, e.g., public announcements, then $\langle\alpha\rangle\varphi$ expresses that a public announcement of α may have the effect that φ holds. If the basic actions are changes in the world, such as spilling milk S or cleaning C, then $[C \; ; \; S]d$ expresses that cleaning up followed by spilling milk always results in a dirty state, while $[S \; ; \; C]\neg d$ expresses that the occurrence of these events in the reverse order always results in a clean state.

Nor does this exhaust the application areas of PDL. In [Blackburn *et al.*, 1993] and [Kracht, 1995], variants of PDL are used for defining a variety of structural relations in syntax trees for natural language, and in [Marx, 2004] PDL is used to analyse *XPath*, a node addressing language of XML documents.

3.2   Semantics

If R_1, R_2 are binary relations on a state set S, then the relational composition $R_1 \circ R_2$ of R_1 and R_2 is given by:

$$R_1 \circ R_2 = \{(t_1, t_2) \in S \times S \mid \exists t_3 \in S \; ((t_1, t_3) \in R_1 \wedge (t_3, t_2) \in R_2)\}.$$

Let I be the identity relation on S. Then the n-fold composition of a binary relation R on S with itself is defined by recursion, as follows:

$$\begin{array}{rcl} R^0 & = & I \\ R^n & = & R \circ R^{n-1} \end{array}$$

The reflexive transitive closure of R is given by:

$$R^* = \bigcup_{n \in \mathbb{N}} R^n.$$

The semantics of PDL over P, A is given relative to a labelled transition system $\mathbf{M} = \langle S, V, R \rangle$ for signature P, A. The formulae of PDL are interpreted as subsets of $S_\mathbf{M}$, the actions a of PDL as binary relations on $S_\mathbf{M}$ (with the interpretation of basic actions a given as $\xrightarrow{a}$), as follows:

$$
\begin{aligned}
\llbracket \top \rrbracket^{\mathbf{M}} &= S_{\mathbf{M}} \\
\llbracket p \rrbracket^{\mathbf{M}} &= \{s \in S_{\mathbf{M}} \mid p \in V_{\mathbf{M}}(s)\} \\
\llbracket \neg\varphi \rrbracket^{\mathbf{M}} &= S_{\mathbf{M}} - \llbracket \varphi \rrbracket^{\mathbf{M}} \\
\llbracket \varphi_1 \vee \varphi_2 \rrbracket^{\mathbf{M}} &= \llbracket \varphi_1 \rrbracket^{\mathbf{M}} \cup \llbracket \varphi_2 \rrbracket^{\mathbf{M}} \\
\llbracket \langle \alpha \rangle \varphi \rrbracket^{\mathbf{M}} &= \{s \in S_{\mathbf{M}} \mid \exists t \; (s,t) \in \llbracket \alpha \rrbracket^{\mathbf{M}} \text{ and } t \in \llbracket \varphi \rrbracket^{\mathbf{M}}\} \\[6pt]
\llbracket a \rrbracket^{\mathbf{M}} &= \overset{a}{\rightarrow}_{\mathbf{M}} \\
\llbracket ?\varphi \rrbracket^{\mathbf{M}} &= \{(s,s) \in S_{\mathbf{M}} \times S_{\mathbf{M}} \mid s \in \llbracket \varphi \rrbracket^{\mathbf{M}}\} \\
\llbracket \alpha_1 \; ; \; \alpha_2 \rrbracket^{\mathbf{M}} &= \llbracket \alpha_1 \rrbracket^{\mathbf{M}} \circ \llbracket \alpha_2 \rrbracket^{\mathbf{M}} \\
\llbracket \alpha_1 \cup \alpha_2 \rrbracket^{\mathbf{M}} &= \llbracket \alpha_1 \rrbracket^{\mathbf{M}} \cup \llbracket \alpha_2 \rrbracket^{\mathbf{M}} \\
\llbracket \alpha^* \rrbracket^{\mathbf{M}} &= (\llbracket \alpha \rrbracket^{\mathbf{M}})^*
\end{aligned}
$$

If $s \in S_{\mathbf{M}}$ then we use $\mathbf{M} \models_s \varphi$ for $s \in \llbracket \varphi \rrbracket^{\mathbf{M}}$.

These definitions specify how formulae of PDL can be used to make assertions about PDL models. The formula $\langle a \rangle \top$, when interpreted at some state in a PDL model, expresses that that state has a successor in the $\overset{a}{\rightarrow}$ relation in that model.

A PDL formula φ is *true* in a model if it holds at every state in that model, i.e. if $\llbracket \varphi \rrbracket^{\mathbf{M}} = S_{\mathbf{M}}$. Truth of the formula $\langle a \rangle \top$ in a model expresses that $\overset{a}{\rightarrow}$ is serial in that model.

A PDL formula φ is *valid* if it holds for all PDL models $\mathbf{M}$ that φ is true in that model, i.e. that $\llbracket \varphi \rrbracket^{\mathbf{M}} = S_{\mathbf{M}}$. An example of a valid formula is $\langle a \; ; \; b \rangle \top \leftrightarrow \langle a \rangle \langle b \rangle \top$.

Note that ? is an operation for mapping formulae to programs. Programs of the form $?\varphi$ are called *tests*; they are interpreted as the identity relation, restricted to the states satisfying the formula.

Programming Constructs The following abbreviations illustrate how PDL expresses the key constructs of imperative programming:

$$
\begin{aligned}
\text{SKIP} &:= \; ?\top \\
\text{ABORT} &:= \; ?\bot \\
\text{IF } \varphi \text{ THEN } \alpha_1 \text{ ELSE } \alpha_2 &:= \; (?\varphi \; ; \; \alpha_1) \cup (?\neg\varphi \; ; \; \alpha_2) \\
\text{WHILE } \varphi \text{ DO } \alpha &:= \; (?\varphi \; ; \; \alpha)^* \; ; \; ?\neg\varphi \\
\text{REPEAT } \alpha \text{ UNTIL } \varphi &:= \; \alpha \; ; \; (?\neg\varphi \; ; \; \alpha)^* \; ; \; ?\varphi.
\end{aligned}
$$

3.3 PDL Equivalences

The two PDL programs $\beta \; ; \; \text{WHILE } \varphi \text{ DO } \beta$ and REPEAT β UNTIL $\neg\varphi$ are equivalent, in the sense that they will receive the same interpretations in all PDL models, for any choice of PDL formula φ and PDL program β. What this means is that for any formula ψ, the formula

$$\langle \beta \ ; \ \text{WHILE } \varphi \text{ DO } \beta\rangle\psi \ \leftrightarrow \ \langle \text{REPEAT } \beta \text{ UNTIL } \neg\varphi\rangle\psi$$

will be true in all PDL models.
   Similarly, the formula

$$\langle \text{ IF } \varphi \text{ THEN } \beta \text{ ELSE } \gamma\rangle\psi \ \leftrightarrow \ \langle \text{ IF } \neg\varphi \text{ THEN } \gamma \text{ ELSE } \beta\rangle\psi$$

will be true in all PDL models, for all choices of $\beta, \gamma, \varphi, \psi$.
   The regular expressions over a finite alphabet Σ are given by (σ ranges over Σ):

$$E \quad ::= \quad \epsilon \mid \sigma \mid E_1 \ ; \ E_2 \mid E_1 \cup E_2 \mid E^*$$

The denotations of regular expressions over Σ are precisely the regular languages over Σ. Two regular expressions are equivalent if they denote the same language. It is clear that if the basic actions are taken as the alphabet Σ, regular expressions correspond to PDL programs (take $?\top$ for the empty string ϵ).
   Regular expression equivalence can be expressed in PDL, as follows. The regular expressions $(A \cup B)^*$ and $(A^* \ ; \ B^*)^*$ are equivalent. This law translates into PDL as the equivalence of the programs $(\alpha \cup \beta)^*$ and $(\alpha^* \ ; \ \beta^*)^*$ (or the equivalence of the formulae $\langle(\alpha \cup \beta)^*\rangle\varphi$ and $\langle(\alpha^* \ ; \ \beta^*)^*\rangle\varphi$). And so on.

3.4   Axiomatisation

The logic of PDL is axiomatised as follows. Axioms are all propositional tautologies, plus the following axioms (we give box ($[\alpha]$)versions here, but every axiom has an equivalent diamond ($\langle\alpha\rangle$) version):

$$
\begin{array}{rll}
(\text{K}) & \vdash & [\alpha](\varphi \rightarrow \psi) \rightarrow ([\alpha]\varphi \rightarrow [\alpha]\psi) \\
(\text{test}) & \vdash & [?\varphi_1]\varphi_2 \leftrightarrow (\varphi_1 \rightarrow \varphi_2) \\
(\text{sequence}) & \vdash & [\alpha_1 \ ; \ \alpha_2]\varphi \leftrightarrow [\alpha_1][\alpha_2]\varphi \\
(\text{choice}) & \vdash & [\alpha_1 \cup \alpha_2]\varphi \leftrightarrow [\alpha_1]\varphi \wedge [\alpha_2]\varphi \\
(\text{mix}) & \vdash & [\alpha^*]\varphi \leftrightarrow \varphi \wedge [\alpha][\alpha^*]\varphi \\
(\text{induction}) & \vdash & (\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi)) \rightarrow [\alpha^*]\varphi
\end{array}
$$

and the following rules of inference:

(modus ponens) From $\vdash \varphi_1$ and $\vdash \varphi_1 \rightarrow \varphi_2$, infer $\vdash \varphi_2$.

(modal generalisation) From $\vdash \varphi$, infer $\vdash [\alpha]\varphi$.

   The first axiom is the familiar K axiom from modal logic. The second captures the effect of testing, the third captures concatenation, the fourth choice. These axioms together reduce PDL formulae without $*$ to formulae of multi-modal logic. The fifth axiom, the so-called mix axiom, expresses the fact that α^* is a reflexive

and transitive relation containing α, and the sixth axiom, the axiom of induction, captures the fact that α^* is the *least* reflexive and transitive relation containing α.

All axioms have dual forms in terms of $\langle \alpha \rangle$, derivable by propositional reasoning. For example, the dual form of the test axiom reads

$$\vdash \langle ?\varphi_1 \rangle \varphi_2 \leftrightarrow (\varphi_1 \wedge \varphi_2).$$

The dual form of the induction axiom reads

$$\vdash \langle \alpha^* \rangle \varphi \rightarrow \varphi \vee \langle \alpha^* \rangle (\neg \varphi \vee \langle \alpha \rangle \varphi).$$

Use $\Gamma \vdash \varphi$ to express that φ is derivable using hypotheses from Γ by means of the axioms and inference rules of PDL. By induction on the length of proofs it can be shown that PDL satisfies the deduction theorem:

$$\Gamma \cup \{\varphi\} \vdash \psi \text{ iff } \Gamma \vdash \varphi \rightarrow \psi.$$

The deduction theorem will be used to facilitate PDL reasoning in what follows.

The following theorem shows that in the presence of the other axioms, the induction axiom is equivalent to the so-called loop invariance rule:

$$\frac{\varphi \rightarrow [\alpha]\varphi}{\varphi \rightarrow [\alpha^*]\varphi}$$

THEOREM 2. *In PDL without the induction axiom, the induction axiom and the loop invariance rule are interderivable.*

Proof. For deriving the loop invariance rule from the induction axiom, assume the induction axiom. Suppose

$$\vdash \varphi \rightarrow [\alpha]\varphi.$$

Then by modal generalisation:

$$\vdash [\alpha^*](\varphi \rightarrow [\alpha]\varphi).$$

Now assume φ. Then:

$$\varphi \vdash \varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi).$$

From this by the induction axiom and propositional reasoning:

$$\varphi \vdash [\alpha^*]\varphi.$$

From this by conditionalisation (the left-to-right direction of the deduction theorem):

$$\vdash \varphi \rightarrow [\alpha^*]\varphi.$$

Now assume the loop invariance rule. We have to establish the induction axiom. Assume φ and $[\alpha^*](\varphi \rightarrow [\alpha]\varphi)$. Then by the mix axiom:

$$\varphi, [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \vdash \varphi \rightarrow [\alpha]\varphi.$$

From this, by propositional reasoning:

$$\varphi, [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \vdash [\alpha]\varphi.$$

Conditionalisation:

$$\vdash (\varphi \land [\alpha^*](\varphi \rightarrow [\alpha]\varphi)) \rightarrow [\alpha]\varphi.$$

Applying the loop invariance rule to this yields the induction axiom:

$$\vdash (\varphi \land [\alpha^*](\varphi \rightarrow [\alpha]\varphi)) \rightarrow [\alpha^*]\varphi.$$

$\blacksquare$

3.5   PDL and Floyd-Hoare Reasoning

Floyd-Hoare correctness assertions are expressible in PDL, as follows. If φ, ψ are PDL formulae and α is a PDL program, then

$$\{\varphi\}\ \alpha\ \{\psi\}$$

translates into

$$\varphi \rightarrow [\alpha]\psi.$$

Clearly, $\{\varphi\}\ \alpha\ \{\psi\}$ holds in a state in a model iff $\varphi \rightarrow [\alpha]\psi$ is true in that state in that model.

The Floyd-Hoare inference rules can now be derived in PDL. As an example we derive the rule for guarded iteration:

$$\frac{\{\varphi \land \psi\}\ \alpha\ \{\psi\}}{\{\psi\}\ \text{WHILE}\ \varphi\ \text{DO}\ \alpha\ \{\neg\varphi \land \psi\}}$$

Let the premise $\{\varphi \land \psi\}\ \alpha\ \{\psi\}$ be given, i.e. assume (4).

(4) $\vdash (\varphi \land \psi) \rightarrow [\alpha]\psi.$

We wish to derive the conclusion

$$\vdash \{\psi\}\ \text{WHILE}\ \varphi\ \text{DO}\ \alpha\ \{\neg\varphi \land \psi\},$$

i.e. we wish to derive (5).

(5) $\vdash \psi \rightarrow [(?\varphi; \alpha)^*\ ;\ ?\neg\varphi](\neg\varphi \land \psi).$

From (4) by means of propositional reasoning:

$$\vdash \psi \rightarrow (\varphi \rightarrow [\alpha]\psi).$$

From this, by means of the test and sequence axioms:

$$\vdash \psi \rightarrow [\varphi\ ;\ \alpha]\psi.$$

Applying the loop invariance rule gives:

$$\vdash \psi \to [(\varphi \; ; \; \alpha)^*]\psi.$$

Since ψ is propositionally equivalent with $\neg\varphi \to (\neg\varphi \wedge \psi)$, we get from this by propositional reasoning:

$$\vdash \psi \to [(\varphi \; ; \; \alpha)^*](\neg\varphi \to (\neg\varphi \wedge \psi)).$$

The test axiom and the sequencing axiom yield the desired result (5).

3.6   Properties

Failure of Compactness

The presence of the * (Kleene star) operator causes true infinitary behaviour. In particular, the compactness theorem, which says that finite satisfiability of an infinite set of formulae Γ implies satisfiability of Γ, fails for PDL. Here is an example of a set of PDL formulae that is finitely satisfiable but not satisfiable:

$$\{\langle a^*\rangle p\rangle\} \cup \{\neg p, \neg\langle a\rangle p, \neg\langle a^2\rangle p, \ldots\}.$$

Finite Model Property

A logic has the finite model property (fmp) if every non-theorem of the logic has a finite counterexample. Having the fmp implies decidability, but not conversely (there are decidable logics without the fmp). We will now show that PDL has the fmp.

For normal modal logic, the fmp can be shown by means of the so-called filtration method [Blackburn *et al.*, 2001, Ch 2], using subformula closed sets of formulae. Because of the presence of the star operator, in the case of PDL closure under subformulae is not enough. We also need to make sure that program modalities are decomposed in an appropriate way. For this, we use so-called Fisher-Ladner closures [Fischer and Ladner, 1979].

Define $\mathrm{FL}(\varphi)$, the Fisher-Ladner closure of a PDL formula φ, as follows. $\mathrm{FL}(\varphi)$ is the smallest set of formulae X containing φ that is closed under the following operations (the definition assumes diamond modalities here; an equivalent formulation in terms of box modalities is also possible):

- if $\neg\psi \in X$ then $\psi \in X$,

- if $(\psi_1 \vee \psi_2) \in X$ then $\psi_1 \in X, \psi_2 \in X$,

- if $\langle\alpha\rangle\psi \in X$ then $\psi \in X$,

- if $\langle\alpha_1 \; ; \; \alpha_2\rangle\psi \in X$ then $\langle\alpha_1\rangle\langle\alpha_2\rangle\psi \in X$,

- if $\langle\alpha_1 \cup \alpha_2\rangle\psi \in X$ then $\langle\alpha_1\rangle\psi \vee \langle\alpha_2\rangle\psi \in X$,

- if $\langle ?\psi_1\rangle\psi_2 \in X$ then $\psi_1 \in X, \psi_2 \in X$,

- if $\langle\alpha^*\rangle\psi \in X$ then $\langle\alpha\rangle\langle\alpha^*\rangle\psi \in X$.

Note that $\mathrm{FL}(\varphi)$ is always finite. E.g., $\mathrm{FL}(\langle(a \; ; \; b)^*\rangle(p \vee q))$ equals

$$\{\langle(a \; ; \; b)^*\rangle(p \vee q), p \vee q, p, q,$$
$$\langle(a \; ; \; b)\rangle\langle(a \; ; \; b)^*\rangle(p \vee q), \langle a\rangle\langle b\rangle\langle(a \; ; \; b)^*\rangle(p \vee q), \langle b\rangle\langle(a \; ; \; b)^*\rangle(p \vee q)\}.$$

Using $\mathrm{FL}(\varphi)$, define filtrations of LTSs, as follows. Let $\mathbf{M} = (S, V, R)$ be an LTS. For every s, let $\bar{s} = \{\psi \in \mathrm{FL}(\varphi) \mid \mathbf{M} \models_s \psi\}$.

Set $\bar{s}\bar{R}_a\bar{t}$ if $\exists u, v \in S$ such that $uR_a v$ and $\bar{u} = \bar{s}$ and $\bar{v} = \bar{t}$. Finally, put $\bar{V}(\bar{s}) = \{p \in P \mid p \in \bar{s}\}$. Let $\overline{\mathbf{M}} = (\bar{S}, \bar{V}, \bar{R})$. Then one can prove:

LEMMA 3 (Filtration Lemma). *For all $\psi \in FL(\varphi)$, all $s \in S$:*

$$\mathbf{M} \models_s \psi \;\; \text{iff} \;\; \overline{\mathbf{M}} \models_{\bar{s}} \psi.$$

Proof. One shows with induction on the complexity of formulae and programs occurring in $\mathrm{FL}(\varphi)$ that:

- $\mathbf{M} \models_s \psi$ iff $\overline{\mathbf{M}} \models_{\bar{s}} \psi$.

- if $sR_\alpha t$ then $\bar{s}\bar{R}_\alpha\bar{t}$.

The crucial step is the following. Suppose that $\langle\alpha\rangle\psi$ is true in $\overline{\mathbf{M}}$ on $\bar{s}$. Then there exists a computation path for α consisting of a finite sequence of atomic transitions

$$\bar{s} \rightarrow \bar{s}_1 \rightarrow \cdots \rightarrow \bar{s}_n = \bar{t},$$

with appropriate atomic $\bar{R}_a$ links between $\bar{s}_i$ and $\bar{s}_{i+1}$, and possible appropriate tests $?\chi_i$ at $\bar{s}_i$, and with ψ true at $\bar{t}$.

By the definition of $\bar{R}_a$, there has to be a corresponding 'pseudo computation path'

$$s \sim u \rightarrow s_1 \sim u_1 \rightarrow \cdots \rightarrow u_n \sim t,$$

where $x \sim y$ expresses that $\bar{x} = \bar{y}$. Moreover, we have by the induction hypothesis that the same test conditions $?\chi_i$ hold at s_i and u_i, and that ψ holds at u_n and t.

Next, prove by induction on α:

If $\langle\alpha\rangle\psi \in \mathrm{FL}(\varphi)$ and there is pseudo computation path for α from s to t with $\mathbf{M} \models_t \psi$ then $\mathbf{M} \models_s \langle\alpha\rangle\psi$.

This clinches the argument.   ∎

Decidability

Decidability follows from the filtration lemma:

THEOREM 4. *Universal validity for PDL is decidable*

Proof. By the filtration lemma, counterexamples for a formula φ must already show up in models with at most $2^{|FL(\varphi)|}$ states. It is possible, in principle, to inspect all of these. ∎

It follows immediately that satisfiability for PDL is decidable too: to check that φ is satisfiable, just find a satisfying model with at most $2^{|FL(\varphi)|}$ states.

Converse

Let ˘ (converse) be an operator on PDL programs with the following interpretation:

$$[\![\alpha^{\smile}]\!]^{\mathbf{M}} \quad = \quad \{(s,t) \mid (t,s) \in [\![\alpha]\!]^{\mathbf{M}}\}.$$

It is easy to see that the following equations hold:

$$\begin{aligned}
(\alpha \; ; \; \beta)^{\smile} &= \beta^{\smile} \; ; \; \alpha^{\smile} \\
(\alpha \cup \beta)^{\smile} &= \alpha^{\smile} \cup \beta^{\smile} \\
(\alpha^*)^{\smile} &= (\alpha^{\smile})^*
\end{aligned}$$

This means that it is enough to add converse to the PDL language for atomic programs only. To see that adding converse in this way increases expressive power, observe that in state 0 in the following picture $\langle a^{\smile}\rangle\top$ is true, while in state 2 in the picture $\langle a^{\smile}\rangle\top$ is false. On the assumption that 0 and 2 have the same valuation, no PDL formula without converse can distinguish the two states.
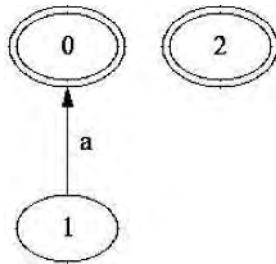
Suitable axioms to enforce that $a^{\smile}$ behaves as the converse of a are well known from temporal logic (read $\langle a\rangle$ as F 'once in the future', $[a]$ as G 'always in the future', $\langle a^{\smile}\rangle$ as P 'once in the past', $[a^{\smile}]$ as H 'always in the past', [Prior, 1957; 1967]):

$$\begin{aligned}
\varphi &\rightarrow [a]\langle a^{\smile}\rangle\varphi \\
\varphi &\rightarrow [a^{\smile}]\langle a\rangle\varphi
\end{aligned}$$

Wellfoundedness, Halting

For deterministic programs α, formula $\langle\alpha\rangle\top$ expresses that α does not loop. For non-deterministic programs α, however, there turns out to be no PDL way to express non-looping behaviour. If α is non-deterministic, $\langle\alpha\rangle\top$ merely says that in the current state there exists a terminating run for α, it does not preclude the existence of diverging runs. For example, formula $\langle(?\top)^*\rangle\top$ will be true at any state, while $(?\top)^*$ has diverging runs from every state.

One way to deal with this situation is to add a predicate to PDL to express wellfoundedness. A relation R is wellfounded in s_0 if there does *not* exist an infinite sequence $s_0, s_1, \ldots$ with

$$s_0 R s_1, s_1 R s_2, \ldots$$

Let **wellfounded** be a predicate for this. Then its interpretation is:

$$[\![\mathbf{wellfounded}(\alpha)]\!]^{\mathcal{M}} \;\;=\;\; \{s_0 \in S_{\mathbf{M}} \mid \neg\exists s_1, s_2, \ldots \forall i \geq 0 (s_i, s_{i+1}) \in [\![\alpha]\!]^{\mathbf{M}}\}.$$

In terms of **wellfounded**, a predicate **halt** for program termination can be defined as follows:

$$
\begin{aligned}
\mathbf{halt}(a) &:\equiv \top \\
\mathbf{halt}(?\varphi) &:\equiv \top \\
\mathbf{halt}(\alpha \;;\; \beta) &:\equiv \mathbf{halt}(\alpha) \wedge [\alpha]\mathbf{halt}(\beta) \\
\mathbf{halt}(\alpha \cup \beta) &:\equiv \mathbf{halt}(\alpha) \wedge \mathbf{halt}(\beta) \\
\mathbf{halt}(\alpha^*) &:\equiv \mathbf{wellfounded}(\alpha) \wedge [\alpha^*]\mathbf{halt}(\alpha)
\end{aligned}
$$

What the definition of **halt** for programs of the form α^* says is that for α^* to halt it has to be the case that α is wellfounded at the present state (so that its execution can not be repeated without end), and also α has to halt at all states that can be reached in a finite number of α steps from the present state. This expresses that α^* can loop for two reasons: (i) because α can be repeated without end, or (ii) because after repeated execution of α there is a state where α itself does not terminate.

Applying this to the example program $(?\top)^*$, we get:

$$
\begin{aligned}
\mathbf{halt}((?\top)^*) &\equiv \mathbf{wellfounded}(?\top) \wedge [(?\top)^*]\mathbf{halt}(?\top) \\
&\equiv \mathbf{wellfounded}(?\top) \wedge [(?\top)^*]\top \\
&\equiv \mathbf{wellfounded}(?\top) \wedge \top \\
&\equiv \bot
\end{aligned}
$$

What this says is that $(?\top)^*$ does not halt because the test $?\top$ is not wellfounded (for $?\top$ can be repeated an arbitrary number of times).

Floyd-Hoare total correctness statements for PDL programs α,

$$\{\varphi\}\ \alpha\ \{\Downarrow \psi\}$$

can now be expressed as:

$$\varphi \rightarrow [\alpha]\psi \wedge \varphi \rightarrow \mathbf{halt}(\alpha).$$

Every state in the infinite model of the following picture satisfies $\mathbf{halt}(a)$, but clearly, any filtration of this model must collapse some of the states, and in these collapsed states $\mathbf{halt}(a)$ will fail. This shows that extending PDL with a $\mathbf{halt}$ predicate (and, a fortiori, extending PDL with a $\mathbf{wellfounded}$ predicate) increases expressive power.
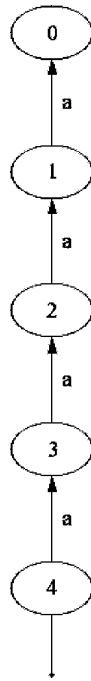
Further Extensions and Variations

Other possible extensions of PDL are with intersection and nominals [Passy and Tinchev, 1991]. The extension with nominals turns PDL into a kind of hybrid logic [Areces *et al.*, 2001]. Replacing the regular programs of PDL by finite automata yields a formalism with the same expressive power but allowing more succinct descriptions: see [Harel *et al.*, 2000]. Replacing the regular programs of PDL with another data structure such as pushdown automata or context free grammars or flowcharts yields more expressive (but also more complex) formalisms.

Complexity

Although satisfiability checking in individual LTSs can be done quite efficiently (i.e. in polynomial time), the above algorithm for checking satisfiability is highly inefficient, because the size of the models to check is exponential in the size of the formula, and the number of these models is doubly exponential in the size of the formula. So the naive satisfiability checking algorithm is doubly exponential in the size of the formula.

Time complexity of the satisfiability problem for PDL is singly exponential: an exponential algorithm is given in [Pratt, 1978]. One cannot do better than this: [Fischer and Ladner, 1979]establishes an exponential-time lower bound for PDL satisfiability, by showing how PDL formulae can encode computations of linear-space-bounded alternating Turing machines. An exponential time satisfiability algorithm for PDL with converse is given in [Streett, 1982]. Intuitively, adding converse does not increase complexity, for converses of atomic programs a can be taken as atoms, and the definition of converse for complex programs is linear in the size of the programs.

Modal μ calculus

For a proper perspective on PDL, it is useful to contrast it with a much more expressive dynamic logic, the modal μ calculus.

Let a set of proposition letters $P = \{p_0, p_1, \ldots\}$, a set of actions $A = \{a_0, a_1, \ldots\}$, and a set of variables $V = \{X_0, X_1, \ldots\}$ be given. Assume p ranges over P, a ranges over A, and X ranges over V. Then the set of μ formulae is given by the following definition:

$$\varphi \quad ::= \quad \top \mid p \mid X \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \langle a \rangle \varphi \mid \mu X.\varphi,$$

with the syntactic restriction on $\mu X.\varphi$ that occurrences of X in φ are positive. An occurrence of X in a formula φ is positive if the occurrence is in the scope of an even number of negation signs.

Interpretation is in LTSs $\mathbf{M}$, relative to an assignment $g : V \to \mathcal{P}(S_\mathbf{M})$. If T is a subset of $S_\mathbf{M}$, $g[X \mapsto T]$ is the assignment that is like g except for the fact that it maps X to T.

$$
\begin{aligned}
[\![\top]\!]_g^\mathbf{M} &= S_\mathbf{M} \\
[\![p]\!]_g^\mathbf{M} &:= \{s \in S_\mathbf{M} \mid p \in V_\mathbf{M}(s)\} \\
[\![X]\!]_g^\mathbf{M} &:= g(X) \\
[\![\neg\varphi]\!]_g^\mathbf{M} &= S_\mathbf{M} - [\![\varphi]\!]_g^\mathbf{M} \\
[\![\varphi_1 \vee \varphi_2]\!]_g^\mathbf{M} &= [\![\varphi_1]\!]_g^\mathbf{M} \cup [\![\varphi_2]\!]_g^\mathbf{M} \\
[\![\langle a \rangle \varphi]\!]_g^\mathbf{M} &= \{s \in S_\mathbf{M} \mid \exists t \; s \xrightarrow{a} t \text{ and } t \in [\![\varphi]\!]_g^\mathbf{M}\} \\
[\![\mu X.\varphi]\!]_g &= \bigcap \{T \subseteq S_\mathbf{M} \mid [\![\varphi]\!]_{g[X \mapsto T]}^\mathbf{M} \subseteq T\}
\end{aligned}
$$

The clause for $\mu X.\varphi$ expresses that the interpretation of this formula is the least fixed point of the operation $T \mapsto \llbracket \varphi \rrbracket^{\mathbf{M}}_{g[T \mapsto S]}$. Thanks to the fact that X only occurs positively in φ, this operation is monotone:

$$\text{if } T \subseteq S \text{ then } \llbracket \varphi \rrbracket^{\mathbf{M}}_{g[X \mapsto T]} \subseteq \llbracket \varphi \rrbracket^{\mathbf{M}}_{g[X \mapsto S]}.$$

It follows, by a theorem of Knaster and Tarski (see, e.g., [Davey and Priestley, 2002]), that the operation has a least fixed point, and that this least fixed point is given by the semantic clause for $\mu X.\varphi$. The proof of this fact is instructive.

For simplicity we use $[\varphi]T$ for $\llbracket \varphi \rrbracket^{\mathbf{M}}_{g[X \mapsto T]}$, and $[\varphi]$ for $T \mapsto [\varphi]T$. Let

$$
\begin{aligned}
W &:= \bigcap \{T \subseteq S_{\mathbf{M}} \mid [\varphi]T \subseteq T\} \\
\mathcal{F} &:= \{T \subseteq S_{\mathbf{M}} \mid [\varphi]T \subseteq T\}.
\end{aligned}
$$

We have to show that W is the least fixed point of $[\varphi]$.

First we show $[\varphi]W \subseteq W$. Observe that for all $U \in \mathcal{F}$ we have $W \subseteq U$ and $[\varphi]U \subseteq U$. By monotonicity of $[\varphi]$, $[\varphi]W \subseteq [\varphi]U$, and therefore, by $[\varphi]U \subseteq U$, $[\varphi]W \subseteq U$. From the fact that for all $U \in \mathcal{F}$ it holds that $[\varphi]W \subseteq U$ we get the desired result $[\varphi]W \subseteq W$.

Next we show $W \subseteq [\varphi]W$. We start out from the previous result $[\varphi]W \subseteq W$. By monotonicity of $[\varphi]$ we get from this that $[\varphi][\varphi]W \subseteq [\varphi]W$. This shows that $[\varphi]W \in \mathcal{F}$, whence $W \subseteq [\varphi]W$.

Finally, to show that W is the least fixpoint, observe that any fixpoint U of $[\varphi]$ is in $\mathcal{F}$, so that $W \subseteq U$.

The modal μ calculus translates into second order predicate logic as follows:

$$
\begin{aligned}
X^{\circ} &:= X(x) \\
(\mu X.\varphi)^{\circ} &:= \forall X(\forall x(\varphi^{\circ} \to X(x)) \to X(x)).
\end{aligned}
$$

This translation is called the *standard translation* into monadic second order logic, *monadic* because the predicate variables X quantified over in the translation are unary.

The μ calculus can be presented in PDL format by distinguishing between formulae and programs, as follows:

$$
\begin{aligned}
\varphi &::= \top \mid p \mid X \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \langle \alpha \rangle \varphi \mid \mu X.\varphi \\
\alpha &::= a \mid ?\varphi \mid \alpha_1 \cup \alpha_2 \mid \alpha_1; \alpha_2 \mid \alpha^*
\end{aligned}
$$

again with the syntactic restriction on $\mu X.\varphi$ formulae that X occurs only positively in φ.

This PDL version of the μ calculus does not have greater expressive power than the original, for we have the following equivalences:

$$
\begin{aligned}
\langle ?\varphi_1 \rangle \varphi_2 &\equiv \varphi_1 \wedge \varphi_2 \\
\langle \alpha_1 \cup \alpha_2 \rangle \varphi &\equiv \langle \alpha_1 \rangle \varphi \vee \langle \alpha_2 \rangle \varphi \\
\langle \alpha_1; \alpha_2 \rangle \varphi &\equiv \langle \alpha_1 \rangle \langle \alpha_2 \rangle \varphi \\
\langle \alpha^* \rangle \varphi &\equiv \mu X.(\varphi \vee \langle \alpha \rangle X).
\end{aligned}
$$

To see that $\langle \alpha^* \rangle \varphi$ and $\mu X.(\varphi \vee \langle \alpha \rangle X)$ are equivalent, observe that the least fixpoint of the operation

$$T \mapsto \llbracket \varphi \rrbracket^{\mathbf{M}} \cup \{s \in S_{\mathbf{M}} \mid \exists t \in T.s \xrightarrow{\alpha} t\}$$

is equal to the set

$$\{s \in S_{\mathbf{M}} \mid \exists t \in \llbracket \varphi \rrbracket^{\mathbf{M}}.s \xrightarrow{\alpha^*} t\}.$$

We will now show that the μ calculus has greater expressive power than PDL. In PDL, there is no way to express that a program is wellfounded. The following formula expresses wellfoundedness of α in the μ calculus:

$$\mu X.[\alpha]X.$$

The meaning of this may not be immediately obvious, so let us analyse this a bit further. Let

$$W := \{s \in S_{\mathbf{M}} \mid \text{ there is no infinite } \alpha \text{ path from } s\}.$$

Then clearly, $\{s \in S_{\mathbf{M}} \mid \text{ if } s \xrightarrow{\alpha} t \text{ then } t \in W\} = W$. If there is no infinite α path starting form s, then there is no infinite α path from any α successor of α, and if at no α successor of s an infinite α path starts, then no infinite α path starts from s. In other words, W is a fixpoint of the operation

$$T \mapsto \{s \in S_{\mathbf{M}} \mid \text{ if } s \xrightarrow{\alpha} t \text{ then } t \in T\}.$$

We still have to show that W is also the least fixpoint of the operation. So suppose U is another solution:

$$\{s \in S_{\mathbf{M}} \mid \text{ if } s \xrightarrow{\alpha} t \text{ then } t \in U\} = U. \qquad (*)$$

We have to show that $W \subseteq U$. Assume, for a contradiction, that there is some $s \in W$ with $s \notin U$. From (*),

$$s \notin \{s \in S_{\mathbf{M}} \mid \text{ if } s \xrightarrow{\alpha} t \text{ then } t \in U\}.$$

It follows that for some $t \in S_{\mathbf{M}}$ we have $s \xrightarrow{\alpha} t$ and $t \notin U$. Continuing like this, we find $t \xrightarrow{\alpha} t'$ with $t' \notin U$, $t' \xrightarrow{\alpha} t''$ with $t'' \notin U$, and so on, an infinite α path starting from s, which contradicts the assumption that $s \in W$.

To define a greatest fixpoint operator dual to μ, use

$$\nu X.\varphi := \neg \mu X.(\varphi[X \mapsto \neg X]),$$

where $\varphi[X \mapsto \neg X]$ denotes the result of replacing every occurrence of X in φ by $\neg X$.

The μ calculus originates in [Kozen, 1983]. It has great expressive power (it subsumes PDL, CTL, LTL and CTL*), it is decidable and has the finite model property [Streett and E.A, 1989], but it has greater complexity than PDL: known decision procedures use doubly exponential time.

Kozen [Kozen, 1983] proposed an elegant proof system: the axioms and rules of multi-modal logic together with the axiom

$$\mu X.\varphi \leftrightarrow \varphi[X \mapsto \mu X.\varphi]$$

and the following rule of inference:

$$\frac{\varphi[X \mapsto \psi] \rightarrow \psi}{\mu X.\varphi \rightarrow \psi}$$

This axiomatisation is sound and complete.

Alternatively, PDL style μ calculus is axiomatised by the axioms and rules of PDL plus the μ axiom and the μ rule of inference.

Bisimulation

PDL and modal μ calculus are both interpreted in LTSs. But the correspondence between LTSs and processes is not one-to-one. The process that produces an infinite number of a transitions and nothing else can be represented as an LTS in lots of different ways. The following representations are all equivalent:
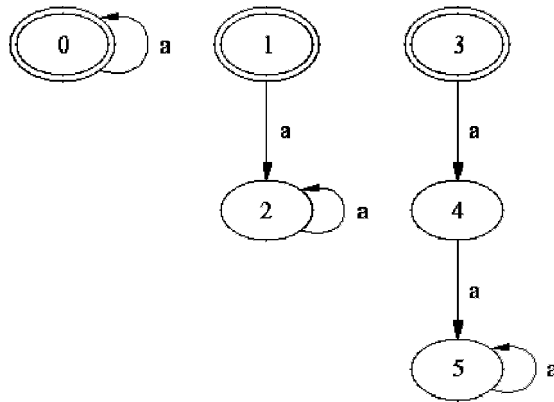
The notion of bisimulation is intended to capture such process equivalences. A bisimulation C between LTSs **M** and **N** is a relation on $S_{\mathbf{M}} \times S_{\mathbf{N}}$ such that if sCt then the following hold:

Invariance $V_{\mathbf{M}}(s) = V_{\mathbf{N}}(t)$ (the two states have the same valuation),

Zig if for some $a \in S_1$ $s \xrightarrow{a} s' \in R_{\mathbf{M}}$ then there is a $t' \in S_2$ with $t \xrightarrow{a} t' \in R_{\mathbf{N}}$ and $s'Ct'$.

Zag same requirement in the other direction.

One uses $\mathbf{M}, s \underleftrightarrow{} \mathbf{N}, t$ to indicate that there is a bisimulation that connects s and t. In such a case one says that s and t are bisimilar.
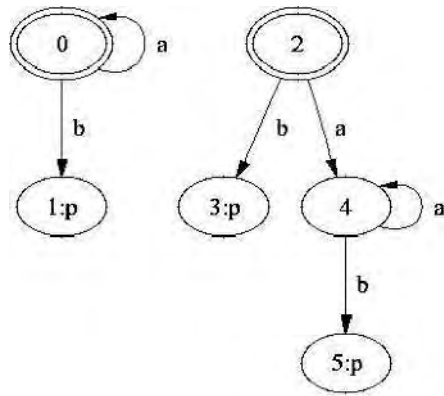
In the LTSs of the picture, $0 \underleftrightarrow{} 2 \underleftrightarrow{} 4$ and $1 \underleftrightarrow{} 3 \underleftrightarrow{} 5$.

Bisimulation is intimately connected to modal logic, as follows. Modal logic is a sublogic of PDL. It is given by restricting the set of programs to atomic programs. Usually, one writes $\diamondsuit_a$ for $\langle a \rangle$:

$$\varphi \quad ::= \quad \top \mid p \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \langle a \rangle \varphi$$

Bisimulations can be viewed as a motivation for modal logic. A global property of LTSs is a function $\mathbf{P}$ that assigns to any LTS $\mathbf{M}$ over a given signature a property $\mathbf{P_M} \subseteq S_\mathbf{M}$. A global property $\mathbf{P}$ is *invariant for bisimulation* if whenever C is a bisimulation between $\mathbf{M}$ and $\mathbf{N}$ with sCt, then $s \in \mathbf{P_M}$ iff $t \in \mathbf{P_N}$.

Modal formulae may be viewed as global properties, for if φ is a modal formula, then $\lambda \mathbf{M}.[\![\varphi]\!]^\mathbf{M}$ is a global property. Similarly for formulae of first order logic.

An example of a first order logic formula that is *not* invariant for bisimulation is the formula $R_a(x, x)$. This formula is true in state 0, but false in bisimilar state 1 in the following picture:
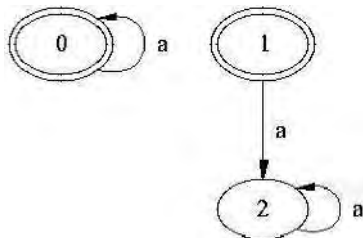
Another example of a first order logic formula that is *not* invariant for bisimulation:

$$\varphi(x) = \exists y(R_a(x, y) \wedge R_b(x, y)).$$

The picture below indicates that $\varphi(x)$ is not invariant for the example bisimulation that links 0 to 2 and 1 to 3 and 4. The state 0 satisfies $\varphi(x)$ while 2 does not, and the two states are bisimilar.
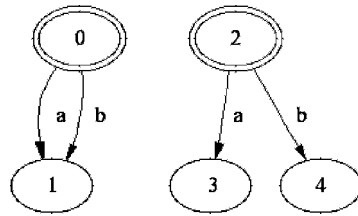
Clearly, all modal formulae are invariant for bisimulation: If φ is a modal formula that is true of a state s, and s is bisimilar to t, then an easy induction on the structure of φ establishes that φ is true of t as well.

More surprisingly, it turns out that all first order formulae that are invariant for bisimulation are translations of modal formulae. If first order logic is given and bisimulation is given, modal logic results from the following theorem:

THEOREM 5 (Van Benthem, [Van Benthem, 1976]). *A first order formula $\varphi(x)$ is* invariant *for bisimulation iff $\varphi(x)$ is equivalent to a modal formula.*

One direction of this can easily be verified by the reader: if φ is a modal formula, it can be proved by induction on formula structure that φ cannot distinguish between bisimilar points.

The argument for the other direction is more involved. We give a sketch of the proof. Define Ψ as the set of modal formulae that are implied by $\varphi(x)$, as follows:

$$\Psi := \{\psi \mid \psi \text{ is a modal formula and } \varphi(x) \models \psi\}.$$

Next, if we can prove that $\Psi \models \varphi(x)$, then the compactness theorem for FOL gives us $\{\psi_1, \ldots, \psi_n\} \subseteq \Psi$ with $\psi_1, \ldots, \psi_n \models \varphi(x)$, and we see that $\varphi(x)$ is equivalent to the modal formula $\psi_1 \wedge \cdots \wedge \psi_n$.

So suppose $\mathbf{M} \models_s \Psi$. We are done if we can show that $\mathbf{M} \models_s \varphi(x)$. For this, consider the modal theory of s, i.e. the set of modal formulae true at s:

$$\Phi := \{\varphi \mid \varphi \text{ is a modal formula and } \mathbf{M} \models_s \varphi\}.$$

Now $\Phi \cup \{\varphi(x)\}$ must be finitely satisfiable (i.e. any finite subset must be satisfiable), for if not then there are $\varphi_1, \ldots, \varphi_n \in \Phi$ with $\varphi(x) \models \neg\varphi_1 \vee \cdots \vee \neg\varphi_n$, which contradicts the fact that $\neg\varphi_1 \vee \cdots \vee \neg\varphi_n$ is false at s. Using the compactness theorem for FOL again, we see that there must be some node t in an LTS $\mathbf{N}$ with $\mathbf{N} \models_t \Phi \cup \{\varphi(x)\}$.

There is one given that we haven't used yet: $\varphi(x)$ is invariant for bisimulation. To use that given, we replace $\mathbf{M}$ and $\mathbf{N}$ by so-called ω saturated elementary extensions $\mathbf{M}^\bullet$ and $\mathbf{N}^\bullet$.

A FOL model $\mathbf{M}$ is ω saturated if whenever $\Phi(x, y_1, \ldots, y_n)$ is a set of first order formulae, and $d_1, \ldots, d_n$ are elements of the domain of $\mathbf{M}$, then $\Phi[x, d_1, \ldots, d_n]$ is finitely satisfiable, i.e. for every finite subset Φ_0 of Φ we can find a d in the domain of $\mathbf{M}$ with $\mathbf{M} \models \Phi[d, d_1, \ldots, d_n]$.

Every FO model has a an ω saturated elementary extension (see Chang and Keisler [Chang and Keisler, 1973, Ch 6] for a proof), so the replacement of $\mathbf{M}, \mathbf{N}$ by $\mathbf{M}^\bullet, \mathbf{N}^\bullet$ is warranted. Moreover, $\mathbf{N}^\bullet \models \varphi(x)$, for truth of $\varphi(x)$ is preserved under the extension.

Lemma: If $\mathbf{M}, \mathbf{N}$ are ω saturated, then the relation of modal equivalence is a bisimulation between them.

Proof of the lemma: Let $\mathbf{M}, \mathbf{N}$ be ω saturated. Let $\equiv$ be the relation of being modally equivalent. Let $\mathbf{M}, s \equiv \mathbf{N}, t$. We show that $s \leftrightarrow t$, by checking the clauses for bisimulation:

Invariance Clearly, s and t have the same valuation.

Zig Suppose $s \xrightarrow{a} s'$. Let Φ be the set of modal formulae that are true at s'. Then for every finite subset Φ_0 of Φ, $\mathbf{M} \models_s \langle a \rangle \bigwedge \Phi_0$. Since $s \equiv t$, $\mathbf{M} \models_t \langle a \rangle \bigwedge \Phi_0$, so there is a t' with $t \xrightarrow{a} t'$ and $\mathbf{M} \models_{t'} \Phi_0$. Thus, Φ is finitely satisfiable in a successors of t. By the fact that $\mathbf{N}$ is ω saturated, it follows that there is a t' with $t \xrightarrow{a} t'$ and $\mathbf{N} \models_{t'} \Phi$.

Zag Same argument in the other direction.

Back to the main proof. $\mathbf{N}^\bullet \models_t \Phi \wedge \varphi(x)$ and $\mathbf{M}^\bullet \models_s \Phi$, where Φ is the modal theory of s. Thus, s, t have the same modal theory, and invoking the lemma we see that $s \leftrightarrow t$. Since $\varphi(x)$ is invariant for bisimulation, $\mathbf{M}^\bullet \models_s \varphi(x)$, hence $\mathbf{M} \models_s \varphi(x)$. ∎

Bisimulations are also intimately connected to PDL, as follows.

A global relation is a function $\mathbf{R}$ that assigns to any LTS $\mathbf{M}$ over a given signature a relation $\mathbf{R}_{\mathbf{M}} \subseteq S_{\mathbf{M}} \times S_{\mathbf{M}}$. A global relation $\mathbf{R}$ is *safe for bisimulation* if whenever C is a bisimulation between $\mathbf{M}$ and $\mathbf{N}$ with sCt, then:

Zig: if $s\mathbf{R}_{\mathbf{M}}s'$ then there is a t' with $t\mathbf{R}_{\mathbf{N}}t'$ and $s'Ct'$,

Zag: vice versa: if $t\mathbf{R}_{\mathbf{N}}t'$ then there is an s' with $s\mathbf{R}_{\mathbf{M}}s'$ and $s'Ct'$.

An example of a relation that is not safe for bisimulation is the relation given by the following first order formula:

$$\varphi(x, y) = R_a(x, y) \wedge x = y.$$

Look at the counterexample picture for invariance of $R_a(x, x)$ again. Formula $\varphi(x, y)$ is true of state pair $(0, 0)$ and false of the state pair $(1, 2)$ in that picture,

but 0 and 1 are bisimilar, and $(0, 0)$ satisfies the zig, and $(1, 2)$ the zag condition for bisimulation.

Another counterexample for safety for bisimulation is provided by the following formula:

$$\psi(x, y) = R_a(x, y) \wedge R_b(x, y).$$

Look at the counterexample picture for invariance of $\exists y(R_a(x, y) \wedge R_b(x, y))$ again. Formula $\psi(x, y)$ is true of state pair $(0, 1)$ and false of state pairs $(2, 3)$ and $(2, 4)$, while 0 and 2 are bisimilar, $(0, 1)$ satisfies the zig condition, and both $(2, 3)$ and $(2, 4)$ satisfy the zag condition for bisimulation.

In fact, invariance for bisimulation and safety for bisimulation are closely connected. If $\varphi(x)$ is invariant for bisimulation then $\varphi(x) \wedge x = y$ is safe for bisimulation. Conversely, if $\varphi(x, y)$ is safe for bisimulation, and P is some unary predicate that does not occur in φ then $\exists y(\varphi(x, y) \wedge P(y))$ is invariant for bisimulation.

Note that the notion of safety for bisimulation generalises the zig and zag conditions of bisimulations, while invariance for bisimulation generalises the invariance condition of bisimulations.

A modal program is a PDL program that does not contain *. Modal programs can be viewed as global relations, for if α is a modal program, then $\lambda \mathbf{M}.[\![\alpha]\!]^{\mathbf{M}}$ is a global relation.

It is not difficult to see that all modal programs are safe for bisimulation. The surprising thing is the converse: all first order relations that are safe for bisimulation turn out to be translations of modal programs.

THEOREM 6 (Van Benthem [van Benthem, 1994]). *A first order formula $\varphi(x, y)$ is* safe *for bisimulation iff $\varphi(x, y)$ is equivalent to a modal program.*

Proofs of this can be found in [van Benthem, 1994; Hollenberg, 1998]. The perspective on Van Benthem's characterisations of modal logic and PDL is from [Hollenberg, 1998]. In fact, Van Benthem gives a slightly different characterisation. He proves that any bisimulation safe first order formula can be generated from atomic tests $?p$, atomic actions a, sequential composition $;$, choice $\cup$ and *dynamic negation* $\sim$, where $\sim\alpha$ is interpreted by:

$$[\![\sim\alpha]\!]^{\mathbf{M}} \quad = \quad \{(s, s) \in S_{\mathbf{M}} \times S_{\mathbf{M}} \mid \neg\exists t(s, t) \in [\![\alpha]\!]^{\mathbf{M}}\}$$

The two characterisations are equivalent, for $\sim\alpha$ is definable as the PDL program $?([\alpha]\bot)$, while any modal PDL test $?\varphi$ can be expressed in terms of dynamic negation using the following translation:

$$
\begin{aligned}
(?\top)^\circ &= \sim\bot \\
(?(\varphi_1 \vee \varphi_2))^\circ &= (?\varphi_1)^\circ \vee (?\varphi_2)^\circ \\
(?\neg\varphi)^\circ &= \sim(?\varphi)^\circ \\
(?\langle\alpha\rangle\varphi)^\circ &= \sim\sim(\alpha \; ; \; (?\varphi)^\circ)
\end{aligned}
$$

Looking at PDL programs from an algebraic perspective, the obvious notion to be axiomatised is that of PDL program equivalence. A calculus that produces

precisely the equations of the form $\alpha_1 = \alpha_2$ for those α_1, α_2 that have the same interpretation in any PDL model is given in [Hollenberg, 1996] (see also [Hollenberg, 1997], where equivalence of modal PDL programs is axiomatised). The axiomatisation has the following quasi-equations between programs:

associativity of ;	$\alpha \;;\; (\beta \;;\; \gamma) = (\alpha \;;\; \beta) \;;\; \gamma$
associativity for $\cup$	$\alpha \cup (\beta \cup \gamma) = (\alpha \cup \beta) \cup \gamma$
commutativity of $\cup$	$\alpha \cup \beta = \beta \cup \alpha$
idempotency of $\cup$	$\alpha \cup \alpha = \alpha$
left distributivity	$(\alpha \cup \beta) \;;\; \gamma = ((\alpha \;;\; \gamma) \cup (\beta \;;\; \gamma))$
right distributivity	$\alpha \;;\; (\beta \cup \gamma) = ((\alpha \;;\; \beta) \cup (\alpha \;;\; \gamma))$
left identity	$?\top \;;\; \alpha = \alpha$
right identity	$\alpha \;;\; ?\top = \alpha$
left zero	$?\bot \;;\; \alpha = ?\bot$
right zero	$\alpha \;;\; ?\bot = ?\bot$
zero sum	$\alpha \cup ?\bot = \alpha$
$*$ expansion	$\alpha^* = ?\top \cup (\alpha \;;\; \alpha^*)$
left induction	$\alpha \;;\; \beta \leq \beta \Rightarrow \alpha^* \;;\; \beta \leq \beta$
right induction	$\beta \;;\; \alpha \leq \beta \Rightarrow \beta \;;\; \alpha^* \leq \beta$
test choice	$?(\varphi \vee \psi) = ?\varphi \cup ?\psi$
test sequence	$?(\varphi \wedge \psi) = ?\varphi \;;\; ?\psi$
domain test	$?\langle\alpha\rangle\top \;;\; \alpha = \alpha$

where $\alpha \leq \beta$ is defined as $\alpha \cup \beta = \beta$, and the following equations between booleans hold:

equations of boolean algebra	
choice	$\langle\alpha \cup \beta\rangle\varphi = \langle\alpha\rangle\varphi \vee \langle\beta\rangle\varphi$
sequence	$\langle\alpha \;;\; \beta\rangle\varphi = \langle\alpha\rangle\langle\beta\rangle\varphi$
iteration	$\langle\alpha^*\rangle\varphi = \varphi \vee \langle\alpha\rangle\langle\alpha^*\rangle\varphi$
induction	$\langle\alpha^*\rangle\varphi = \varphi \vee \langle\alpha^*\rangle(\neg\varphi \wedge \langle\alpha\rangle\varphi)$
test diamond	$\langle?\varphi\rangle\psi = \varphi \wedge \psi$

If one restricts attention to the modal part of PDL (PDL without *, for this is equivalent to multi-modal logic), the quasi-equations for * drop out, and an equational axiomatisation of modal PDL results.

We end with mentioning an intimate connection between modal μ calculus and bisimulation:

THEOREM 7 (Janin and Walukiewicz [Janin and Walukiewicz, 1996]). *A monadic second order formula $\varphi(x)$ is invariant for bisimulation iff it is equivalent to the standard translation in monadic second order logic of a μ sentence.*

4   ANALYSING THE DYNAMICS OF COMMUNICATION

Dynamic logic is the logic of action and the results of action, but it is also a branch of modal logic, and it enjoys the same breadth of applications as modal logic.

What happens if we reinterpret the atomic action modalities as something else? In epistemic logic, atomic accessibilities denote epistemic similarity relations of agents in a multi-agent epistemic setting. Epistemic PDL is the result of reinterpreting the basic action modalities as epistemic relations. Now $[a; b]\varphi$ means that agent a knows that agent b knows that φ. This is more expressive than multi-agent epistemic logic. E.g., $[(a \cup b)^*]\varphi$ expresses that φ is common knowledge among a and b, and it is well known that common knowledge for a, b cannot be expressed in terms of basic modalities $[a], [b]$ alone.

As an aside, expressing *implicit knowledge* would require extending epistemic PDL with an intersection operation. Implicit knowledge among a, b that φ can be expressed in this extended language as $[a \cap b]\varphi$. This extension results in a logic that is still decidable, but the invariance for bisimulation gets lost. Implicit knowledge will not concern us in what follows.

Interestingly, the shift of application from computation to epistemics turns PDL into a description tool for static situations, for under this interpretation LTSs denote multi-agent epistemic situations instead of sets of computations within a set of states. Still, at a higher level, there is again a dynamic turn. We can study how multi-agent epistemic situations evolve as a result of communicative actions. An important example of such actions is public announcement. What happens to the knowledge of a set of participating agents if it is suddenly announced to all that φ is the case? On the assumption that none of the agents takes φ to be impossible, this should result in a new epistemic state of affairs where it is common knowledge among the agents that φ. In this section we will see that epistemic PDL (PDL, with the basic modalities interpreted as epistemic relations) is eminently suited for the analysis of the dynamics of communication.

Dynamic epistemic logic (cf., e.g., [Baltag, 2002; Baltag and Moss, 2004; Baltag *et al.*, 1999; 2003]) analyses the changes in epistemic information among sets of agents that result from various communicative actions, such as public announcements, group messages and individual messages. The logics studied in [Baltag *et al.*, 2003] add information update operations to epistemic description languages with a common knowledge operator, in such a way that the addition increases expressive power. This makes axiomatisations complicated and completeness proofs hard. In [Kooi and van Benthem, 2004] it is demonstrated how update axioms can be made susceptible to reduction axioms, by the simple means of switching to more expressive epistemic description languages. In particular, it is shown in [Kooi and van Benthem, 2004] how generic updates with epistemic actions can be axiomatised in automata PDL [Harel *et al.*, 2000, Chapter 10.3].

We will follow [van Eijck, 2004] in giving a direct reduction of the logic of generic updates with epistemic actions in the style of [Baltag *et al.*, 1999; 2003] to PDL.

4.1   System

Let $\mathcal{L}$ be a language that can be interpreted in labelled transition systems. Then action models for $\mathcal{L}$ look like this:

DEFINITION 8 (Action models for $\mathcal{L}$, Ag). Let a set of agents Ag and an LTS language $\mathcal{L}$ with label set Ag be given. An action model for $\mathcal{L}, Ag$ is a triple

$$A = ([s_0, \ldots, s_{n-1}], \text{pre}, T)$$

where $[s_0, \ldots, s_{n-1}]$ is a finite list of action states, $\text{pre} : \{s_0, \ldots, s_{n-1}\} \to \mathcal{L}$ assigns a precondition to each action state, and $T : Ag \to \mathcal{P}(\{s_0, \ldots, s_{n-1}\}^2)$ assigns an accessibility relation $\xrightarrow{a}$ to each agent $a \in Ag$.

$\mathcal{L}$ actions can be executed in labelled transition systems for $\mathcal{L}$, by means of the following product construction:

DEFINITION 9 (Action Update). Let an LTS $\mathbf{M} = (W, V, R)$, a world $w \in W$, and a pointed action model (A, s), with $A = ([s_0, \ldots, s_{n-1}], \text{pre}, T)$, be given. Then the result of executing (A, s) in $(\mathbf{M}, w)$ is the model $(\mathbf{M} \otimes A, (w, s))$, with $\mathbf{M} \otimes A = (W', V', R')$, where

$$
\begin{aligned}
W' &= \{(w, s) \mid s \in \{s_0, \ldots, s_{n-1}\}, w \in [\![\text{pre}(s)]\!]^{\mathbf{M}}\} \\
V'(w, s) &= V(w) \\
R'(a) &= \{((w, s), (w', s')) \mid (w, w') \in R(a), (s, s') \in T(a)\}.
\end{aligned}
$$

For the set of basic propositions P and the set of agents Ag, the language of PDL^{DEL} (which we will call 'update PDL') over P, Ag is like that for standard PDL over P, Ag, but with a construct for action update added: if φ is an update PDL formula, and $[A, s]$ is a single pointed action model, then $[A, s]\varphi$ is an update PDL formula. If B is a set of agents $\{b_1, \ldots, b_n\}$, then we abbreviate $b_1 \cup \cdots \cup b_n$ as B. Now $[B]\varphi$ expresses that φ is *general* knowledge among B (they all know φ, but they need not know that the others know φ) and $[B^*]\varphi$ expresses that φ is *common* knowledge among B (they all know φ and they all know that the others know φ).

The semantics of PDL^{DEL} is given by the standard PDL clauses, with the following clause for update added:

$$[\![[A, s]\varphi]\!]^{\mathbf{M}} = \{w \in W_{\mathbf{M}} \mid \text{if } \mathbf{M} \models_w \text{pre}(s) \text{ then } (w, s) \in [\![\varphi]\!]^{\mathbf{M} \otimes A}\}.$$

Using $\langle A, s \rangle \varphi$ as shorthand for $\neg[A, s]\neg\varphi$, we see that the interpretation for $\langle A, s \rangle \varphi$ turns out as:

$$[\![\langle A, s \rangle \varphi]\!]^{\mathbf{M}} = \{w \in W_{\mathbf{M}} \mid \mathbf{M} \models_w \text{pre}(s) \text{ and } (w, s) \in [\![\varphi]\!]^{\mathbf{M} \otimes A}\}.$$

Updating with multiple pointed update actions is also possible. A multiple pointed action is a pair (A, S), with A an action model, and S a subset of the state set of A. Extend the language with updates $[A, S]\varphi$, and interpret this as follows:

$$[\![[A, S]\varphi]\!]^{\mathbf{M}} = \{w \in W_{\mathbf{M}} \mid \forall s \in S(\text{ if } \mathbf{M} \models_w \text{pre}(s) \text{ then } \mathbf{M} \otimes A \models_{(w, s)} \varphi)\}.$$

The reason to employ multiple pointed models for updating is that it allows us to handle choice. Suppose we want to model the action of testing whether φ followed by a public announcement of the result. More precisely:

A test is performed to check whether φ holds in the actual world. If the outcome of the test is affirmative, then φ gets announced. If the test reveals that φ does not hold, then $\neg\varphi$ gets announced.

Single pointed update models do not allow us to model this.

THEOREM 10 (Preservation of bisimulation; Baltag, Moss, Solecki). *The action update operation $\otimes$ preserves bisimulation on epistemic models:*

$$\text{if } \mathbf{M} \underleftrightarrow{} \mathbf{N} \text{ then } \mathbf{M} \otimes \mathbf{A} \underleftrightarrow{} \mathbf{N} \otimes \mathbf{A}.$$

We can also look at the update models modulo action bisimulation. An action bisimulation is like an ordinary bisimulation, with the clause for 'same valuations' replaced by a clause for 'equivalent preconditions'.

THEOREM 11 (Preservation of action bisimulation). *The action update operation preserves action bisimulation:*

$$\text{if } \mathbf{A} \underleftrightarrow{} \mathbf{B} \text{ then } \mathbf{M} \otimes \mathbf{A} \underleftrightarrow{} \mathbf{M} \otimes \mathbf{B}.$$

Proof. Let Z be a bisimulation between $\mathbf{A}$ and $\mathbf{B}$. Define a relation relation on $\mathbf{M} \otimes \mathbf{A} \times \mathbf{M} \otimes \mathbf{B}$ by means of

$$(u, s)C(v, t) \text{ iff } u = v \text{ and } sZt.$$

It is easily shown that this is a bisimulation.   ■

4.2   Logics of Communication

In terms of the system just defined a variety of types of communicative actions can be described. The two most important ones are public announcements and group announcements.

Public Announcements

The language of **public announcements** is the language that one gets if one allows action models for public announcement. The action model for public announcement that φ consists of a single state s_0 with precondition φ and epistemic relation $\{s_0 \xrightarrow{a} s_0 \mid a \in Ag\}$. Call this model P_φ.

The following equivalence shows how public announcement relates to common knowledge among set of agents B:

(6) $[P_\varphi, s_0][B^*]\psi \leftrightarrow [(?\varphi; B)^*][P_\varphi, s_0]\psi.$

What this says is that after public announcement with φ it is common knowledge among B that ψ if and only if before the update it holds at the end of every $(?\varphi \; ; \; B)^*$ path through the model that a public update with φ will result in

ψ. Axiomatisations of public announcement logic are given in [Plaza, 1989] and [Gerbrandy, 1999b; 1999a], for a language that cannot express common knowledge. An axiomatisation for a language with a common knowledge operator is given in [Kooi and van Benthem, 2004]. Below we will show how this equivalence emerges in the axiomatisation of PDL$^{\mathrm{DEL}}$ from [van Eijck, 2004].

Group Announcements

The language of **group announcements** is the result of allowing action models for group messages. These will be defined below. Similarly, we can define the languages of **secret group communications**, of **individual messages**, of **tests**, of **lies**, and so on [Baltag, 2002]. All these languages are comprised in the language of PDL$^{\mathrm{DEL}}$, because all these communicative actions can be characterised by appropriate action models.

4.3 *Program Transformation*

We will now show how PDL$^{\mathrm{DEL}}$ formulae can be reduced to PDL formulae. For every action model A with states $s_0, \ldots, s_{n-1}$ we define a set of n^2 program transformers $T_{i,j}^A$ $(0 \leq i < n, 0 \leq j < n)$, as follows:

$$T_{ij}^A(a) = \begin{cases} ?\mathrm{pre}(s_i) \; ; \; a & \text{if } s_i \xrightarrow{a} s_j, \\ ?\bot & \text{otherwise} \end{cases}$$

$$T_{ij}^A(?\varphi) = \begin{cases} ?\varphi & \text{if } i = j, \\ ?\bot & \text{otherwise} \end{cases}$$

$$T_{ij}^A(\pi_1; \pi_2) = \bigcup_{k=0}^{n-1} (T_{ik}^A(\pi_1) \; ; \; T_{kj}^A(\pi_2))$$

$$T_{ij}^A(\pi_1 \cup \pi_2) = T_{ij}^A(\pi_1) \cup T_{ij}^A(\pi_2)$$

$$T_{ij}^A(\pi^*) = K_{ijn}^A(\pi)$$

where $K_{ijk}^A(\pi)$ is a (transformed) program for all the π^* paths from s_i to s_j that can be traced through A while avoiding a pass through intermediate states s_k and higher. Thus, $K_{ijn}^A(\pi)$ is a program for all the π^* paths from s_i to s_j that can be traced through A, period.

$K_{ijk}^A(\pi)$ is defined by recursion on k, as follows:

$$
K_{ij0}^A(\pi) \;\; = \;\;
\begin{cases}
?\top \cup T_{ij}^A(\pi) & \text{if } i = j, \\[2ex]
T_{ij}^A(\pi) & \text{otherwise}
\end{cases}
$$

$$
K_{ij(k+1)}^A(\pi) \;\; = \;\;
\begin{cases}
(K_{kkk}^A(\pi))^* & \text{if } i = k = j, \\[2ex]
(K_{kkk}^A(\pi))^*; K_{kjk}^A(\pi) & \text{if } i = k \neq j, \\[2ex]
K_{ikk}^A(\pi) \; ; \; (K_{kkk}^A(\pi))^* & \text{if } i \neq k = j, \\[2ex]
K_{ijk}^A(\pi) \cup (K_{ikk}^A(\pi) \; ; \; (K_{kkk}^A(\pi))^* \; ; \; K_{kjk}^A(\pi)) & \text{otherwise} \\
& (i \neq k \neq j).
\end{cases}
$$

For some runs through example applications of these definitions, see section 4.5 below.

LEMMA 12 (Kleene Path). *Suppose $(w, w') \in [\![T_{ij}^A(\pi)]\!]^{\mathbf{M}}$ iff there is a π path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$. Then $(w, w') \in [\![K_{ijn}^A(\pi)]\!]^{\mathbf{M}}$ iff there is a π^* path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$.*

Proof. Use the definition of K_{ijk}^A to prove by induction on k that $(w, w') \in [\![K_{ijk}^A(\pi)]\!]^{\mathbf{M}}$ iff there is a π^* path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$ that does not pass through any pairs (v, s) with $s \in \{s_k, \dots, s_{n-1}\}$.

Base case, $i = j$: A π^* path from (w, s_i) to (w', s_j) that does not visit any intermediate states is either the empty path or a single π step from (w, s_i) to (w', s_j). Such a path exists iff $(w, w') \in [\![?\top \cup T_{ij}^A]\!]^{\mathbf{M}}$ iff $(w, w') \in [\![K_{ij0}^A(\pi)]\!]^{\mathbf{M}}$.

Base case, $i \neq j$: A π^* path from (w, s_i) to (w', s_j) that does not visit any intermediate states is a single π step from (w, s_i) to (w', s_j). Such a path exists iff $(w, w') \in [\![T_{ij}^A]\!]^{\mathbf{M}}$ iff $(w, w') \in [\![K_{ij0}^A(\pi)]\!]^{\mathbf{M}}$.

Induction step. Assume that $(w, w') \in [\![K_{ijk}^A(\pi)]\!]^{\mathbf{M}}$ iff there is a π^* path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$ that does not pass through any pairs (v, s) with $s \in \{s_k, \dots, s_{n-1}\}$.

Case $i = k = j$. A path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$ that does not pass through any pairs (v, s) with $s \in \{s_{k+1}, \dots, s_{n-1}\}$ now consists of an arbitrary number of π^* paths from s_k to s_k that do not visit any intermediate states with action component s_k or higher. By the induction hypothesis, such a path exists iff $(w, w') \in [\![(K_{kkk}^A(\pi))^*]\!]^{\mathbf{M}}$ iff $(w, w') \in [\![K_{ij(k+1)}^A(\pi)]\!]^{\mathbf{M}}$.

Case $i = k \neq j$. A path from (w, s_i) to (w', s_j) in $\mathbf{M} \otimes A$ that does not pass through any pairs (v, s) with $s \in \{s_{k+1}, \dots, s_{n-1}\}$ now consists of a π^* path starting in (w, s_k) visiting states of the form (u, s_k) an arbitrary number of times, but never touching on states with action component s_k or higher in between, and ending in (v, s_k), followed by a π^* path from (v, s_k) to (w', s_j) that does not pass through any pairs (v, s) with $s \in \{s_k, \dots, s_{n-1}\}$. By the induction hypothesis,

a path from (w, s_k) to (v, s_k) of the first kind exists iff $(w, v) \in [\![(K_{kkk}^A(\pi))^*]\!]^{\mathbf{M}}$.
Again by the induction hypothesis, a path from (v, s_k) to (w', s_j) of the second
kind exists iff $(v, w') \in [\![K_{kjk}^A]\!]^{\mathbf{M}}$. Thus, the required path from (w, s_i) to (w', s_j)
in $\mathbf{M} \otimes A$ exists iff $(w, w') \in [\![(K_{kkk}^A(\pi))^*; K_{kjk}^A(\pi)]\!]^{\mathbf{M}}$ iff $(w, w') \in [\![K_{ij(k+1)}^A(\pi)]\!]^{\mathbf{M}}$.

The other two cases are similar.   ∎

The Kleene path lemma is the key ingredient in the following program trans-
formation lemma.

LEMMA 13 (Program Transformation). *Assume A has n states $s_0, \ldots, s_{n-1}$. Then:*

$$\mathbf{M} \models_w [A, s_i][\pi]\varphi \ \textit{iff} \ \mathbf{M} \models_w \bigwedge_{j=0}^{n-1} [T_{ij}^A(\pi)][A, s_j]\varphi.$$

Proof. Induction on the complexity of π.
    Basis, epistemic link case:

$$\mathbf{M} \models_w [A, s_i][a]\varphi$$

iff $\mathbf{M} \models_w \operatorname{pre}(s_i)$ implies $\mathbf{M} \otimes A \models_{(w,s_i)} [a]\varphi$

iff $\mathbf{M} \models_w \operatorname{pre}(s_i)$ implies for all $s_j \in A$, all $w' \in \mathbf{M}$:

   if $s_i \xrightarrow{a} s_j, w \xrightarrow{a} w'$, then $\mathbf{M} \models_{w'} [A, s_j]\varphi$

iff   for all $s_j \in A$: if $s_i \xrightarrow{a} s_j$ then $\mathbf{M} \models_w [\operatorname{pre}(s_i) \ ; \ a][A, s_j]\varphi$

iff $\mathbf{M} \models_w \bigwedge_{j=0}^{n-1} [T_{ij}^A(a)][A, s_j]\varphi.$

Basis, test case:

$$\mathbf{M} \models_w [A, s_i][?\psi]\varphi$$

iff $\mathbf{M} \models_w \operatorname{pre}(s_i)$ implies $\mathbf{M} \otimes A \models_{(w,s_i)} [?\psi]\varphi$

iff $\mathbf{M} \models_w \operatorname{pre}(s_i)$ implies $\mathbf{M} \models_w [?\psi][A, s_i]\varphi$

iff $\mathbf{M} \models_w \bigwedge_{j=0}^{n-1} [T_{ij}^A(?\psi)][A, s_j]\varphi.$

Induction step, cases $\pi_1 \ ; \ \pi_2$ and $\pi_1 \cup \pi_2$ are straightforward. The case of π^* is
settled with the help of the Kleene path lemma.   ∎

4.4   Reduction Axioms for Update PDL

The program transformations can be used to translate PDL$^{\mathrm{DEL}}$ to PDL by means
of the following mutually recursive definitions of translations t for formulae and r
for programs:

$$
\begin{aligned}
t(\top) &= \top \\
t(p) &= p \\
t(\neg\varphi) &= \neg t(\varphi) \\
t(\varphi_1 \wedge \varphi_2) &= t(\varphi_1) \wedge t(\varphi_2) \\
t([\pi]\varphi) &= [r(\pi)]t(\varphi) \\
t([A,s]\top) &= \top \\
t([A,s]p) &= t(\mathrm{pre}(s)) \rightarrow p \\
t([A,s]\neg\varphi) &= t(\mathrm{pre}(s)) \rightarrow \neg t([A,s]\varphi) \\
t([A,s](\varphi_1 \wedge \varphi_2)) &= t([A,s]\varphi_1) \wedge t([A,s]\varphi_2) \\
t([A,s_i][\pi]\varphi) &= \bigwedge_{j=0}^{n-1} [T_{ij}^A(r(\pi))]t([A,s_j]\varphi) \\
t([A,s][A',s']\varphi) &= t([A,s]t([A',s']\varphi))
\end{aligned}
$$

$$
\begin{aligned}
r(a) &= a \\
r(?\varphi) &= ?t(\varphi) \\
r(\pi_1;\pi_2) &= r(\pi_1);r(\pi_2) \\
r(\pi_1 \cup \pi_2) &= r(\pi_1) \cup r(\pi_2) \\
r(\pi^*) &= (r(\pi))^*.
\end{aligned}
$$

The correctness of this translation follows from direct semantic inspection, using the program transformation lemma for the translation of $[A, s_i][\pi]\varphi$ formulae. The translation points the way to appropriate reduction axioms, as follows.

Take all axioms and rules of PDL [Segerberg, 1982; Fischer and Ladner, 1979; Parikh, 1978], plus the following reduction axioms:

$$
\begin{aligned}
[A,s]p &\leftrightarrow (\mathrm{pre}(s) \Rightarrow p) \\
[A,s]\neg\varphi &\leftrightarrow (\mathrm{pre}(s) \Rightarrow \neg[A,s]\varphi) \\
[A,s](\varphi_1 \wedge \varphi_2) &\leftrightarrow ([A,s]\varphi_1 \wedge [A,s]\varphi_2) \\
[A,s_i][\pi]\varphi &\leftrightarrow \bigwedge_{j=0}^{n-1} [T_{ij}^A(\pi)][A,s_j]\varphi
\end{aligned}
$$

and necessitation for action model modalities. The reduction axioms for $[A,s]p$, $[A,s]\neg\varphi$ and $[A,s](\varphi_1 \wedge \varphi_2)$ are as in [Kooi and van Benthem, 2004]. The final reduction axiom is based on program transformation and is new. Note that if we allow multiple action models, we need the following reduction axiom for those:

$$
[A,S]\varphi \leftrightarrow \bigwedge_{s \in S} [A,s]\varphi
$$

If updates with multiple pointed action models are also in the language, we need the following additional reduction axiom:

$$[A, S]\varphi \quad \leftrightarrow \quad \bigwedge_{s \in S} [A, s]\varphi$$

THEOREM 14 (Completeness). *If* $\models \varphi$ *then* $\vdash \varphi$.

Proof. The proof system for PDL is complete, and every formula in the language of PDL$^{\mathrm{DEL}}$ is provably equivalent to a PDL formula. ∎

4.5 Special Cases

Public Announcement and Common Knowledge

As introduced above, in section 4.2.1, the action model for public announcement that φ consists of a single state s_0 with precondition φ and epistemic relation $\{s_0 \xrightarrow{a} s_0 \mid a \in Ag\}$. We call this model P_φ.
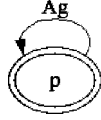
We are interested in how public announcement that φ brings about common knowledge of ψ among group of agents B, i.e. we want to compute $[P_\varphi, s_0][B^*]\psi$. For this, we need $T_{00}^{P_\varphi}(B^*)$, which is defined as $K_{001}^{P_\varphi}(B)$.

To work out $K_{001}^{P_\varphi}(B)$, we need $K_{000}^{P_\varphi}(B)$, and for $K_{000}^{P_\varphi}(B)$, we need $T_{00}^{P_\varphi}(B)$, which turns out to be $\bigcup_{b \in B}(?\varphi \ ; \ b)$, or equivalently, $?\varphi; B$. Working upward from this, we get:

$$K_{000}^{P_\varphi}(B) = ?\top \cup T_{00}^{P_\varphi}(B) = ?\top \cup (?\varphi; B),$$

and therefore:

$$
\begin{aligned}
K_{001}^{P_\varphi}(B) &= (K_{000}^{P_\varphi}(B))^* \\
&= (?\top \cup (?\varphi; B))^* \\
&= (?\varphi; B)^*.
\end{aligned}
$$

Thus, the reduction axiom for the public announcement action P_φ with respect to the program for common knowledge among agents B, works out as follows:

$$
\begin{aligned}
[P_\varphi, s_0][B^*]\psi &\leftrightarrow [P_\varphi, s_0][B^*]\psi \\
&\leftrightarrow [T_{00}^{P_\varphi}(B^*)][P_\varphi, s_0]\psi \\
&\leftrightarrow [K_{001}^{P_\varphi}(B)][P_\varphi, s_0]\psi \\
&\leftrightarrow [(?\varphi; B)^*][P_\varphi, s_0]\psi.
\end{aligned}
$$

This expresses that every B path consisting of φ worlds ends in a $[P_\varphi, s_0]\psi$ world, i.e. it expresses exactly what is captured by the special purpose operator $C_B(\varphi, \psi)$ introduced in [Kooi and van Benthem, 2004]. Indeed, the authors remark in a footnote that their proof system for $C_B(\varphi, \psi)$ essentially follows the usual PDL treatment for the PDL transcription of this formula.

Secret Group Communication and Common Belief

The logic of secret group communication is the logic of email CCs. The action model for a secret group message to B that φ consists of two states s_0, s_1, where s_0 has precondition φ and s_1 has precondition $\top$, and where the accessibilities T are given by:

$$
\begin{aligned}
T \quad = \quad & \{s_0 \xrightarrow{b} s_0 \mid b \in B\} \\
& \cup \{s_0 \xrightarrow{a} s_1 \mid a \in Ag - B\} \\
& \cup \{s_1 \xrightarrow{a} s_1 \mid a \in Ag\}.
\end{aligned}
$$

The actual world is s_0. The members of B are aware that action φ takes place; the others think that nothing happens. In this they are mistaken, which is why CC updates generate KD45 models: i.e. CC updates make knowledge degenerate into belief.
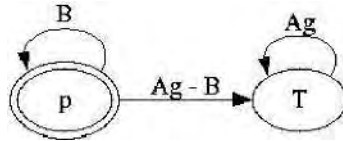
We work out the program transformations that this update engenders for common knowledge among some group of agents D. Call the action model CC_φ^B.

We will have to work out $K_{002}^{CC_\varphi^B} D$, $K_{012}^{CC_\varphi^B} D$, $K_{112}^{CC_\varphi^B} D$, $K_{102}^{CC_\varphi^B} D$.

For these, we need $K_{001}^{CC_\varphi^B} D$, $K_{011}^{CC_\varphi^B} D$, $K_{111}^{CC_\varphi^B} D$, $K_{101}^{CC_\varphi^B} D$.

For these in turn, we need $K_{000}^{CC_\varphi^B} D$, $K_{010}^{CC_\varphi^B} D$, $K_{110}^{CC_\varphi^B} D$, $K_{100}^{CC_\varphi^B} D$.

For these, we need:

$$
\begin{aligned}
T_{00}^{CC_\varphi^B} D \quad &= \quad \bigcup_{d \in B \cap D} (?\varphi \; ; \; d) \; = \; ?\varphi \; ; \; (B \cap D) \\
T_{01}^{CC_\varphi^B} D \quad &= \quad \bigcup_{d \in D - B} (?\varphi \; ; \; d) \; = \; ?\varphi \; ; \; (D - B) \\
T_{11}^{CC_\varphi^B} D \quad &= \quad D \\
T_{10}^{CC_\varphi^B} D \quad &= \quad ?\bot
\end{aligned}
$$

It follows that:

$$K_{000}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\top \cup (?\varphi \;;\; (B \cap D))$$
$$K_{010}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\varphi \;;\; (D - B)$$
$$K_{110}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\top \cup D,$$
$$K_{100}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\bot$$

From this we can work out the K_{ij1}, as follows:

$$K_{001}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; (?\varphi \;;\; (B \cap D))^*$$
$$K_{011}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; (?\varphi \;;\; (B \cap D))^* \;;\; (D - B)$$
$$K_{111}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\top \cup D$$
$$K_{101}^{\mathrm{CC}_\varphi^B} D \;\;=\;\; ?\bot.$$

Finally, we get K_{002} and K_{012} from this:

$$
\begin{aligned}
K_{002}^{\mathrm{CC}_\varphi^B} D \;\;&=\;\; K_{001}^{\mathrm{CC}_\varphi^B} D \cup K_{011}^{\mathrm{CC}_\varphi^B} D \;;\; (K_{111}^{\mathrm{CC}_\varphi^B} D)^* \;;\; K_{101}^{\mathrm{CC}_\varphi^B} D \\
&=\;\; K_{001}^{\mathrm{CC}_\varphi^B} D \;\; \text{(since the right-hand expression evaluates to } ?\bot) \\
&=\;\; (?\varphi \;;\; (B \cap D))^* \\
K_{012}^{\mathrm{CC}_\varphi^B} D \;\;&=\;\; K_{011}^{\mathrm{CC}_\varphi^B} D \cup K_{011}^{\mathrm{CC}_\varphi^B} D \;;\; (K_{111}^{\mathrm{CC}_\varphi^B} D)^* \\
&=\;\; K_{011}^{\mathrm{CC}_\varphi^B} D \;;\; (K_{111}^{\mathrm{CC}_\varphi^B} D)^* \\
&=\;\; (?\varphi \;;\; (B \cap D))^* \;;\; (D - B) \;;\; D^*.
\end{aligned}
$$

Thus, the program transformation for common belief among D works out as follows:

$$[\mathrm{CC}_\varphi^B, s_0][D^*]\psi$$
$$\leftrightarrow \;\; [(?\varphi \;;\; (B \cap D))^*][\mathrm{CC}_\varphi^B, s_0]\psi \wedge [(?\varphi \;;\; (B \cap D))^* \;;\; (D - B) \;;\; D^*][\mathrm{CC}_\varphi^B, s_1]\psi.$$

Compare [Ruan, 2004] for a direct axiomatisation of the logic of CCs.

Group Messages and Common Knowledge

The action model for a group message to B that φ consists of two states s_0, s_1, where s_0 has precondition φ and s_1 has precondition $\top$, and where the accessibilities T are given by:

$$
\begin{aligned}
T \;\;=\;\; & \{s_0 \xrightarrow{b} s_0 \mid b \in B\} \\
& \cup \{s_1 \xrightarrow{b} s_1 \mid b \in B\} \\
& \cup \{s_0 \xrightarrow{a} s_1 \mid a \in Ag - B\} \\
& \cup \{s_1 \xrightarrow{a} s_0 \mid a \in Ag - B\}.
\end{aligned}
$$

This captures the fact that the members of B can distinguish the φ update from the $\top$ update, while the other agents (the members of $Ag - B$) cannot. The actual action is s_0. Call this model G_φ^B.
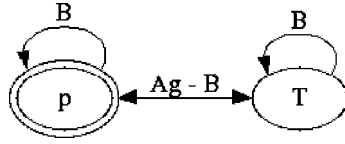
A difference with the CC case is that group messages are S5 models. Since updates of S5 models with S5 models are S5, group messages engender common knowledge (as opposed to mere common belief). Let us work out the program transformation that this update engenders for common knowledge among some group of agents D.

We will have to work out $K_{002}^{G_\varphi^B} D$, $K_{012}^{G_\varphi^B} D$, $K_{112}^{G_\varphi^B} D$, $K_{102}^{G_\varphi^B} D$.

For these, we need $K_{001}^{G_\varphi^B} D$, $K_{011}^{G_\varphi^B} D$, $K_{111}^{G_\varphi^B} D$, $K_{101}^{G_\varphi^B} D$.

For these in turn, we need $K_{000}^{G_\varphi^B} D$, $K_{010}^{G_\varphi^B} D$, $K_{110}^{G_\varphi^B} D$, $K_{100}^{G_\varphi^B} D$.

For these, we need:

$$T_{00}^{G_\varphi^B} D \;=\; \bigcup_{d \in D} (?\varphi \;;\; d) = ?\varphi \;;\; D,$$

$$T_{01}^{G_\varphi^B} D \;=\; \bigcup_{d \in D-B} (?\varphi \;;\; d) = ?\varphi \;;\; (D - B),$$

$$T_{11}^{G_\varphi^B} D \;=\; D,$$

$$T_{10}^{G_\varphi^B} D \;=\; D - B.$$

It follows that:

$$K_{000}^{G_\varphi^B} D \;=\; ?\top \cup (?\varphi \;;\; D),$$

$$K_{010}^{G_\varphi^B} D \;=\; ?\varphi \;;\; (D - B),$$

$$K_{110}^{G_\varphi^B} D \;=\; ?\top \cup D,$$

$$K_{100}^{G_\varphi^B} D \;=\; D - B.$$

From this we can work out the K_{ij1}, as follows:

$$K_{001}^{G_\varphi^B} D \;=\; (?\varphi \;;\; D)^*,$$

$$K_{011}^{G_\varphi^B} D \;=\; (?\varphi \;;\; D)^* \;;\; ?\varphi \;;\; D - B,$$

$$K_{111}^{G_\varphi^B} D \;=\; ?\top \cup D \cup (D - B; (?\varphi \;;\; D)^*; ?\varphi \;;\; D - B),$$

$$K_{101}^{G_\varphi^B} D \;=\; D - B \;;\; (?\varphi; D)^*.$$

Finally, we get K_{002} and K_{012} from this:

$$K_{002}^{G_\varphi^B} D = K_{001}^{G_\varphi^B} D \cup K_{011}^{G_\varphi^B} D \;\; ; \;\; (K_{111}^{G_\varphi^B} D)^* \;\; ; \;\; K_{101}^{G_\varphi^B} D$$
$$= (?\varphi \;\; ; \;\; D)^* \cup$$
$$(?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B;$$
$$(D \cup (D - B; (?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B))^* \;\; ; \;\; D - B \;\; ; \;\; (?\varphi \;\; ; \;\; D)^*,$$
$$K_{012}^{G_\varphi^B} D = K_{011}^{G_\varphi^B} D \;\; ; \;\; (K_{111}^{G_\varphi^B} D)^*$$
$$= (?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B \;\; ; \;\; (D \cup (D - B \;\; ; \;\; (?\varphi; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B))^*.$$

Abbreviating $D \cup (D - B \;\; ; \;\; (?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B)$ as π, we get the following transformation for common knowledge among D after a group message to B that φ:

$$[G_\varphi^B, s_0][D^*]\psi$$
$$\leftrightarrow$$
$$[(?\varphi \;\; ; \;\; D)^* \cup ((?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B \;\; ; \;\; \pi^* \;\; ; \;\; D - B \;\; ; \;\; (?\varphi \;\; ; \;\; D)^*)][G_\varphi^B, s_0]\psi$$
$$\wedge$$
$$[(?\varphi \;\; ; \;\; D)^* \; ; \; ?\varphi \;\; ; \;\; D - B \;\; ; \;\; \pi^*][G_\varphi^B, s_1]\psi.$$

This equivalence gives a precise characterisation of two path requirements that have to hold in the original model in order for common knowledge among D to result from the group message to B. The formula may look complicated, but mechanical verification of the requirement is quite easy.

5 QUANTIFIED DYNAMIC LOGIC

The second core system of dynamic logic that will be discussed in detail is that of *quantified dynamic logic* (QDL). QDL was developed by Harel [1979] and Goldblatt [1982]. Both monographs were inspired by Pratt [1976]. Further information about the development of QDL can be found in [Harel, 1984; Harel *et al.*, 2000; Goldblatt, 1992/1987].

Quantified dynamic logic can be viewed as the first order version of propositional dynamic logic. It is less abstract than PDL, for program atoms now get further analysed as assignments of values to program variables or as relational tests, and states take the concrete shape of mappings from program variables to appropriate values. At the background is a first order structure **M** consisting of a domain plus interpretations of relation and function symbols.

Recall that the assignment programs of WHILE looked like $v := t$, with v a program variable and t a term of the WHILE language. In QDL, the basic actions are:

- assigning a random value to a variable:

$$v :=?,$$

- assigning a definite value to a variable:

$$v := t,$$

- and testing for the truth of a formula:

$$?\varphi.$$

Various versions of QDL result from imposing further restrictions on testing, e.g., by only allowing tests on boolean combinations of relational and equational atoms.

Consider a state where x has value 3 and y value 2. Assuming we are computing on the natural numbers, random assignment of a new value to x causes infinite branching to the states with

$$\{x \mapsto 0, y \mapsto 2\}, \{x \mapsto 1, y \mapsto 2\}, \{x \mapsto 2, y \mapsto 2\}, \{x \mapsto 3, y \mapsto 2\},$$

and so on. The subsequent test $x = y$ only succeeds for the state with $\{x \mapsto 2, y \mapsto 2\}$. The nett effect of $x :=? \; ; \; ?(x = y)$ is a transition from $\{x \mapsto 3, y \mapsto 2\}$ to $\{x \mapsto 2, y \mapsto 2\}$.
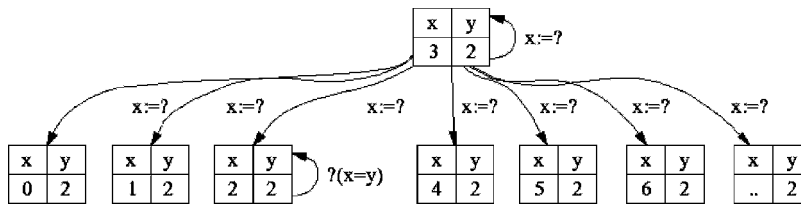
5.1 Language

Take a signature for first order logic. Define terms, formulae and programs, as follows:

$$
\begin{aligned}
t &\; ::= \; v \mid ft_1 \cdots t_n \\
\varphi &\; ::= \; \top \mid Rt_1 \cdots t_n \mid t_1 = t_2 \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \exists v\varphi \mid \langle \pi \rangle \varphi \\
\pi &\; ::= \; v := ? \mid v := t \mid ?\varphi \mid \pi_1 \; ; \; \pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^*
\end{aligned}
$$

Abbreviations are as in the case of PDL. In particular, the *SKIP, ABORT, WHILE, REPEAT, IF-THEN-ELSE* constructs are also defined as in the case of PDL. What Quantified Dynamic Logic gives us is a fleshed out version of PDL, with assignments (random and definite) and tests as basic actions. The assignments change relational structures, and therefore the appropriate assertion language is built from first order predicate logic rather than propositional logic, as in PDL.

Floyd-Hoare correctness statements

for WHILE programs can be expressed directly in QDL. Recall the example of the correctness statement for the factorial program from section 2.4:

$$x! = Z \rightarrow [y := 1 \ ; \ \text{WHILE } x \neq 1 \text{ DO } (y := y * x \ ; \ x := x \dot{-} 1)] y = Z.$$

This expresses partial correctness of the factorial program. Total correctness of the factorial program can be expressed in QDL as the conjunction of the above with the following:

$$\langle y := 1 \ ; \ \text{WHILE } x \neq 1 \text{ DO } (y := y * x \ ; \ x := x \dot{-} 1) \rangle \top.$$

5.2 Semantics

A first order signature is a pair $(\mathbf{f}, \mathbf{R})$ where $\mathbf{f}$ is a list of function symbols with their arities and $\mathbf{R}$ is a list of relation symbols with their arities. Nullary function symbols are individual constants, nullary relation symbols are propositional constants, unary relation symbols are predicates.

A model for a signature $(\mathbf{f}, \mathbf{R})$ is a structure of the form

$$\mathbf{M} = (E^{\mathbf{M}}, f^{\mathbf{M}}, \ldots, R^{\mathbf{M}}, \ldots),$$

where E is a non-empty set, the $f^{\mathbf{M}}$ are interpretations in E for the members of $\mathbf{f}$ (i.e. if f is an n-ary function symbol, then $f^{\mathbf{M}} : E^n \rightarrow E$), and the $R^{\mathbf{M}}$ are interpretations in E for the members of $\mathbf{R}$ (i.e. if R is an n-ary relation symbol, then $R^{\mathbf{M}} \subseteq E^n$).

Let V be the set of variables of the language. As usual $g \sim_v h$ expresses that state h differs at most from state g on v. Interpretation of terms in $\mathbf{M}$ is defined relative to a variable assignment $g : V \rightarrow E^{\mathbf{M}}$, as follows:

$$\begin{aligned}
\llbracket v \rrbracket_g^{\mathbf{M}} &= g(v) \\
\llbracket f t_1 \cdots t_n \rrbracket_g^{\mathbf{M}} &= f^{\mathbf{M}}(\llbracket t_1 \rrbracket_g^{\mathbf{M}}, \ldots, \llbracket t_n \rrbracket_g^{\mathbf{M}})
\end{aligned}$$

Truth in $\mathbf{M}$ for formulae and relational meaning in $\mathbf{M}$ for programs are defined by simultaneous recursion:

$$\begin{aligned}
\mathbf{M} &\models_g \top & &\text{always} \\
\mathbf{M} &\models_g R t_1 \cdots t_n & \text{iff} \quad & (\llbracket t_1 \rrbracket_g^{\mathbf{M}}, \ldots, \llbracket t_n \rrbracket_g^{\mathbf{M}}) \in R^{\mathbf{M}} \\
\mathbf{M} &\models t_1 = t_2 & \text{iff} \quad & \llbracket t_1 \rrbracket_g^{\mathbf{M}} \text{ is the same as } \llbracket t_2 \rrbracket_g^{\mathbf{M}} \\
\mathbf{M} &\models_g \neg\varphi & \text{iff} \quad & \text{not } \mathbf{M} \models_g \varphi \\
\mathbf{M} &\models_g \varphi_1 \vee \varphi_2 & \text{iff} \quad & \mathbf{M} \models_g \varphi_1 \text{ or } \mathbf{M} \models_g \varphi_2 \\
\mathbf{M} &\models_g \exists v \varphi & \text{iff} \quad & \text{for some } h \text{ with } g \sim_v h, \mathbf{M} \models_h \varphi \\
\mathbf{M} &\models_g \langle \pi \rangle \varphi & \text{iff} \quad & \text{for some } h \text{ with } {}_g\llbracket \pi \rrbracket_h^{\mathbf{M}}, \mathbf{M} \models_h \varphi
\end{aligned}$$

$$
\begin{aligned}
{}_g[\![v := ?]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad g \sim_v h \\
{}_g[\![v := t]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad h \text{ equals } g[v \mapsto [\![t]\!]_g^{\mathbf{M}}] \\
{}_g[\![?\varphi]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad g = h \text{ and } \mathbf{M} \models_g \varphi \\
{}_g[\![\pi_1 \ ; \ \pi_2]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad \text{there is an assignment } f \text{ with} \\
&\qquad {}_g[\![\pi_1]\!]_f^{\mathbf{M}} \text{ and } {}_f[\![\pi_2]\!]_h^{\mathbf{M}} \\
{}_g[\![\pi_1 \cup \pi_2]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad {}_g[\![\pi_1]\!]_h^{\mathbf{M}} \text{ or } {}_g[\![\pi_2]\!]_h^{\mathbf{M}} \\
{}_g[\![\pi^*]\!]_h^{\mathbf{M}} \quad &\text{iff} \quad (g, h) \in ([\![\pi]\!]^{\mathbf{M}})^*
\end{aligned}
$$

Validity of QDL formulae over a given signature is defined in terms of truth in all models for the signature. A QDL formula φ over a given signature is satisfiable if there is model $\mathbf{M}$ for that signature together with a variable assignment g in the domain of that model, such that $\mathbf{M} \models_g \varphi$.

Note that the presence of $v := ?$ does not increase the expressive power of the language. Indeed, we have the following validities:

$$
\begin{aligned}
\exists v \varphi \quad &\leftrightarrow \quad \langle v := ? \rangle \varphi \\
\forall v \varphi \quad &\leftrightarrow \quad [v := ?] \varphi
\end{aligned}
$$

Next, if v does not occur in t, definite assignment of t to v is equivalent to random assignment to v followed by a test of the equality $v = t$. In other words, if v does not occur in t we have the following validities:

$$
\begin{aligned}
\langle v := t \rangle \varphi \quad &\leftrightarrow \quad \langle v := ? \ ; \ ?v = t \rangle \varphi \\
[v := t] \varphi \quad &\leftrightarrow \quad [v := ? \ ; \ ?v = t] \varphi.
\end{aligned}
$$

Substitution and Assignment

The computational process of assigning a value to a variable is intimately linked to the syntactic process of making a substitution of a term for a variable.

Recall the situation in first order logic. There, the basic truth definition is phrased in terms of a first order model $\mathbf{M}$, a variable assignment g, and a formula φ: $\mathbf{M} \models_g \varphi$ means that variable assignment g makes φ true in $\mathbf{M}$. Let t_s^v be the result of replacing variable v everywhere in term t by term s. Then the following term substitution lemma holds for FOL and for QDL:

LEMMA 15 (Term substitution). $[\![t_s^v]\!]_g^{\mathbf{M}} = [\![t]\!]_{g[v \mapsto [\![s]\!]_g^{\mathbf{M}}]}^{\mathbf{M}}$.

This is easily proved with induction on the structure of t.

Using this, one can prove the substitution lemma for FOL. Recall that a term t is substitutable for v in φ (or: free for v in φ) if the substitution process does not cause accidental capture of variables in t. Use φ_t^v for the result of substituting t for all free occurrences of v in φ. The following holds for FOL:

LEMMA 16 (Substitution). *If t is free for v in φ then*

$$\mathbf{M} \models_g \varphi_t^v \ \textit{iff} \ \mathbf{M} \models_{g[v \mapsto [\![t]\!]_g^{\mathbf{M}}]} \varphi.$$

The proof uses induction on the structure of φ, using the term substitution lemma for the atomic case. In the case of QDL, we can rephrase this as follows:

LEMMA 17 (Assignment).

$$\mathbf{M} \models_g [v := t]\varphi \ \textit{iff} \ \mathbf{M} \models_{g[v \mapsto [\![t]\!]_g^{\mathbf{M}}]} \varphi.$$

What this means is that in QDL we can replace syntactic substitutions φ_t^v by $[v := t]\varphi$.

Below we will be interested in the subsystem of QDL defined by

$$\pi \quad ::= \quad ?Rt_1 \cdots t_2 \mid ?t_1 = t_2 \mid v :=? \mid \sim\pi \mid \pi_1 \ ; \ \pi_2.$$

where $\sim\pi$ is an abbreviation of $?[\pi]\bot$.

It turns out that this subsystem, baptised DPL in [Groenendijk and Stokhof, 1991], has the same expressive power as first order logic, but its quantifier $v :=?$ has different binding behaviour from the quantifiers of first order logic. [Groenendijk and Stokhof, 1991] proposes to employ the dynamic binding behaviour of the DPL quantifiers for analysing anaphoric linking (establishing the links between pronouns and their antecedents) in natural language.

Expressiveness

We can immediately see that the expressive power of QDL is greater than that of FOL. The following formula in the language of natural number arithmetic expresses induction on the natural numbers:

(7) $\quad \forall y \langle x := 0 \ ; \ \text{WHILE } x \neq y \text{ DO } x := x + 1 \rangle \top.$

This asserts that for all y the program $x := 0 \ ; \ \text{WHILE } x \neq y \text{ DO } x := x+1$ has a terminating execution. That is, every y can be reached by starting from 0 and repeatedly applying the successor function. This defines the natural numbers up to isomorphism, and no first order formula can do that. Let $\varphi_{\mathbb{N}}$ be the conjunction of formula (7) with the Peano axioms for arithmetic except the induction axiom. Then the valid QDL sentences of the form $\varphi_{\mathbb{N}} \to \psi$, with ψ a first order sentence, specify the first order sentences ψ that are true on $\mathbb{N}$. But we know from Gödel's incompleteness theorem and Church's Thesis that this set of sentences cannot be effectively enumerated.

5.3 *Interpreted versus Uninterpreted Reasoning*

As was the case with the WHILE-language and other systems, we are often interested in computation with respect to some standard structure, such as the natural

numbers. In such cases, we evaluate QDL formulae and programs in this structure, and talk, e.g., about $\mathbb{N}$-validity: truth for all variable assignments in $\mathbb{N}$, and so on.

Note that all WHILE programs over a given signature are QDL programs over that same signature. Thus, we can use QDL for making assertions about the behaviour of WHILE programs. When interpreting with respect to $\mathbb{N}$, we can specify Euclid's GCD algorithm as the following QDL program:

$$\pi = \text{ WHILE } x \neq y \text{ DO IF } x > y \text{ THEN } x := x - y \text{ ELSE } y := y - x.$$

Clearly, Floyd-Hoare correctness statements about WHILE programs can be expressed in QDL. E.g., the following QDL statements about the GCD program, expressing the total correctness of the program, are valid in $\mathbb{N}$:

$$(x = x' \wedge y = y' \wedge x \times y > 0) \to [\pi] \ x = \gcd(x', y').$$

$$x \times y > 0 \to \langle \pi \rangle \top$$

The first of these says that if program π over $\mathbb{N}$ terminates then in the output state x holds the value of the GCD of x' and y'. The second of these expresses that the program does indeed terminate for all states with $x \times y > 0$, for $\langle \pi \rangle \top$ expresses termination for all deterministic programs.

In the case of uninterpreted reasoning we are interested in truth in all structures. The following is valid in all models:

$$(x = x' \wedge y = y') \to [z := x \ ; \ x := y \ ; \ y := z](x = y' \wedge y = x').$$

5.4   Undecidability and Completeness

QDL is a proper extension of classical FOL, and, as we have seen, its validity problem is not effectively enumerable. This means that there can be no proof theory for QDL based on an enumerable set of axioms and an enumerable set of decidable inference rules. A proof theory will have to be based on infinitary (hence undecidable) inference rules.

The following axioms relate random assignment to quantification and definite assignment to substitution:

$$
\begin{array}{ll}
\forall v \varphi \leftrightarrow [v := ?]\varphi & \\
\forall v \varphi \to [v := t]\varphi & \\
\forall w [v := w]\varphi \to \forall v \varphi & w \notin \{v\} \cup var(\varphi) \\
\forall v \varphi \to [v := t]\forall v \varphi & \\
\forall w [v := t]\varphi \to [v := t]\forall w \varphi & w \notin \{v\} \cup var(t) \\
\langle v := t \rangle \varphi \leftrightarrow [v := t]\varphi & \\
[v := t]Rt_1 \cdots t_n \leftrightarrow Rt_{1t}^{v} \cdots t_{nt}^{v} & \\
[v := t]t_1 = t_2 \leftrightarrow t_{1t}^{v} = t_{2t}^{v} & \\
[v := t][v := s]\varphi \to [v := s_t^v]\varphi & \\
[v := t][w := s]\varphi \to [w := s_t^v][v := t]\varphi & w \notin \{v\} \cup var(t) \\
s = t \to ([v := t]\varphi \leftrightarrow [v := s]\varphi) &
\end{array}
$$

Now take as axiom schemes the following:

- All instances of valid FOL formulae,

- all instances of valid PDL formulae,

- the assignment axiom schemes above,

and as rules of inference:

- modus ponens

- quantifier generalisation

$$\frac{\varphi}{\forall v \varphi}$$

- program generalisation

$$\frac{\varphi}{[\pi]\varphi}$$

- and infinitary convergence:

$$\frac{\varphi \rightarrow [\pi^n]\psi, \quad n \in \mathbb{N}}{\varphi \rightarrow [\pi^*]\psi}$$

where π^n is given by $\pi^0 = ?\top$, $\pi^{n+1} = \pi \;;\; \pi^n$

A proof in this calculus may have infinitely many premises. This infinitary proof system is sound and complete (Harel [1984] or Goldblatt [1992/1987]):

THEOREM 18. *For any QDL formula φ,*

$$\models \varphi \; \text{iff} \; \vdash \varphi.$$

6 DPL AS A FRAGMENT OF QDL

In the introduction we mentioned that dynamic logic is also used in linguistics, in particular in the analysis of various phenomena involving information flow in discourse (text, conversation). In this section we turn to the study of a particular formalism, that of Dynamic Predicate Logic (DPL), that has played a prominent role in the development of dynamic semantic theories for natural language.

The DPL system is a representative instance of a whole variety of systems that have been developed in formal semantics of natural language to deal with dynamic aspects of meaning and information flow: the contribution of declaratives to the 'common ground', presuppositional phenomena, anaphoric links across sentence boundaries, the temporal structure of discourse, the semantic effects of imperatives, and so on. DPL is an illustrative example in the present context because of its obvious affinities with systems developed in other areas, in particular with PDL and QDL. The formal properties of the DPL system have been studied quite extensively (cf., e.g., [van Benthem, 1996] and the references below). Also, DPL provides a nice illustration of some of the central concepts of QDL. A more eleborate discussion of the specific linguistic issues involved can be found in section 7.

6.1   System

DPL is the subsystem of QDL that is given by the following syntax:

DEFINITION 19 (DPL syntax).

$$t \quad ::= \quad v \mid c \mid f t_1 \cdots t_n$$
$$\pi \quad ::= \quad ? R t_1 \cdots t_2 \mid ? t_1 = t_2 \mid v := ? \mid \sim\pi \mid \pi_1 \; ; \quad \pi_2.$$

Semantics: as in the definition of QDL. The meaning of $\sim\pi$ is given by:

$$_g[\![\sim\pi]\!]_h \quad \text{iff} \quad g \text{ equals } h \text{ and for no } g' \text{ it holds that } _g[\![\pi]\!]^{\mathbf{M}}_{g'}.$$

As was noted earlier, $\sim\pi$ can be taken as an abbreviation of $?[\pi]\bot$.
FOL can be interpreted in DPL, as follows:

$$
\begin{aligned}
(R t_1 \cdots t_n)^{\bullet} &= ? R t_1 \cdots t_n \\
(t_1 = t_2)^{\bullet} &= ? t_1 = t_2 \\
(\neg\varphi)^{\bullet} &= \sim\varphi^{\bullet} \\
(\varphi_1 \vee \varphi_2)^{\bullet} &= \sim(\sim\varphi_1^{\bullet} \; ; \quad \sim\varphi_2^{\bullet}) \\
(\exists v \varphi)^{\bullet} &= \sim\sim(v := ? \; ; \quad \varphi^{\bullet})
\end{aligned}
$$

DPL and FOL

An inspection of the DPL semantics yields:

LEMMA 20 (Embedding). *For all FOL formulae φ, all models $\mathbf{M}$ for the signature of φ, all assignments g, h in $\mathbf{M}$:*

$$\mathbf{M} \models_g \varphi \text{ and } g = h \quad \text{iff} \quad _g[\![\varphi^{\bullet}]\!]^{\mathbf{M}}_h.$$

DPL programs can be reversed, as follows:

$$
\begin{aligned}
(? R t_1 \cdots t_n)^{\smile} &= ? R t_1 \cdots t_n \\
(? t_1 = t_2)^{\smile} &= ? t_1 = t_2 \\
(v := ?)^{\smile} &= v := ? \\
(\sim\pi)^{\smile} &= \sim\pi \\
(\pi_1 \; ; \; \pi_2)^{\smile} &= \pi_2^{\smile} \; ; \; \pi_1^{\smile}
\end{aligned}
$$

This definition shows that $^{\smile}$ is definable in DPL, because $? R t_1 \cdots t_n$, $? t_1 = t_2$, $\sim\pi$, $v := ?$ and $\sim\pi$ are all symmetric and hence self-converse. What this means is that adding a converse operator to DPL does not increase expressive power. The following reversal lemma is proved by induction on DPL program structure:

LEMMA 21 (Reversal). *For all DPL programs π, all models $\mathbf{M}$, all assignments g, h in $\mathbf{M}$:*

$$_g[\![\pi]\!]^{\mathbf{M}}_h \text{ iff } _h[\![\pi^{\smile}]\!]^{\mathbf{M}}_g.$$

DPL and DPL′

One of the features of DPL is that it does not have the distinction between programs (interpreted as binary relations on a set of appropriate valuations) and formulae (interpreted as predicates on a set of appropriate valuations). Still, it is sometimes useful to be able to make statements about DPL programs. For this, we define DPL′ formulae as follows (π ranges over DPL programs):

$$\varphi \quad ::= \quad \top \mid Rt_1 \cdots t_n \mid t_1 = t_2 \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \exists v\varphi \mid \langle\pi\rangle\varphi.$$

Statements about DPL programs can now be made in DPL′. The formula $\langle\pi\rangle\top$ characterises the assignments where π succeeds. In [Groenendijk and Stokhof, 1991] this is called the satisfaction set of π. The set of possible output assignments for π (the production set of π) is characterised by $\langle\pi^{\smile}\rangle\top$. The following formula expresses that π_1 and π_2 have the same satisfaction and production sets:

(8) $(\langle\pi_1\rangle\top \leftrightarrow \langle\pi_2\rangle\top) \ \wedge \ (\langle\pi_1^{\smile}\rangle\top \leftrightarrow \langle\pi_2^{\smile}\rangle\top).$

Note that it does not follow from (8) that π_1 and π_2 are equivalent. Let π_1 be $?x = x$ and let π_2 be $x :=?$. Then $\langle\pi_1\rangle\top \leftrightarrow \top \leftrightarrow \langle\pi_1\rangle\top$ and $\langle\pi_1^{\smile}\rangle\top \leftrightarrow \top \leftrightarrow \langle\pi_1^{\smile}\rangle\top$, but the two programs are not equivalent. The interpretation of π_1 is the identity relation on the set of assignments, that of π_2 is the set of all pairs g, h such that $g \sim_x h$.

6.2   Proof theory

Reduction to FOL

There are various proof systems for DPL or closely related logics. An early example is the Floyd-Hoare-type system of Van Eijck and De Vries [van Eijck and de Vries, 1992]. Basically, this calculus uses Floyd-Hoare rules to reduce DPL to FOL. We can also use QDL to reduce DPL to FOL. Here is a translation function from DPL′ to FOL:

$$
\begin{aligned}
(\top)^{\circ} &= \top \\
(Rt_1 \cdots t_n)^{\circ} &= Rt_1 \cdots t_n \\
(t_1 = t_2)^{\circ} &= t_1 = t_2 \\
(\neg\varphi)^{\circ} &= \neg\varphi^{\circ} \\
(\varphi_1 \vee \varphi_2)^{\circ} &= \varphi_1^{\circ} \vee \varphi_2^{\circ} \\
(\exists v\varphi)^{\circ} &= \exists v\varphi^{\circ} \\
(\langle ?Rt_1 \cdots t_n\rangle\varphi)^{\circ} &= Rt_1 \cdots t_n \wedge \varphi^{\circ} \\
(\langle ?t_1 = t_2\rangle\varphi)^{\circ} &= t_1 = t_2 \wedge \varphi^{\circ} \\
(\langle v :=?\rangle\varphi)^{\circ} &= \exists v\varphi^{\circ} \\
(\langle \sim\pi\rangle\varphi)^{\circ} &= \neg(\langle\pi\rangle\top)^{\circ} \wedge \varphi^{\circ} \\
(\langle \pi_1 \ ; \ \pi_2\rangle\varphi)^{\circ} &= (\langle\pi_1\rangle\langle\pi_2\rangle\varphi)^{\circ}.
\end{aligned}
$$

Direct inspection of the semantics reveals that this translation is correct, in the following sense:

LEMMA 22 (Translation Correctness). *For all DPL′ formulae φ, all FO models $\mathbf{M}$ for the signature of φ, all variable assignments g in $\mathbf{M}$:*

$$\mathbf{M} \models_g \varphi \text{ iff } \mathbf{M} \models_g (\varphi)^{\circ}.$$

It follows from this that the following reduction axioms for DPL are sound:

test relation	$\langle ?Rt_1 \cdots t_n \rangle \varphi \leftrightarrow Rt_1 \cdots t_n \wedge \varphi$
test equality	$\langle ?t_1 = t_2 \rangle \varphi \leftrightarrow t_1 = t_2 \wedge \varphi$
random assignment	$\langle v := ? \rangle \varphi \leftrightarrow \exists v \varphi$
dynamic negation	$\langle \sim \pi \rangle \varphi \leftrightarrow \neg \langle \pi \rangle \top \wedge \varphi$
sequence	$\langle \pi_1 \ ; \ \pi_2 \rangle \varphi \leftrightarrow \langle \pi_1 \rangle \langle \pi_2 \rangle \varphi.$

The boxed counterparts of these axioms can be derived by propositional reasoning:

test relation	$[?Rt_1 \cdots t_n]\varphi \leftrightarrow (Rt_1 \cdots t_n \rightarrow \varphi)$
test equality	$[?t_1 = t_2]\varphi \leftrightarrow (t_1 = t_2 \rightarrow \varphi)$
random assignment	$[v := ?]\varphi \leftrightarrow \forall v \varphi$
dynamic negation	$[\sim \pi]\varphi \leftrightarrow ([\pi]\bot \rightarrow \varphi)$
sequence	$[\pi_1 \ ; \ \pi_2]\varphi \leftrightarrow [\pi_1][\pi_2]\varphi.$

The calculus for DPL′ can now consist of the axioms for FOL, the axioms for test relation, test equality, dynamic negation and sequence (either in their box or in their diamond versions), and the inference rules of FOL: modus ponens and generalisation. It follows from the translation lemma that this axiomatisation is sound. The axiomatisation is also complete.

THEOREM 23 (DPL′ completeness).
For all DPL′ formulae φ: if $\models \varphi$ then $\vdash \varphi$.

Proof. The proof system for FOL is complete, and every DPL′ formula φ is provably equivalent to some FOL formula. ∎

By way of example of the application of the calculus we give the derivation of the FOL counterpart to the DPL rendering of so-called 'donkey sentences' (cf., section 7.2 below for more extensive discussion of this type of phenomenon):

1. If a farmer owns a donkey then he beats it.

DPL translates this using a defined operator for dynamic implication, given by:

$$\varphi \Rightarrow \psi \ :\equiv \ \sim(\varphi \ ; \ \sim\psi).$$

The DPL rendering of (1) looks like this:

$$(x := ? \ ; \ ?Fx \ ; \ y := ? \ ; \ ?Dy \ ; \ ?Oxy) \Rightarrow ?Bxy.$$

Here is the reduction to FOL using the reduction axioms:

$$\langle \sim(x :=? \; ; \; ?Fx \; ; \; y :=? \; ; \; ?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy)\rangle\top$$
$$\leftrightarrow \quad [x :=? \; ; \; ?Fx \; ; \; y :=? \; ; \; ?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy]\bot$$
$$\leftrightarrow \quad \forall x[?Fx \; ; \; y :=? \; ; \; ?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy]\bot$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow [y :=? \; ; \; ?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy]\bot)$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y[?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy]\bot)$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y([?Dy \; ; \; ?Oxy \; ; \; \sim?Bxy]\bot))$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y(Dy \rightarrow [?Oxy \; ; \; \sim?Bxy]\bot))$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y(Dy \rightarrow (Oxy \rightarrow [\sim?Bxy]\bot)))$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y(Dy \rightarrow (Oxy \rightarrow ([?Bxy]\bot \rightarrow \bot))))$$
$$\leftrightarrow \quad \forall x(Fx \rightarrow \forall y(Dy \rightarrow (Oxy \rightarrow Bxy))).$$

Clearly, this is the desired universal reading of the example.

Axiomatisation

Axiomatising DPL becomes more of a challenge if one is after an axiomatisation at the level of programs, without recourse to a static assertion language like FOL. Such a direct axiomatisation is provided in Van Eijck [van Eijck, 1999]. Key element of the calculus is an appropriate treatment of substitution in DPL.

For readability, it is useful to slightly rephrase the DPL language, by leaving out the spurious test operators and by using quantifier notation for random assignment:

DEFINITION 24 (DPL syntax again).

$$\pi \quad ::= \quad \top \mid Rt_1 \cdots t_2 \mid t_1 = t_2 \mid \exists v \mid \sim\pi \mid \pi_1 \; ; \; \pi_2.$$

Types of Variable Occurrences   Let V be the variables of the DPL language. The set of variables which have a fixed occurrence in a DPL program π is given by a function $free : \text{DPL} \rightarrow \mathcal{P}V$, the set of variables which are introduced in a formula is given by a function $intro : \text{DPL} \rightarrow \mathcal{P}V$, and the set of variables which have a classically bound occurrence in a formula is given by a function $cbnd : \text{DPL} \rightarrow \mathcal{P}V$.

The introduced variables of π (called 'blocked' variables in [Visser, 1998]) are the variables y such that π contains an $\exists y$ not in the scope of a negation. The free variables of π are the variables on which input valuations have to agree on output valuations. The classically bound variables of π are the variables that behave like the bound variables of FOL. Let $var(Pt_1 \cdots t_n)$ be the set of all variables among $t_1 \cdots t_n$.

DEFINITION 25 (free, intro, cbnd).

- $free(\top) := \emptyset, intro(\top) := \emptyset, cbnd(\top) := \emptyset.$

- $free(\exists v \ ; \ \pi) := free(\pi) - \{v\}$,
  $intro(\exists v \ ; \ \pi) := \{v\} \cup intro(\pi)$,
  $cbnd(\exists v \ ; \ \pi) := cbnd(\pi)$.

- $free(Pt_1 \cdots t_n \ ; \ \pi) := var(Pt_1 \cdots t_n) \cup free(\pi)$,
  $intro(Pt_1 \cdots t_n \ ; \ \pi) := intro(\pi)$,
  $cbnd(Pt_1 \cdots t_n \ ; \ \pi) := cbnd(\pi)$.

- $free(\sim(\pi_1) \ ; \ \pi_2) := free(\pi_1) \cup free(\pi_2)$,
  $intro(\sim(\pi_1) \ ; \ \pi_2) := intro(\pi_2)$,
  $cbnd(\sim(\pi_1) \ ; \ \pi_2) := intro(\pi_1) \cup cbnd(\pi_1) \cup cbnd(\pi_2)$.

- $free((\pi_1\pi_2) \ ; \ \pi_3) := free(\pi_1 \ ; \ (\pi_2 \ ; \ \pi_3))$,
  $intro((\pi_1\pi_2) \ ; \ \pi_3) := intro(\pi_1 \ ; \ (\pi_2 \ ; \ \pi_3))$,
  $cbnd((\pi_1\pi_2) \ ; \ \pi_3) := cbnd(\pi_1 \ ; \ (\pi_2 \ ; \ \pi_3))$.

Some examples may clarify this definition. Let

$$\pi := \exists v \ ; \ \exists w \ ; \ Ruvw.$$

Then $intro(\pi) = \{v, w\}$, $free(\pi) = \{u\}$, $cbnd(\pi) = \emptyset$. The occurrence of u in $Ruvw$ is free.

Variables introduced within the scope of negation become classically bound. Let

$$\pi := \ \sim(\exists v \ ; \ \exists w \ ; \ Ruvw).$$

Then $intro(\pi) = \emptyset$, $free(\pi) = \{u\}$, $cbnd(\pi) = \{v, w\}$. The occurrence of u in $Ruvw$ is still free.

A variable can have fixed, bound and introduced occurrences in an expression. Let

$$\pi := Px \ ; \ \exists x \ ; \ \sim Px \ ; \ \sim(\exists x \ ; \ Qx).$$

Then $intro(\pi) = \{x\}$, $free(\pi) = \{x\}$, $cbnd(\pi) = \{x\}$. The leftmost occurrence of x is free, the other occurrences are not.

Binding   Note that for all DPL programs π, $intro(\pi) \cap free(\pi) = \emptyset$. Let $g \sim_X h$ if variable assignments g and h differ at most in the values of variables among X. Let $g[X]h$ if $g \sim_{V-X} h$, where V is the set of all variables. Thus, $g[X]h$ expresses that g and h agree on the values of variables in X.

LEMMA 26 (DPL binding). *If $_g[\![\varphi]\!]_h^{\mathbf{M}}$ then $g \sim_{intro(\varphi)} h$ and $g[free(\varphi)]h$.*

Thus, the leftmost occurrence of x in $Px \ ; \ \exists x \ ; \ \sim Px \ ; \ \sim(\exists x \ ; \ Qx)$ is free, the other occurrences are not. Use π_t^v for the result of substituting t for all free occurrences of v in π:

DEFINITION 27 (π_t^v).

$$
\begin{array}{lcl}
\top_t^v & := & \top \\
(Rt_1 \cdots t_n \ ; \ \pi)_t^v & := & R{t_1}_t^v \cdots {t_n}_t^v \ ; \ \pi_t^v \\
(t_1 = t_2 \ ; \ \pi)_t^v & := & {t_1}_t^v = {t_n}_t^v \ ; \ \pi_t^v \\
(\exists v \ ; \ \pi)_t^v & := & \exists v \ ; \ \pi \\
(\exists w \ ; \ \pi)_t^v & := & \exists w \ ; \ \pi_t^v \\
(\sim(\pi_1) \ ; \ \pi_2)_t^v & := & \sim({\pi_1}_t^v) \ ; \ {\pi_2}_t^v \\
((\pi_1 \ ; \ \pi_2) \ ; \ \pi_3)_t^v & := & (\pi_1 \ ; \ (\pi_2 \ ; \ \pi_3))_t^v
\end{array}
$$

Note that this definition of substitution takes the dynamic binding force of $\exists v$ over the text that follows into account (cf. the clause for $(\exists v \ ; \ \pi)_t^v$, where the occurrence of $\exists v$ blocks off the π that follows). Visser [Visser, 1998] calls this substitution 'left' substitution.

Sequent Deduction Rules Figure 2 gives a set of sequent deduction rules for DPL, using the format $\varphi \implies \psi$, where $\implies$ is the sequent separator. Note that $\varphi \implies \bot$ expresses that φ is inconsistent. The calculus defines a relation $\implies \subseteq \text{DPL}^2$ by means of: $\varphi \implies \psi$ iff $\varphi \implies \psi$ is at the root of a finite tree with sequents at its nodes, such that the sequents at a leaf node are axioms of the calculus, and the sequents at the internal nodes follow by means of a rule of the calculus from the sequent(s) at the daughter node(s) of that internal node.

In the calculus, C, with and without subscripts, is used as a variable over contexts, where a context is a formula or the empty list ϵ. Substitution and evaluation are extended to contexts in the obvious way. If C is a context and φ a formula, then we use $C\varphi$ for the formula given by: $C\varphi := \varphi$ if $C = \epsilon$, $C\varphi := \psi; \varphi$ if $C = \psi$. Similarly for φC, and for $C_1 \varphi C_2$.

It is convenient to extend the definition of substitution to sequents.

DEFINITION 28 ($(C \implies \varphi)_t^v$). Induction on the structure of C:

$$
(\epsilon \implies \varphi)_t^v \quad := \quad \epsilon \implies \varphi_t^v
$$

$$
(\psi \implies \varphi)_t^v \quad := \quad \left\{ \begin{array}{ll} \psi_t^v \implies \varphi & \text{if } v \in intro(\psi) \\ \psi_t^v \implies \varphi_t^v & \text{otherwise.} \end{array} \right.
$$

Substitution for sequents carries in its wake a notion of being free for a variable in a sequent:

DEFINITION 29 (t is free for v in $C \implies \psi$).

1. t is free for v in $\epsilon \implies \psi$ if t is free for v in ψ.

2. t is free for v in $\varphi \implies \psi$ if t is free for v in φ, and either $v \in intro(\varphi)$ or t is free for v in ψ.

Figure 2. The Calculus for DPL

test axiom	$$\overline{T \Longrightarrow T}$$
transitivity	$$\frac{\varphi \Longrightarrow \psi \qquad \psi \Longrightarrow \chi}{\varphi \Longrightarrow \chi} \; \mathit{intro}(\psi) \cap \mathit{free}(\chi) = \emptyset$$
test swap	$$\frac{C_1 T_1 \;\; ; \;\; T_2 C_2 \Longrightarrow \varphi}{C_1 T_2 \;\; ; \;\; T_1 C_2 \Longrightarrow \varphi}$$
quantifier move	$$\frac{C_1 T \;\; ; \;\; \exists v \; C_2 \Longrightarrow \varphi}{C_1 \exists v \;\; ; \;\; T C_2 \Longrightarrow \varphi} \; v \notin \mathit{free}(T)$$ $$\frac{C_1 \exists v \;\; ; \;\; T C_2 \Longrightarrow \varphi}{C_1 T \;\; ; \;\; \exists v \; C_2 \Longrightarrow \varphi} \; v \notin \mathit{free}(T)$$
quantifier intro	$$\frac{\varphi \Longrightarrow \psi_t^v}{\varphi \Longrightarrow \exists v \;\; ; \;\; \psi} \; t \text{ free for } v \text{ in } \psi$$
var refreshment	$$\frac{C_1 \exists v \; C_2 \Longrightarrow \varphi}{C_1 \exists w (C_2 \Longrightarrow \varphi)_w^v} \; w \notin \mathit{intro}(C_1) \cup \mathit{free}(C_1)$$
sequencing	$$\frac{\psi \Longrightarrow \chi}{\varphi \;\; ; \;\; \psi \Longrightarrow \chi}$$ $$\frac{\varphi \Longrightarrow \psi \qquad \varphi \Longrightarrow \chi}{\varphi \Longrightarrow \psi \;\; ; \;\; \chi} \; \mathit{intro}(\psi) \cap \mathit{free}(\chi) = \emptyset$$
negation	$$\frac{\varphi \Longrightarrow \psi}{\varphi \;\; ; \;\; \sim\!\psi \Longrightarrow \bot} \qquad \frac{\varphi \;\; ; \;\; \psi \Longrightarrow \bot}{\varphi \Longrightarrow \sim\!\psi}$$
double negation	$$\frac{\varphi \Longrightarrow \sim\!\sim\!\psi}{\varphi \Longrightarrow \psi} \qquad \frac{\varphi \;\; ; \;\; \sim\!\sim\!\psi \Longrightarrow \bot}{\varphi \;\; ; \;\; \psi \Longrightarrow \bot}$$

When a rule mentions a substitution φ^v_t in the consequent of a sequent then the standard assumption is made that t is free for v in φ. When a rule mentions a substitution $C_1(C_2 \Longrightarrow \varphi)^v_t$ then it is assumed that t is free for v in $C_2 \Longrightarrow \varphi$.

In the rules of Figure 2 T is used as an abbreviation of formulae φ with $intro(\varphi) = \emptyset$ (T for *Test* formula).

Here is an example application of the quantifier intro rule.

$$\frac{Rxx \Longrightarrow Rxx}{Rxx \Longrightarrow \exists y \; ; \; Rxy}$$

Rxx equals $(Rxy)^y_x$, so this is indeed a correct application of the rule.

Variable refreshment allows the liberation of a captured variable, e.g., of the first two occurrences of x in $\exists x \; ; \; Px \; ; \; \exists x \; ; \; Qx$, by means of replacement by a variable that does not occur as an introduced or free variable in the left context in the given sequent:

$$\frac{\exists x \; ; \; Px \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx}{\exists y \; ; \; Py \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx}$$

It is also possible to change the other occurrences of x in the same example. The following is also a correct application of the rule:

$$\frac{\exists x \; ; \; Px \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx}{\exists x \; ; \; Px \; ; \; \exists y \; ; \; Qy \Longrightarrow Qy}$$

Note that the rule can also be used to recycle a variable:

$$\frac{\exists y \; ; \; Py \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx}{\exists x \; ; \; Px \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx}$$

This application is also correct, for

$$(\exists x \; ; \; Px \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx) = (\exists x \; ; \; (Py \; ; \; \exists x \; ; \; Qx \Longrightarrow Qx)^y_x).$$

An example application of the rule for $;$ right is:

$$\frac{Rxx \Longrightarrow \exists y \; ; \; Ryx \qquad Rxx \Longrightarrow \exists z \; ; \; Rxz}{Rxx \Longrightarrow \exists y \; ; \; Ryx \; ; \; \exists z \; ; \; Rxz} \quad ; \;\; \text{right}$$

In case the condition on the rule for $;$ right is not satisfied, e.g. for the two sequents $\sim Px \; ; \; \exists x \; ; \; Px \Longrightarrow \exists x \; ; \; \sim Px$ and $\sim Px \; ; \; \exists x \; ; \; Px \Longrightarrow Px$, this can always be remedied by one or more applications of $\exists$ Right to the second premise.

It is not hard to see that the rules of the calculus are sound. The calculus is also complete. For the proof — a modification of the standard Henkin style completeness proof for classical first order logic — we refer to [van Eijck, 1999].

6.3   Computational DPL

In [Apt and Bezem, 1999] a computational interpretation of standard first order logic is proposed, with as key ingredient a new interpretation of identity statements (in suitable contexts) as assignment actions. Computation states are partial maps of variables to values. The gist of the proposal is this: in a state α that is defined for a term t but undefined for a variable v, an identity statement $v = t$ or $t = v$ is interpreted as an instruction to assign the value t^α to the variable v.

Let $\mathcal{M} = (M, I)$ be a FO model, and let V be a set of variables. Let $\mathbf{A} := \{\alpha \in M^X \mid X \subseteq V\}$. If $\alpha \in M^X$, then call X the domain of α; a term t is α-closed if all variables in t are in X, an atom $Pt_1 \cdots t_n$ is α-closed if all t_i are α-closed, and an identity $t_1 = t_2$ is α-closed if both of t_1, t_2 are. Use $\uparrow$ for 'undefined' and $\downarrow$ for 'defined'. Term interpretation in model $\mathbf{M} = (M, I)$ with respect to valuation α now has to take the possibility into account that the value of the term under α is undefined.

$$
\begin{aligned}
v^\alpha &:= \begin{cases} \alpha(v) & \text{if } v \text{ is } \alpha\text{-closed} \\ \uparrow & \text{otherwise} \end{cases} \\
(ft_1 \cdots t_n)^\alpha &:= \begin{cases} I(f)t_1^\alpha \cdots t_n^\alpha & \text{if } t_1, \ldots, t_n \ \alpha\text{-closed} \\ \uparrow & \text{otherwise} \end{cases}
\end{aligned}
$$

An identity $t_1 = t_2$ is an α-assignment if either $t_1 \equiv v$, $t_1^\alpha =\uparrow$, $t_2^\alpha =\downarrow$, or $t_2 \equiv v$, $t_1^\alpha =\downarrow$, $t_2^\alpha =\uparrow$. An α-assignment can be used as a statement that extends a valuation α with a new value.

A first order predicate with its arguments $Pt_1 \cdots t_n$ is interpreted as a test that can either fail or succeed, provided that all of the t_i are defined for the input state α; otherwise an error is generated. The empty conjunction is interpreted as the instruction to succeed in any state α, with output α.

This is then extended to finite conjunctions of implications, negations, disjunctions and existential quantifications, according to the following rule set: $[\![\varphi]\!]_\alpha$ denotes the computation tree for φ on input α. A tree is successful if it contains at least one leaf consisting of just a variable map, it fails if all its leafs equal *fail*.

$$
\begin{array}{c}
\exists v\varphi \wedge \psi, \alpha \\
\mid \\
[\![\varphi \wedge \psi]\!]_\alpha
\end{array}
\qquad \text{if } v \notin \text{dom}(a), \ v \text{ not free in } \psi.
$$

$$
\begin{array}{c}
\neg\varphi \wedge \psi, \alpha \\
\mid \\
[\![\psi]\!]_\alpha
\end{array}
\qquad \text{if } \varphi \ \alpha\text{-closed}, \ [\![\varphi]\!]_\alpha \text{ failed.}
$$

$$
\begin{array}{c}
\neg\varphi \wedge \psi, \alpha \\
\mid \\
\textit{fail}
\end{array}
\qquad \text{if } \varphi \ \alpha\text{-closed}, \ [\![\varphi]\!]_\alpha \text{ successful.}
$$

$$(\varphi_1 \to \varphi_2) \wedge \psi, \alpha$$
$$\llbracket \psi \rrbracket_\alpha \qquad \text{if } \varphi_1 \text{ } \alpha\text{-closed, } \llbracket \varphi_1 \rrbracket_\alpha \text{ failed.}$$

$$(\varphi_1 \to \varphi_2) \wedge \psi, \alpha$$
$$\llbracket \varphi_2 \wedge \psi \rrbracket_\alpha \qquad \text{if } \varphi_1 \text{ } \alpha\text{-closed,} \llbracket \varphi_1 \rrbracket_\alpha \text{ successful.}$$

$$(\varphi_1 \vee \varphi_2) \wedge \psi, \alpha$$
$$\llbracket \varphi_1 \wedge \psi \rrbracket_\alpha \qquad \llbracket \varphi_2 \wedge \psi \rrbracket_\alpha$$

All cases not listed generate an error.

This computation procedure has the property that for any φ and any input valuation α, the valuations at success nodes in $\llbracket \varphi \rrbracket_\alpha$ are extensions of α. Computations never *change* the input valuations. In particular, $\exists x \varphi \wedge \psi$ is treated as equivalent with $\varphi \wedge \psi$ provided the variable conditions hold. Thus, the quantifier has no computational effect, but acts as a prohibition sign: its only function is to rule out occurrences of x in the outside context of $\exists x \varphi$.

The computational engine can be adapted to a setting where quantifiers are read dynamically, by giving assignments $v :=?$ an appropriate computational meaning. The relational interpretation of $v :=?$ is computationally infeasible, for the instruction to replace the value of register v by an arbitrary new value is awkward if one is computing over an infinite domain, say the domain of natural numbers. As a statement on $\mathbb{N}$, $v :=?$ is an instruction to pick an arbitrary natural number and assign it to v. Since this can be done in an infinite number of ways, this does not represent any finite computational procedure.

In the computational interpretation of DPL one therefore changes the quantifier action as follows. Instead of letting the quantifier action $v :=?$ perform its full duty, the action $v :=?$ is split into two tasks:

1. throwing away the old value of v, and

2. identifying appropriate new values for v.

On infinite domains any attempt to perform task (2) immediately will cause an infinite branching transition, and therefore this task is *postponed*. The duty of finding an appropriate new value for v is relegated to an appropriate *identifying statement* for v further on. This move is inspired by the computational interpretation of identity statements from [Apt and Bezem, 1999]. See [van Eijck *et al.*, 2001] and [Heguiabehere, 2001] for more information on computing with DPL.

6.4 Extensions of DPL

DPL can be viewed as the most basic of a hierarchy of formulae-as-programs languages. We will now look at extensions of DPL with the six operations $\cup$, $\breve{}$, σ, $\breve{\sigma}$, $\cap$, $\exists\!\!\exists$. Extensions of DPL with $\cap$ (relation intersection) and $\exists\!\!\exists$ (local variable declaration) are studied in [Visser, 1998], while in [van Eijck *et al.*, 2001], an extension of DPL with $\cup$ (relation union) and σ (simultaneous substitution) is axiomatised, and ω-completeness is proved for the extension of DPL with $\cup$, σ and Kleene star.

Extended Semantics

A substitution is a finite set of bindings $x \mapsto t$, with the usual conditions that no binding is trivial (of the form $x \mapsto x$) and that every x in the set has at most one binding (substitutions are functional). Examples of substitutions are $\{x \mapsto f(x)\}$ ("set new x equal to f-value of old x"), $\{x \mapsto y, y \mapsto x\}$ ("swap values of x and y"). If a substitution contains just a single binding we omit the curly brackets and write just the *assignment statement* $x := t$. Note that if x occurs in t, the assignment $x := t$ is not expressible in DPL. Similarly , there is no DPL program that is equivalent to $\{x \mapsto y, y \mapsto x\}$.

Left-to-right substitutions σ have right-to-left counterparts $\breve{\sigma}$ (converse substitutions). For pre- and postcondition reasoning with extensions of DPL, converse substitution and relation converse $\breve{}$ are attractive.

A converse substitution is a finite set of converse bindings $(x \mapsto t)^{\breve{}}$, with the same conditions as those for substitutions. An example is $(x \mapsto f(x))^{\breve{}}$ ("set old x equal to f-value of new x", i.e. "look at all inputs g that differ from the output h only in x, and that satisfy $f(g(x)) = h(x)$").

The semantics definition for the new operators runs:

$$
\begin{aligned}
\llbracket \sigma \rrbracket^{\mathbf{M}} &= \{(g, g^{x_1 \cdots x_n}_{d_1 \cdots d_n}) \mid \{x_1, \ldots, x_n\} = \mathrm{dom}(\sigma) \text{ and } d_i = \sigma(x_i)^{\mathbf{M},g}\} \\
\llbracket \breve{\sigma} \rrbracket^{\mathbf{M}} &= \{(g^{x_1 \cdots x_n}_{d_1 \cdots d_n}, g) \mid \{x_1, \ldots, x_n\} = \mathrm{dom}(\sigma) \text{ and } d_i = \sigma(x_i)^{\mathbf{M},g}\} \\
\llbracket \exists\!\!\exists x(\pi) \rrbracket^{\mathbf{M}} &= \{(g, k^x_{g(x)}) \mid \text{for some } d : (g^x_d, k) \in \llbracket \pi \rrbracket^{\mathbf{M}}\} \\
\llbracket \pi_1 \cap \pi_2 \rrbracket^{\mathbf{M}} &= \llbracket \pi_1 \rrbracket^{\mathbf{M}} \cap \llbracket \pi_2 \rrbracket^{\mathbf{M}} \\
\llbracket \pi^{\breve{}} \rrbracket^{\mathbf{M}} &= \{(g, h) \mid (h, g) \in \llbracket \pi \rrbracket^{\mathbf{M}}\}
\end{aligned}
$$

The $\exists\!\!\exists$ operator allows for the declaration of local variables. Simultaneous substitution permits performing certain computations without the use of auxiliary variables. Converse and converse simultaneous substitution are useful for pre- and postcondition reasoning, as they allow us to define the inverses of programs under certain conditions [Gries, 1981, Chapter 21].

Left-to-Right and Right-to-Left Substitution

Because the semantics of DPL programs is completely symmetric, performing a substitution in a DPL program can be done in two directions: left-to-right and

right-to-left [Visser, 1998] (see also [Vermeulen, 2001], where substitutions for DPL with a stack semantics are studied). Left-to-right substitutions affect the left-free variable occurrences, right-to-left substitutions the right-free (or 'actively bound') variable occurrences.

DPL has two directional analogues to the substitution lemma from FOL: one for left-to-right substitution and one for right-to-left substitution. For left-to-right substitution we get that $g[\![\sigma(\pi)]\!]^{\mathbf{M}}h$ iff $g_\sigma[\![\pi]\!]^{\mathbf{M}}h$. Viewing the substitution itself as a state change, we can decompose this into $g[\![\sigma]\!]^{\mathbf{M}}g'[\![\pi]\!]^{\mathbf{M}}h$. This uses $g[\![\sigma]\!]^{\mathbf{M}}k$ iff $k = g_\sigma$.

The right-to-left substitution lemma for DPL says that $g[\![\breve{\sigma}(\pi)]\!]^{\mathbf{M}}h$ iff $g[\![\pi]\!]^{\mathbf{M}}h_\sigma$. Viewing the right-to-left substitution itself as a state change, we can decompose this into $g[\![\pi]\!]^{\mathbf{M}}h'[\![\breve{\sigma}]\!]^{\mathbf{M}}h$. This uses $k[\![\breve{\sigma}]\!]^{\mathbf{M}}h$ iff $k = h_\sigma$. Again, since in general $\breve{\sigma}$ is not expressible in DPL, we have a motivation to extend the language with converse substitutions.

Use $\circ$ for relational composition of substitution expressions, defined as follows:

DEFINITION 30 (Composition of substitutions). Let

$$\sigma = \{v_1 \mapsto t_1, \ldots, v_n \mapsto t_n\} \text{ and } \rho = \{w_1 \mapsto r_1, \ldots, w_m \mapsto r_m\}$$

be substitutions. Then $\sigma \circ \rho$ is the result of removing from the set

$$\{w_1 \mapsto \sigma(r_1), \ldots, w_m \mapsto \sigma(r_m), v_1 \mapsto t_1, \ldots, v_n \mapsto t_n\}$$

the bindings $w_1 \mapsto \sigma(r_i)$ for which $\sigma(r_i) = w_i$, and the bindings $v_j \mapsto t_j$ for which $v_j \in \{w_1, \ldots, w_m\}$.

It is easily proved now that $\sigma \; ; \; \rho$ is equivalent to $\sigma \circ \rho$. E.g., $x := x+1 \; ; \; y := x$ is equivalent to $\{x \mapsto x+1, y \mapsto x+1\}$, and $x := y \; ; \; x := x+1$ is equivalent to $x := y+1$.

Every DPL$(\cup, \sigma)$ formula can be written with $;$ associating to the right, as a list of predicates, quantifications, negations, choices and substitutions, with a substitution ρ at the end (possibly the empty substitution). Left-to-right substitution in DPL$(\cup, \sigma)$ is defined by:

$$
\begin{aligned}
\sigma(\rho) &:= \sigma \circ \rho \\
\sigma(\rho \; ; \; \pi) &:= \sigma \circ \rho \; ; \; \pi \\
\sigma(\exists v \; ; \; \pi) &:= \exists v \; ; \; \sigma'\pi \text{ where } \sigma' = \sigma \backslash \{v \mapsto t \mid t \in T\} \\
\sigma(P\bar{t} \; ; \; \pi) &:= P\sigma\bar{t}; \sigma\pi \\
\sigma(t_1 = t_2 \; ; \; \pi) &:= \sigma t_1 = \sigma t_2 \; ; \; \sigma\pi \\
\sigma(\sim(\pi_1) \; ; \; \pi_2) &:= \neg(\sigma\pi_1) \; ; \; \sigma\pi_2 \\
\sigma((\pi_1 \cup \pi_2); \pi_3) &:= \sigma(\pi_1; \pi_3) \cup \sigma(\pi_2; \pi_3)
\end{aligned}
$$

A term t is left-to-right free for v in π if all variables in t are input-constrained in all positions of the left-free occurrences of v in π. A substitution σ is safe for π if all bindings $v \mapsto t$ of σ are such that t is left-to-right free for v in π. This allows us to prove:

LEMMA 31 (Left-to-Right Substitution). *If σ is safe for π then $g[\![\sigma(\pi)]\!]h$ iff $g_\sigma[\![\pi]\!]h$.*

Right-to-left substitution is defined in a symmetric fashion, now reading the formulae in a left-associative manner, with a converse substitution at the front, and overloading the notation by also using $\circ$ for the relational composition of converse substitutions (defined as one would expect, to get $\breve\sigma \circ \breve\rho = (\rho \circ \sigma)^\vee$):

$$
\begin{aligned}
\breve\sigma(\breve\rho) &:= \breve\sigma \circ \breve\rho \\
\breve\sigma(\pi \; ; \; \breve\rho) &:= \pi \; ; \; \breve\sigma \circ \breve\rho \\
\breve\sigma(\pi; \exists v) &:= \breve\sigma'\pi \; ; \; \exists v \text{ where } \breve\sigma' = \breve\sigma \backslash \{(v \mapsto t)^\vee \mid t \in T\} \\
\breve\sigma(\pi \; ; \; P\bar t) &:= \breve\sigma\pi \; ; \; P\sigma\bar t \\
\breve\sigma(\pi \; ; \; t_1 = t_2) &:= \breve\sigma\pi \; ; \; \sigma t_1 = \sigma t_2 \\
\breve\sigma(\pi_1 \; ; \; \sim(\pi_2)) &:= \breve\sigma\pi_1 \; ; \; \sim(\breve\sigma\pi_2) \\
\breve\sigma(\pi_1 \; ; \; (\pi_2 \cup \pi_3)) &:= \breve\sigma(\pi_1; \pi_2) \cup \breve\sigma(\pi_1; \pi_3)
\end{aligned}
$$

A term t is right-to-left free for v in π if all variables in t are output-constrained in all positions of the right-free (actively bound) occurrences of v in π. A converse substitution $\breve\sigma$ is safe for π if all converse bindings $(v \mapsto t)^\vee$ of $\breve\sigma$ are such that t is right-to-left free for v in π. This allows us to prove:

LEMMA 32 (Right-to-Left Substitution). *If $\breve\sigma$ is safe for π then $g[\![\breve\sigma(\pi)]\!]h$ iff $g[\![\pi]\!]h_\sigma$.*

Expressive Power

The following results are from [ten Cate *et al.*, 2001];indexHeguiabehere, J. many of the proofs are adapted from proofs given in [Visser, 1998].

THEOREM 33. *$DPL(\exists)$ is equally expressive as $DPL(\cup, \cap, \breve{}, \sigma, \breve\sigma, \exists)$.*

Proof. Let a formula π be given, and let V be the set of variables occurring in π. Furthermore, let V' and V'' be sets of variables, such that V, V' and V'' are mutually disjoint and of equal cardinality. Let $V = \{x_1, \ldots, x_n\}, V' = \{x'_1, \ldots, x'_n\}$, and $V'' = \{x''_1, \ldots x''_n\}$. The following function C translates a formula from $DPL(\cup, \cap, \breve{}, \sigma, \breve\sigma, \exists)$ into a test from DPL.

$$\begin{aligned}
C(\exists y) \quad &= \bigwedge_{x \in V \setminus \{y\}} x' = x \\
C(Rt_1 \ldots t_n) &= \bigwedge_{x \in V} x' = x \;\; ; \quad Rt_1 \ldots t_n \\
C(t_1 = t_2) \quad &= \bigwedge_{x \in V} x' = x \;\; ; \quad t_1 = t_2 \\
C(\sim\pi) \quad &= \bigwedge_{x \in V} x' = x \;\; ; \;\; \sim(\exists x_1'; \ldots; \exists x_n' \;\; ; \;\; C(\pi)) \\
C(\pi_1; \pi_2) \quad &= \sim\sim(\exists x_1''; \ldots; \exists x_n''; C(\pi_1)^{[x_1'/x_1'', \ldots, x_n'/x_n'']} \;\; ; \quad C(\pi_2)^{[x_1/x_1'', \ldots, x_n/x_n'']}) \\
C(\pi_1 \cap \pi_2) \quad &= C(\pi_1); C(\pi_2) \\
C(\pi_1 \cup \pi_2) \quad &= \sim(\sim C(\pi_1); \sim C(\pi_2)) \\
C(\pi^{\smile}) \quad &= C(\pi)^{[x_1/x_1', \ldots, x_n/x_n', x_1'/x_1, \ldots, x_n'/x_n]} \\
C(\sigma) \quad &= \bigwedge_{x \in dom(\sigma)} x' = \sigma(x); \bigwedge_{x \in V \setminus dom(\sigma)} x' = x \\
C(\breve{\sigma}) \quad &= \bigwedge_{x \in dom(\sigma)} x = \sigma(x)^{[x_1/x_1', \ldots, x_n/x_n']}; \bigwedge_{x \in V \setminus dom(\sigma)} x' = x \\
C(\exists x.\pi) \quad &= \sim\sim(\exists x; \exists x'; C(\pi)); x' = x
\end{aligned}$$

Here, $\bigwedge$ is used as a shorthand for a long composition, which is non-ambiguous because the order of the particular sentences involved doesn't matter. By induction, it can be shown that every π containing only variables in V, is equivalent to $\exists x_1' \ldots x_n'(C(\pi); x_1 := x_1'; \ldots; x_n := x_n')$. ∎

THEOREM 34. *$DPL(^*, \exists)$ is equally expressive as $DPL(^*, \cup, \cap, \smile, \sigma, \breve{\sigma}, \exists)$*

Proof. As the proof of Theorem 33, now adding the following clause to the definition of C.

$$C(\pi^*) = \neg\neg(\exists x_1'' \;\; ; \;\; \ldots \;\; ; \;\; \exists x_n''; (C(\pi)^{[x_i'/x_i'']} \;\; ; \;\; \bigwedge_{x \in V} x := x'')^* \;\; ; \;\; \bigwedge_{x \in V} x = x')$$

∎

It follows immediately that every formula $\pi \in DPL(\cup, \cap, \smile, \sigma, \breve{\sigma}, \exists)$ is equivalent to a first order logic formula, in the sense that π can be executed in $\mathcal{M}$ with input assignment g iff the first order translation of π is true in $\mathcal{M}$ under g.

THEOREM 35 (Visser). *$DPL(\exists)$ can be embedded into $DPL(\cap)$.*

Proof. Let π be of the form $\exists x(\psi)$, and let $\{y_1, \ldots, y_n\} = I(\pi) \setminus \{x\}$, where $I(\pi)$ are the introduced variables of π, i.e. the variables in $intro(\pi)$, i.e. the variables y such that π contains an $\exists y$ not in the scope of a negation. Then π is equivalent to $(\exists x; \psi; \exists x) \cap (\exists y_1; \ldots; \exists y_n)$ ∎

In a similar way, the following can be proved:

THEOREM 36. *$DPL(^*, \exists)$ can be embedded into $DPL(^*, \cap)$.*

It is also easy to show that * gets us beyond first order expressive power:

THEOREM 37. *The formula*

$$\neg(\exists y \;\; ; \;\; y = 0 \;\; ; \;\; (\exists z \;\; ; \;\; z = f(y) \;\; ; \;\; \exists y \;\; ; \;\; y = f(z))^* \;\; ; \;\; x = y)$$

cannot be expressed in $DPL(\cup, \cap, \smile, \sigma, \breve{\sigma}, \exists)$.

Proof. On the natural numbers (interpreting f as the successor relation), this formula defines the odd numbers. Oddness on the natural numbers cannot be captured in a first order formula with only successor. ∎

DEFINITION 38. A substitution $\{x_1 \mapsto t_1, \ldots, x_n \mapsto t_n\}$ is *full* if every x_i occurs in some t_j and every t_i contains some x_j.

Examples of full substitutions are $x := f(x)$ and $\{x \mapsto y, y \mapsto x\}$, while the substitution $x := y$ is not full. It is easy to see that full substitutions are closed under composition. Note that a substitution without function symbols is full iff it is a renaming. Also, note that any formula of $\mathrm{DPL}(\sigma)$ or any of its extensions can be transformed into a formula in the same language containing only full substitutions, by replacing bindings of the form $x \mapsto t$, where t does not contain variables, by $\exists x \; ; \quad x = t$.

LEMMA 39. *Every formula $\pi \in DPL(\sigma)$ is equivalent to a formula of one of the following forms (for some ψ, x, χ, σ, where σ is full):*

1. *$\neg\neg\chi; \sigma$.*

2. *$\psi; \exists x; \neg\neg\chi; \sigma$.*

Proof. First rewrite π into a formula that contains only full substitutions. After that, the only non-trivial case in the translation instruction is the case of $\tau \; ; \quad \psi$, where τ is full and ψ is of the first form, i.e. where ψ is equivalent to $\neg\neg\chi; \sigma$, for some χ, σ, with σ full. In this case, $\tau \; ; \quad \psi$ is equivalent to $\neg\neg(\tau; \chi) \; ; \quad \tau \circ \sigma$, where $\tau \circ \sigma$ is full because σ and τ are. ∎

THEOREM 40. $(\exists x \cup \exists y)$ *cannot be expressed in $DPL(\sigma)$.*

Proof. Suppose $\pi \in DPL(\sigma)$ is equivalent to $(\exists x \cup \exists y)$. Take a model with as domain the natural numbers, and let R be the interpretation of π. By Lemma 39, it follows that π is equivalent to $\psi; \exists z; \neg\neg\chi; \sigma$, for some formulae ψ, χ, some variable z and some full substitution σ (otherwise, π would be deterministic). Two cases can be distinguished.

1. z does not occur in σ. Without loss of generality, assume that $z \neq x$. Take any pair of assignments g, h such that $g \neq h$ and $g \sim_x h$. Then gRh. Take any $k \neq h$ such that $k \sim_z h$. Then gRk, but g and k differ with respect to two variables (x and z), which is in contradiction with the fact that π is equivalent to $(\exists x \cup \exists y)$.

2. z occurs in σ. By the fact that there are no function symbols involved, and by the fact that σ is full, there must be exactly one binding in σ of the form $u \mapsto z$. We can apply the same argument as before, now using u instead of z, and again we arrive at a contradiction.

∎

Every substitution is equivalent to a DPL formula containing only full substitutions, and since every full substitution without function symbols is a renaming, and therefore has a converse that is also a renaming, we get:

LEMMA 41. *Every converse substitution containing no function symbols is equivalent to a formula in $DPL(\sigma)$.*

This immediately gives:

THEOREM 42. $(\exists x \cup \exists y)$ *cannot be expressed in* $DPL(\sigma, \breve{})$.

LEMMA 43. *Every formula in $DPL(\sigma, \cup)$ is equivalent to a formula of the form $\pi_1 \cup \ldots \cup \pi_n$ $(n \geq 1)$ where each $\pi_i \in DPL(\sigma)$.*

THEOREM 44. $(x \mapsto f(x))^{\breve{}}$ *cannot be expressed in* $DPL(\sigma, \cup)$.

Proof. Suppose $\pi \in DPL(\sigma, \cup)$ is equivalent to $(x \mapsto f(x))^{\breve{}}$. By Lemma 43, we can assume that π is of the form $\pi_1 \cup \ldots \cup \pi_n$, where each $\pi_i \in DPL(\sigma)$. Consider the model with as domain $\{0, \ldots, n\}$, and where f is interpreted as the "successor modulo $n+1$" function.

Let us say that a relation R fixes a variable x if for $\forall gh \in cod(R)$: $g \sim_x h$ implies that $h = g$. Analysing each π_i, we can distinguish the following two cases.

- π_i is equivalent to $\neg\neg\chi; \sigma$, with σ full. Then $[\![\pi_i]\!]$ fixes x.

- π_i is equivalent to $\psi; \exists y; \neg\neg\chi; \sigma$, again with σ full. If y occurs in σ, then let z_i be the (unique) variable such that σ contains a binding of the form $z_i \mapsto f^k(y)$. If σ does not contain y then let $z_i = y$. Then it must be the case that $[\![\pi_i]\!]$ fixes z_i, for otherwise $[\![\pi_i]\!]$ is not injective.

Thus, we have that every π_i fixes some variable z_i. Let $\{z_1, \ldots, z_m\}$ be all variables that are fixed by some π_i (where $m \leq n$).

Consider all possible ways of assigning objects from the domain to the variables $z_1, \ldots z_m$ (assigning 0 to all other variables). This gives us $(n+1)^m$ assignments, each of which is in the co-domain of π. Now, of this space of assignments, each π_i can cover only a small part: at most $(n+1)^{m-1}$ (since one variable is fixed). So, together, $\pi_1, \ldots, \pi_n$ can cover at most $n * (n+1)^{m-1} = (n+1)^m - (n+1)^{m-1} < (n+1)^m$ assignments, which means that some assignments are not in the co-domain of π. This is in contradiction with the fact that π is equivalent to $(x \mapsto f(x))^{\breve{}}$. ∎

By symmetry, we get the following

THEOREM 45. $x \mapsto f(x)$ *cannot be expressed in* $DPL(\breve{\sigma}, \cup)$.

Finally we have

THEOREM 46. $\exists y(y = x; \exists x \; ; \; Rxy)$ *cannot be expressed in* $DPL(\cup, \sigma, \breve{})$.

Proof. The same proof as for Theorem 44 can be used. Assume a signature without function symbols. Let the domain of the model be the set $\{0, \ldots, n\}$. Let R be interpreted as "successor modulo $n+1$". Then R is interpreted in the same

way as f was in the proof of Theorem 44. Notice that, under this interpretation, $\exists y(y = x \; ; \; \exists x \; ; \; Rxy)$ means the same as $(x \mapsto f(x))^\vee$ did in the proof of Theorem 44. It follows that $\exists y(y = x; \exists x \; ; \; Rxy)$ cannot be expressed in $DPL(\cup, \sigma)$. Since the signature contains no function symbols, it follows by Lemma 41 that this formula cannot be expressed in $DPL(\cup, \sigma, ^\vee)$ either. ∎

DPL and Dynamic Relational Algebra

Yet another way in which the logic of DPL and sundry systems has been studied is by looking at the connection with dynamic relational algebra.

A dynamic relation algebra is an algebra for the signature $\{\bot, \sim, ;\}$, i.e. it consists of all binary relations on a set B (all members of $\mathcal{P}(B \times B)$), with $\bot$ interpreted as the empty relation , ; as relation composition, and $\sim$ as dynamic negation. A dynamic relation algebra is completely determined by its carrier set B.

Note that this is different from the usual relational algebra in the sense of [Tarski, 1941], where the signature consists of the Boolean operations $\{-, \cap, \cup, \bot, \top\}$ and the order operations plus the identity relation $\{\circ, ^\vee, \text{id}\}$. In fact, dynamic relation algebra can be viewed as a small non-Boolean fragment of relation algebra. Dynamic negation can be defined in ordinary relation algebra by means of:

$$\sim R := \text{id} \cap -(R; \top)$$

Hollenberg [Hollenberg, 1997] gives the following axiomatisation of dynamic relation algebra:

$$
\begin{aligned}
\sim R; R &= \bot \quad \text{(falsum definition)} \\
R; \bot &= \bot \quad \text{(falsum right)} \\
\text{id}; R &= R \quad \text{(identity left)} \\
R; (S; T) &= (R; S); T \quad \text{(associativity)} \\
\sim R; \sim S &= \sim S; \sim R \quad \text{(test permutation)} \\
R &= (\sim\sim R); R \quad \text{(domain test)} \\
\sim\sim(\sim R; \sim S) &= \sim R; \sim S \quad \text{(test composition)} \\
\sim(R; S); R &= (\sim(R; S); R); \sim S \quad \text{(modus ponens)} \\
\sim(R; (S \vee T)) &= \sim((R; S) \vee (R; T)) \quad \text{(distribution)},
\end{aligned}
$$

where $R \vee S$ is an abbreviation of $\sim(\sim R; \sim S)$.

Note that $\sim R; R = \bot$ can be viewed as a definition of $\bot$. Order is important, for $R; \sim R$ does not always denote the empty relation.

Tests are subsets of the identity relation. $\sim R$ is always a test, and R is a test iff $\sim\sim R = R$, so $\sim\sim(\sim R; \sim S) = \sim R; \sim S$ expresses that the composition of two tests is again a test.

The fact that $\sim(R; S); R = (\sim(R; S); R); \sim S$ is called *modus ponens* is explained by defining $R \Rightarrow S$ as $\sim(R; \sim S)$ and substituting $\sim S$ for S. This gives:

$$(R \Rightarrow S); R = (R \Rightarrow S); R; \sim\sim S.$$

Hollenberg [Hollenberg, 1997] has a proof that this axiomatisation is sound and complete for dynamic relation algebra. In [Hollenberg and Visser, 1997] it is proved that in any model $(M, \perp, \sim, ;)$ of this axiom system, dynamic negation is fully determined by the underlying monoid $(M, ;)$.

In [van Benthem and Cepparello, 1994] it was shown that DPL-negation $\sim$ is the only permutation-invariant operator in dynamic relational algebra that satifies the following conditions:

$$
\begin{aligned}
\sim\perp &= \text{id} \\
\sim(\cup_i R_i) &= \cup_i(\sim R_i) \\
\sim\sim R \cup (R; \top) &= R; \top \\
\sim R; R &= R.
\end{aligned}
$$

Permutation-invariant operators are operators O satisfying

$$\pi(O(R, S)) = O(\pi(R), \pi(S))$$

for every permutation π on the state set on which the relations are defined.

This result about DPL-negation led [van Benthem and Cepparello, 1994] to conjecture that DPL is complete for dynamic relational algebra, in the sense that counterexamples to relational identities in the vocabulary $\{\perp, \sim, ;\}$ are expressible in DPL. This conjecture was proved in [Visser, 1997].

THEOREM 47 (Visser). *Schematic validity in DPL is complete for dynamic relational algebra.*

Proof. Suppose some relational equation E in the vocabulary $\{\perp, \sim, ;\}$ is refuted by a family of binary relations $\{R_a \mid a \in A\}$ over some carrier set B, where A is the set of atomic relation symbols occurring in the equation E.

We will consider DPL formulae over the variables x, y. Consider the space $B^{\{x,y\}}$ of all assignments in B to x and y.

DPL formulae in x, y denote relations between input and output assignments to $\{x, y\}$. For each R_a we define a new relation $\widehat{R}_a$ on $B^{\{x,y\}}$, by setting

$$\widehat{R}_a = \{(\{x \mapsto s_1, y \mapsto s_2\}, \{x \mapsto s_3, y \mapsto s_4\}) \mid R_a s_1 s_3\}.$$

The crucial insight is that the function $g \mapsto g(x)$ is a functional bisimulation (also known as: a p-morphism) from the transition system of the $\widehat{R}_a$ on $B^{\{x,y\}}$ to the transition system of the R_a on B, since $\sim$ and ; are safe for bisimulation.

Let the new relation symbol I denote identity in $(B, \{R_a \mid a \in A\})$. Then the relations $\widehat{R}_a$ can be defined in DPL by means of:

$$\exists y; R_a xy; \exists x; I xy; \exists y.$$

If the relations at the lefthand and the righthand side of E are different, their originals under g are different too. Thus, an inequality defined in terms of $\sim$ and ; on $(B, \{R_a \mid a \in A\})$ corresponds to an inequality on

$$(B^{\{x,y\}}, \{\widehat{R}_a \mid a \in A\}).$$

This shows that the left- and righthand sides of the equation E yield a pair of non-equivalent DPL formulae.   ■

7   DYNAMIC LOGIC AND NATURAL LANGUAGE SEMANTICS

7.1   Introduction

As we saw in Section 6 the difference between dynamic predicate logic (DPL) and quantified dynamic logic (QDL) is that whereas the latter makes a distinction, both in the syntax and in the semantics, between static formulae and dynamic programs, the former has basically only one kind of construct: programs. All formulae are programs, so there is no distinction either in syntactic category or in semantic type, between different kinds of linguistic constructions: all constructs are given a dynamic interpretation. The motivation for this is not a matter of expressive power, but one of 'ideology'. The difference can be characterised as follows: whereas QDL acknowledges two different notions of meaning: one descriptive and one imperative, DPL embodies a unified conception: all meanings are relations between states. By doing so, DPL instantiates a conception of meaning that has become prominent in natural language semantics from the early eighties onward and that sometimes is summarised in the slogan 'Meaning is context change potential'.

This view on meaning is often referred to as 'dynamic semantics'. Various people have contributed to it, motivated by various concerns. Broadly speaking we may discern two main trends. First of all there is work that focuses on epistemic and pragmatic issues that arise in connection with presuppositions, the structure of information exchange, but also with conditionals and modal expressions. Very influential in this trend is the early work by Stalnaker on assertion and presuppositions [Stalnaker, 1974; 1979]. Other early work is that of Veltman [Veltman, 1984]. A second influx of ideas derives from issues concerning semantics, in particular pronominal reference and quantification. This is exemplified by work of Heim [Heim, 1982; 1983] and Kamp [Kamp, 1981; Kamp and Reyle, 1993]. Somewhat orthogonal to these two trends is the work on game-theoretical semantics for natural language explored by Hintikka and others [Hintikka, 1983]. Another approach that has clear affinities with a dynamic approach is that of situation semantics [Barwise and Perry, 1983].

The variety of empirical subjects that prompted the use of dynamic concepts have resulted in an analogous variety of systems. Also, different authors entertain different views on how the use of these concepts affect the notion of meaning as it applies to natural language. Some maintain a truth conditional, propositional notion of meaning and relegate dynamics to the realm of language use, i.e. pragmatics, whereas others argue that the notion of meaning as such needs to be viewed

as a dynamic concept. Yet others take a middle position and locate the dynamic aspects in the construction of representations that themselves have a static interpretation. Cf., [Stalnaker, 1998; Kamp, 1990; Groenendijk and Stokhof, 2000] for discussion. In what follows we focus on those systems in which the use of dynamic concepts directly interacts with the concept of meaning that is modelled.

The general characteristic of dynamic systems is that formulae are interpreted as entities that change the context. In natural language semantics and pragmatics, 'context' is an umbrella concept, that covers a wide variety of elements that are somehow tied to the use and the interpretation of expressions. Speaker and addressee, time and place, elements from preceding discourse, objects and properties introduced in conversation, information of speech participants about the world, themselves, each other, and so on, — all these factors may be involved in linguistic exchanges.

Within a particular system the relevant aspects of the context are represented in the system as states. Which aspects counts as relevant depends on the specific application and/or the expressive resources of the system. For example, in DPL states are simply assignments of values to variables, and this reflects that DPL is focused on those aspects of context that concern binding relationships between antecedents, i.e. quantified noun phrases and proper names, and anaphoric expressions, i.e. pronouns. When one extends or alters the scope of application, the notion of a state changes as well, resulting in a modification or extension of the original system. In this type of system states consist of objects and their properties and relationships and dynamic interpretation changes them by adding new objects, establishing new relationships, and so on.

As we noted, another important aspect of the context is the information of the speech participants. On a dynamic view the utterance of a sentence is to be regarded as an instruction to the speech participants to update their information with the content of the utterance. (Hence the name 'update semantics'.) A system modelling this will have states that represent the informational states of speech participants, e.g., as sets of propositions, sets of worlds, possibilities, or situations. Utterances then are interpreted as updates of such states. For example, a dynamic ('update') semantics for a conditional $\varphi \to \psi$ would (roughly) be defined as an operation that checks whether every update of a given set of possibilities with the antecedent satisfies satisfy the consequent.

Actually, these points of view are not incompatible. For example, we can look upon DRT- and DPL-like systems as concerned with information as well, viz., with information *about the discourse*: the entities that have been introduced, their properties and relationships, and the various possibilities that are available for anaphoric reference. Information in the update sense is then information *about the world*: information about the actual state of things as well as possibilities that are still open. As a matter of fact, combining these two perspectives is a more interesting exercise than just putting two orthogonal systems together: there are interesting interactions between the two.

In the remainder of this section we start with the use of dynamic logic in accounting for certain problems in semantics . Then we will turn to systems motivated by epistemic-pragmatics concerns. Finally, we will briefly look at combined systems.

7.2   Dynamic Semantics

Dynamic Phenomena

Discourse Representation Theory (DRT,[Kamp, 1981; Kamp and Reyle, 1993; van Eijck and Kamp, 1997]), File Change Semantics (FCS, [Heim, 1983]), dynamic predicate logic (DPL, [Groenendijk and Stokhof, 1991]) are systems that originated in the late eighties, early nineties of the last century. Their initial motivation was linguistic. They grew out of attempts to deal with certain facts concerning anaphora and binding that had resisted adequate treatment in the Montague framework that dominated natural language semantics at the time. Other important areas of application are tense and aspect, presupposition, plurality. For more extensive discussion of the linguistic applications of these systems, cf., [Chierchia, 1995], [van Benthem *et al.*, 1997], and the references given above. Here it suffices to give just a brief illustration of one example of the kind of phenomena these systems were intended to deal with: scope and binding. Basically, in this area there are two groups of problems: cross-sentential anaphoric relationships and so-called 'donkey'-constructions, which present a particular form of intra-sentential binding.

Cross-sentential anaphora refers to constructions such as:

A man entered the pub. He wore a black hat.

The pronoun 'He' in the second sentence is most naturally taken to refer back to, i.e. as an anaphoric reference to, the referent of 'a man' in the first sentence. At the time there was a preference for dealing with anaphora – antecedents relationships in terms of variable binding: the antecedent 'a man' semantically operates as a quantifier, binding the variable that corresponds to the pronoun. The problem with this type of cross-sentential antecedent – anaphora relationships is, of course, that the binding can be established only when the discourse is finished. And even then, one must take care with such antecedents as 'One man', so as not to end up with the wrong interpretation ('One man φ. He ψ' is not the same as 'One man φ and ψ')

Donkey anaphora is connected with intra-sentential binding, e.g., between antecedent and consequent in conditional constructions:

If John spots a good investment opportunity, he grasps it.

The fact to be accounted for here is the binding of the anaphoric pronoun in the consequent by the indefinite noun phrase in the antecedent in such a way that the indefinite gets 'universal' force: the sentence is most naturally taken to express that John grasps every opportunity he sees. (Not all sentences with this structure

have a universal (also called 'strong') reading: 'If I have a quarter, I'll put it in the parking meter'. Cf., [Kanazawa, 1994] for extensive discussion of so-called 'weak' and 'strong' readings of these kinds of constructions.)

Note that in each case the problem is not finding an adequate representation of the meanings of these sentences in (first) order logic. Rather, the problem is coming up with such a representation while using the standard meanings of the expressions involved, and deriving the representation in an 'on line', i.e. incremental fashion, without delayed interpretation or after the fact re-analysis.

DPL again

Although it was not the first system to be developed, we focus on DPL because it is the most 'pure' instantiation of a dynamic view on meaning. It was developed because of a certain dissatisfaction with the representational, non-compositional nature of, e.g., DRT. It intends to do away with dynamically constructed representations as part of the semantics and wants to locate the dynamics purely in the meanings themselves.

The system   The standard reference is [Groenendijk and Stokhof, 1991], earlier similar views were developed in [Barwise, 1987] and [Staudacher, 1987]. The original DPL-system stayed as close as possible to standard first order logic FOL: it employed the same language and only changed the semantics. In section 6 the system was given in a form that stayed close to that of QDL. What follows is the original formulation, i.e. with the syntax of FOL and an adapted semantics.

$$
\begin{array}{rcl}
t & ::= & v \mid c \\
\varphi & ::= & Rt_1 \ldots t_n \mid t_1 = t_2 \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \exists v\varphi
\end{array}
$$

The other connectives and the universal quantifier can be defined in the usual fashion. (But note that compared to FOL the choice of base logical constants is limited.)

The semantics uses the same ingredients as that of FOL. A model M is a pair $\langle E, F \rangle$, where E is a non-empty set and $F^{\mathbf{M}}(c) \in E$ and $F^{\mathbf{M}}(R^n) \subseteq E^n$. States $g \in S$ are assignments $V \to E$. As usual $g \sim_v h$ denotes the state h that differs from g at most on v.

Interpretation of terms is given by: $[\![t]\!]_g^{\mathbf{M}} = g(t), F^{\mathbf{M}}(t)$ for variables and constants respectively. Formulae denote subsets of $S \times S$:

$$
\begin{array}{lll}
{}_g[\![Rt_1 \ldots t_n]\!]_h^{\mathbf{M}} & \text{iff} & g = h \;\&\; \langle [\![t_1]\!]_g^{\mathbf{M}} \ldots [\![t_n]\!]_g^{\mathbf{M}} \rangle \in F^{\mathbf{M}}(R) \\[4pt]
{}_g[\![t_i = t_j]\!]_h^{\mathbf{M}} & \text{iff} & g = h \;\&\; [\![t_i]\!]_g^{\mathbf{M}} = [\![t_j]\!]_g^{\mathbf{M}} \\[4pt]
{}_g[\![\neg\varphi]\!]_h^{\mathbf{M}} & \text{iff} & g = h \;\&\; \text{there exists no } g' : {}_g[\![\varphi]\!]_{g'}^{\mathbf{M}} \\[4pt]
{}_g[\![\varphi_1 \wedge \varphi_2]\!]_h^{\mathbf{M}} & \text{iff} & \text{there exists a } g' : {}_g[\![\varphi_1]\!]_{g'}^{\mathbf{M}} \;\&\; {}_{g'}[\![\varphi_2]\!]_h^{\mathbf{M}} \\[4pt]
{}_g[\![\exists v\varphi]\!]_h^{\mathbf{M}} & \text{iff} & \text{there exists a } g' : g \sim_v g' \;\&\; {}_{g'}[\![\varphi]\!]_h^{\mathbf{M}}
\end{array}
$$

Note that although all formulae denotes relations between states (assignments), only conjunction and existentially quantified formulae actually change states, the others are tests. Conjunction is effectively re-interpreted as program composition, and an existential quantified formula has the cumulative effect of re-setting the state with respect to the variable and feeding the result into the formulae. It is easy to see that

$$\text{for all } \mathbf{M}, g, h \colon {}_g[\![\exists x \varphi]\!]_h^{\mathbf{M}} \text{ iff } {}_g[\![x :=? \; ; \; \varphi]\!]_h^{\mathbf{M}}$$

The definitions of truth and validity as given in section 6 carry over, as do the notions of production set and satisfaction set. Equivalence as identity of interpretation transcends identity of input (satisfaction set) and output (production set). Cf. section 6 for an example in DPL′. $\neg(Px \land \neg Px)$ and $\exists x \neg(Px \land \neg Px)$ both have S as their satisfaction set and as their production set. But their meanings are different: the identity relation on S, and the set of all pairs g, h such that $g \sim_x h$, respectively. Note the meaning of a test *can* be completely characterised in terms of its satisfaction set and its production set and that all valid tests denote the identity relation on S.

Some characteristic examples   The following two examples exhibit characteristic properties of the semantics of DPL. Both concern the extended binding force of the existential quantifier.

The first one concerns the interaction of the existential quantifier and conjunction. In $\exists x Px \land Qx$ the existential quantifier $\exists x$ randomly assigns a value to x that is passed on to Px, and tested. If it succeeds, conjunction, which is relational composition, passes it on to Qx, to be tested again. (We leave out reference to the model M whenever this does not lead to confusion.)

$$\begin{aligned}{}_g[\![\exists x Px \land Qx]\!]_h \quad &\text{iff} \quad \text{there exists a } g' \colon {}_g[\![\exists x Px]\!]_{g'} \; \& \; {}_{g'}[\![Qx]\!]_h \\ &\text{iff} \quad \text{there exists a } g' \colon g \sim_x g' \; \& \; g'(x) \in F(P) \; \& \; g'(x) \in F(Q)\end{aligned}$$

This allows DPL to deal with cross-sentential anaphora of the kind: 'A man . . . . He . . . '

Note that extended binding can also occur across other quantifiers, as e.g., in $\exists x Px \land \exists y Rxy$, where the occurrence of x in Rxy is bound by $\exists x$; and across negation: in $\exists x Px \land \neg Qx$ the x in $\neg Qx$ is also bound by $\exists x$. Note that since we do not prohibit the same quantifier to occur more than once we have to be careful which occurrence of a quantifier binds a particular variable occurrence: in $\exists x Px \land Qx \land \exists x Hx$ the occurrence of x in Hx is bound by the last occurrence of $\exists x$.

The second example of extended binding concerns the behaviour of the existential quantifier in conditional constructions. Consider the formula $\exists x Px \to Qx$, which is shorthand for $\neg(\exists x Px \land \neg Qx)$. Here we have an existential quantifier in the antecedent of a conditional and an occurrence of x in the consequent that in FOL would be free. However, if we compute its meaning, we see that the second

occurrence is bound by the existential quantifier, and, moreover, that the latter gets universal force:

$$
\begin{aligned}
{}_g[\![\exists x P x \rightarrow Q x]\!]_h \quad &\text{iff} \quad {}_g[\![\neg(\exists x P x \wedge \neg Q x)]\!]_h \\
&\text{iff} \quad \text{there exists no } g' \colon {}_g[\![\exists x P x]\!]_{g'} \ \&_{g'} \ [\![\neg Q x]\!]_h \\
&\text{iff} \quad \text{for all } g' \colon \text{if } {}_g[\![\exists x P x]\!]_{g'} \text{ then } {}_{g'}[\![Q x]\!]_h
\end{aligned}
$$

So, every way of re-setting the value of x to one that satisfies P is one that satisfies Q.

Note that the extended binding force of the existential quantifier is blocked by negation: in $\neg \exists x P x \wedge Q x$ the occurrence of x in $Q x$ is free. This is because the negation turns $\exists x P x$ into a test: the value assigned by $\exists x$ to x remains local to $P x$, and is not passed on to $Q x$. Thus in $\exists x P x \rightarrow Q x$ the binding of the existential quantifier in the antecedent extends to the consequent, but not beyond the formula as a whole.

Thus we can distinguish between formulae that are *internally dynamic*, i.e. in which an existential quantifier binds variables outside its scope, but only in the formula itself; and those that are *externally dynamic*, in which existential quantifiers have the power to bind variables in additional formulae that are added to its right. The latter are responsible for DPL's treatment of extra-sentential, i.,e., discourse binding; the former deal with internal binding from antecedent to consequent.

Other properties   Other characteristic properties of the DPL-logic follow in a straightforward manner from the semantics. Double negation fails in view of negation blocking dynamic binding; conjunction and the existential quantifier can not be defined in terms of, e.g., negation, disjunction and the universal quantifier, because of the asymmetry of the respective expressions w.r.t. binding; conjunction is not unconditionally commutative and idempotent; the existential and universal quantifiers are not fully interdefinable; and finally, we can not take alphabetic variants of existentially quantified formulae.

As for entailment, neither inclusion of truth conditions, nor meaning inclusion, provide a suitable definition. The reason is that we want existential quantifiers in the premises of an argument to be able to bind variables in the conclusion, in view of the possibilities of antecedent – anaphora links in natural language reasoning: from 'A man came in carrying a stick' we want to be able to conclude 'So, he was carrying a stick'. So ψ follows from $\varphi_1 \ldots \varphi_n$ iff in all models every interpretation of the premises (in sequential order, of course) leads to a successful interpretation of the conclusion:

$$
\begin{aligned}
&\varphi_1, \ldots, \varphi_n \models \psi \text{ iff} \\
&\quad \text{for all } \mathbf{M}, g, h \colon \text{if } {}_g[\![\varphi_1 \wedge \cdots \wedge \varphi_n]\!]_h^{\mathbf{M}}, \text{then there exists an } h' \colon {}_h[\![\psi]\!]_{h'}^{\mathbf{M}}
\end{aligned}
$$

In terms of DPL$'$ (see section 6):

$$\varphi_1, \ldots, \varphi_n \models \psi \quad \text{iff} \quad \text{for all } \mathbf{M} \colon [\![[\varphi_1 \; ; \; \cdots \; ; \; \varphi_n] \langle \psi \rangle \top]\!]^{\mathbf{M}}$$

equals the set of all assignments.

It is easily checked that, e.g., $\exists x Px \models Px$, as required. Further we have:

$$\varphi_1, \ldots, \varphi_n \models \psi \quad \text{iff} \quad \models (\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$$

Notice that if no binding occurs from premises to conclusion, the notion of entailment defined boils down to the truth-conditional one. It is easily checked that entailment is not reflexive and also not transitive.

DPL being a first order language, it differs from FOL in its non-standard binding behaviour. As we saw in section 6, FOL can be embedded in DPL in a straightforward way. Since DPL$'$ can be translated into FOL (cf., section 6), the same holds for DPL.

Context   As was noted above, contexts in DPL are assignments of values to variables, satisfying certain descriptive conditions. What they represent are the individuals and their properties that have been introduced in a discourse (a text, a conversation), e.g., by proper names or descriptions, or by indefinite NPs. Other expressions, such as pronouns, may draw from this pool of available referents. In DPL this is accounted for via the use of (indexed) variables. Context-change is represented through operations on assignments, as, for example, by the existential quantifier, which 'resets' the context with regard to a particular variable. (Cf., the formulation of DPL in section 6, that brings this out more explicitly, by regarding the existential quantifier as a construct of its own.)

Discourse Representation Theory

Now we briefly introduce a very streamlined and basic version of Discourse Representation Theory (DRT). For an extensive introduction, the reader is referred to the standard [Kamp and Reyle, 1993]. The differences between DPL and DRT are quite like those between DPL and DPL$'$ or DPL and QDL: whereas DPL is a 'pure' language in which no distinction is made between programs and statements, DRT, like DPL$'$ and QDL, does make such a distinction, between what are called 'conditions' and what are called 'discourse representation structures' (DRSs). This syntactic distinction is reflected in the semantics, and is motivated by what Kamp in his seminal paper on DRT [Kamp, 1981] claims is essential for a proper account of natural language meaning, viz., that it 'combines a definition of truth with a systematic account of semantic representations' (*op.cit.*, p.1). Thus, the dynamics in DRT takes place in the building of semantic representations.

The system   The canonical format of DRT uses so-called box-notation (see below for some examples). In order to facilitate comparison, however, we recast the

syntax and semantics of DRT in a linear format. The non-logical vocabulary consists of n-place predicates, individual constants, and variables. Logical constants are negation $\neg$, implication $\rightarrow$, and identity $=$.

DRT terms are constants and variables:

$$t \quad ::= \quad x \mid c$$

Conditions φ and DRSs Φ are defined as follows:

$$
\begin{aligned}
\varphi &\quad ::= \quad Rt_1 \ldots t_n \mid t_1 = t_2 \mid \neg \Phi \mid \Phi_1 \Rightarrow \Phi_2 \\
\Phi &\quad ::= \quad [x_1 \ldots x_k][\varphi_1 \ldots \varphi_n]
\end{aligned}
$$

Disjunction of DRSs can be defined in the usual way.

In the box notation, a DRS looks like this:

$$
\boxed{
\begin{array}{c}
x_1 \ldots x_k \\
\hline
\varphi_1 \\
\vdots \\
\varphi_n
\end{array}
}
$$

where the φ_i are conditions and the x_i introduced variables. An example of a conditional DRS built from two other DRSs in box notation looks like this:

$$
\boxed{
\begin{array}{c}
x,\ y \\
\hline
Px,\ Qy,\ Rxy
\end{array}
}
\quad \Rightarrow \quad
\boxed{
\begin{array}{c}
\ \\
\hline
Sxy
\end{array}
}
$$

The language of DRT resembles that of QDL and DPL$'$ in its 'mixed mode' nature. This carries over to the semantics.

Models for the DRS-language are the same as those for DPL, as are assignments and the interpretation of terms. Conditions are interpreted as FOL-formulae, whereas DRSs get a relational meaning. Thus, like in the case of QDL (cf., section 6)), the semantics is defined by simultaneous recursion. Note that we use total assignments instead of partial ones, as is customarily the case in DRT. For present purposes, the difference can be neglected.

$$
\begin{aligned}
\mathbf{M} \models_g Rt_1 \ldots t_n \quad &\text{iff} \quad \langle \llbracket t_1 \rrbracket_g^{\mathbf{M}} \ldots \llbracket t_n \rrbracket_g^{\mathbf{M}} \rangle \in F_{\mathbf{M}}(R) \\
\mathbf{M} \models_g t_1 = t_2 \quad &\text{iff} \quad \llbracket t_1 \rrbracket_g^{\mathbf{M}} = \llbracket t_2 \rrbracket_g^{\mathbf{M}} \} \\
\mathbf{M} \models_g \neg \Phi \quad &\text{iff} \quad \text{there exists no } h \colon {}_g \llbracket \Phi \rrbracket_h^{\mathbf{M}} \\
\mathbf{M} \models_g \Phi_1 \Rightarrow \Phi_2 \quad &\text{iff} \quad \text{for all } h \colon \text{ if } {}_g \llbracket \Phi \rrbracket_h^{\mathbf{M}} \text{ there exists a } k \colon {}_h \llbracket \Phi_2 \rrbracket_k^{\mathbf{M}} \\
{}_g \llbracket [x_1 \ldots x_k][\varphi_1 \ldots \varphi_n] \rrbracket_h^{\mathbf{M}} \quad &\text{iff} \quad g \sim_{x_1 \ldots x_k} h \ \& \ \mathbf{M} \models_h \varphi_1 \ldots \mathbf{M} \models_h \varphi_n
\end{aligned}
$$

DRT and DPL The close link between DRT and DPL is illustrated by the following embedding of DRT into DPL:

$$
\begin{aligned}
(Rt_1 \ldots t_n)\dagger &= Rt_1 \ldots t_n \\
(t_i = t_j)\dagger &= t_i = t_j \\
(\neg\Psi)\dagger &= \neg(\Psi\dagger) \\
(\Phi_1 \Rightarrow \Phi_2)\dagger &= \Phi_1\dagger \to \Phi_2\,\dagger \\
([x_1 \ldots x_k][\varphi_1 \ldots \varphi_n])\dagger &= \exists x_1 \ldots \exists x_n[\varphi_1\,\dagger \wedge \ldots \wedge \varphi_n\dagger]
\end{aligned}
$$

The embedding is meaning-preserving in the following sense:

$$
\begin{aligned}
\mathbf{M} \models_g \varphi &\quad \text{iff} \quad \text{there exists an } h\colon {}_g[\![\varphi\dagger]\!]^{\mathbf{M}}_h \\
{}_g[\![\Phi]\!]^{\mathbf{M}}_h &\quad \text{iff} \quad {}_g[\![\Phi\dagger]\!]^{\mathbf{M}}_h
\end{aligned}
$$

Context As it turns out, the notion of a context in DRT does not differ all that much from the one DPL is concerned with: both model basically the same features of a discourse context. But the two systems model context in different ways: DPL uses only assignments and operations on them, DRT uses special types of expressions in its syntax.

Variations and extensions

A number of variations on DRT, DPL and other systems have been proposed in the literature. Some are motivated by reasons of formal simplicity and elegance, others by conceptual and descriptive reasons. It is beyond the scope of this article to discuss them extensively; here it suffices to point to a number of issues motivating these alternatives.

Partial assignments One difference between DPL and DRT is the use that the former makes of total assignment functions, instead of the partial ones used by DRT. The choice for partial assignments, that interpret only the variables that are explicitly introduced in a discourse, is a natural one from the perspective of a procedural interpretation, which was one of the motivations of the original DRT-system (cf. above). The use of total assignments in the original DPL system was mainly motivated by a wish to stay as close as possible to the semantics of standard first order logic. Reformulating the DPL-semantics using partial assignments is an easy exercise. We simply let the interpretation be undefined in case a formula contains occurrences of variables that are not in the domain of the assignment function. The only interesting case is the existential quantifier. Here we should let the quantifier extend the domain of the assignment function, if necessary, and let it assign an arbitrary value to the new element in its domain. Cf., e.g., [Vermeulen, 1995] and the system in section 7.4 below.

Fresh variables   One of the advantages of using partial assignments is that it becomes more natural to constrain the use of variables in the syntax. Recall some of the more awkward logical properties of DPL, such as the failure of reflexivity of entailment:

$$Px \wedge \exists x Px \not\models Px \wedge \exists x Px$$

This essential depends on the possibility of a variable occurring in the same formula first free and then bound by an existential quantifier. One way of preventing this (and similar) issues, is to require the existential quantifier to always use a 'fresh' variable. Cf., also the discussion below, on incremental semantics.

Compositionality   As the preceding discussion will have made clear, the discussion between DPL and DRT centres on compositionality. In DRT the representational level of DRSs plays an essential role, and the cognitive plausibility of the resulting system depends on their presence (cf., the discussion in [Kamp, 1981, section 1]). Other formulations of a compositional alternative for DRT have been proposed by, among others, Zeevat [Zeevat, 1989], Muskens [Muskens, 1996], and Van Eijck and Kamp [van Eijck and Kamp, 1997]. DPL's reliance on an indexing mechanism on variables to account for anaphoric binding has been criticised since it diminishes the plausibility of the appeal to compositionality considerations, and has spurred a number of alternative approaches, such as Dekker's 'predicate logic with anaphora' [Dekker, 2002], [Butler and Mathieu, 2004]. Cf., also the incremental system discussed below in section 7.2, and the combination of update semantics and dynamic semantics in section 7.4.

Stacks and registers   The use of DPL as a theory of testing and resetting registers was explored by Visser [Visser, 1998] and Vermeulen [Vermeulen, 1995; 2000]. The basic idea of a stack semantics for DPL is developed in [Vermeulen, 2001]. The idea is to replace the destructive assignment of ordinary DPL, which throws away old values when resetting, by a stack valued one, that allows old values to be re-used. Stack valued assignments assign to each variable a stack of values, the top of the stack being the current value. Existential quantification pushes a new value on the stack, but there is also the possibility of popping the stack, to re-use a previously assigned value. Adding explicit 'push' and 'pop'-operators to the language, has some interesting consequences. An illustrative example concerns its efficiency in expressing mixed scopes.The idea is as follows. We add $[_x$ and $_x]$ as two new programs and define their semantics as follows:

$$_g[\![\,[_x\,]\!]_h \quad \text{iff} \quad g[x\rangle h$$
$$_g[\![\,_x]\,]\!]_h \quad \text{iff} \quad h[x\rangle g$$

where $g[x\rangle h$ holds by definition iff there is a d in the domain with $h(x) = d : g(x)$, (i.e. $h(x)$ equals the result of pushing d on top of the x-stack of g), and $h(y) = g(y)$ for all y with $y \neq x$. Clearly, the programs $[_x$ and $_x]$ then function as pop and push for the x-stack.

Now consider the FOL-statement:

$$\exists x \exists y \exists z \exists u (Rxy \wedge Ryz \wedge Rzu \wedge Rux)$$

This can be expressed in DPL more succinctly as:

$$\exists x \exists y (\exists z Rxy \wedge \exists x (Ryx \wedge Rxz) \wedge Rzx)$$

But using the push and pop programs we can express the same in terms of only two variables.

$$[_x [_y Rxy [_x Ryx [_y Rxy_x] Ryx_y]_x]_y]$$

The variable free indexing of [van Eijck, 2001] is a special case of the Vermeulen method, where there is just a single variable. Below we take a variation on DPL with variable free indexing as point of departure for the development of a fragment of dynamic Montague grammar.

Incremental Semantics

Destructive assignment is the main weakness of DPL as a basis for a compositional semantics of natural language: in DPL, the semantic effect of a quantifier action $\exists x$ is such that the previous value of x gets lost. In what follows we first replace DPL by the strictly incremental system from [van Eijck, 2001]. Subsequently, we develop its type theoretic version. This will allow us to give of a fully compositional and incremental semantics that is without the destructive assignment flaw. Similar ideas were developed in [Dekker, 1994; 1996].

We start with a slight variation of the DPL language, in which $\exists$ is a separate expression and ; is used for dynamic conjunction. Assume a first order model $M = (D, F)$. We will use contexts $c \in D^*$, and replace variables by indices into contexts. The set of terms of the language is $\mathbb{N}$. We use $|c|$ for the length of context c.

Given a model $M = (D, F)$ and a context $c = c[0] \cdots c[n-1]$, where $n = |c|$ (the length of the context), we interpret terms of the language by means of $[\![i]\!]_c = c[i]$. Note that $[\![i]\!]_c$ is undefined for $i \geq |c|$; we will therefore have to make sure that indices are only evaluated in appropriate contexts. $\uparrow$ will be used for 'undefined'. This allows us to define the two relations

$$\mathbf{M} \models_c Ri_1 \cdots i_n \text{ and } \mathbf{M} =\!\!|_c Ri_1 \cdots i_n$$

by means of:

$$\mathbf{M} \models_c Ri_1 \cdots i_n \ :\Leftrightarrow \ \forall j (1 \leq j \leq n \rightarrow [\![i_j]\!]_c \neq \ \uparrow) \ \text{ and } \ \langle [\![i_1]\!]_c, \ldots, [\![i_n]\!]_c \rangle \in F(R),$$

$$\mathbf{M} =\!\!|_c Ri_1 \cdots i_n \ :\Leftrightarrow \ \forall j (1 \leq j \leq n \rightarrow [\![i_j]\!]_c \neq \ \uparrow) \ \text{ and } \ \langle [\![i_1]\!]_c, \ldots, [\![i_n]\!]_c \rangle \notin F(R),$$

and similarly for the relations:

$$\mathbf{M} \models_c i_1 = i_2, \ \ M =\!\!|_c i_1 = i_2$$

If $c \in D^n$ and $d \in D$ we use $c \hat{\ } d$ for the context $c' \in D^{n+1}$ that is the result of appending d at the end of c.

The interpretation of formulae can now be given as a map in $D^* \hookrightarrow \mathcal{P}(D^*)$ (a partial function, because of the possibility of undefinedness):

$$
\llbracket \exists \rrbracket(c) := \{ c \hat{\ } d \mid d \in D \}
$$

$$
\llbracket R i_1 \cdots i_n \rrbracket(c) :=
\begin{cases}
\uparrow & \text{if } \exists j (1 \leq j \leq n \text{ and } \llbracket i_j \rrbracket_c = \uparrow) \\
\{c\} & \text{if } M \models_c P i_1 \cdots i_n \\
\emptyset & \text{if } M =\!\mid_c P i_1 \cdots i_n
\end{cases}
$$

$$
\llbracket i_1 = i_2 \rrbracket(c) :=
\begin{cases}
\uparrow & \text{if } \llbracket i_1 \rrbracket_c = \uparrow \text{ or } \llbracket i_1 \rrbracket_c = \uparrow \\
\{c\} & \text{if } M \models_c i_1 = i_2 \\
\emptyset & \text{if } M =\!\mid_c i_1 = i_2
\end{cases}
$$

$$
\llbracket \neg \varphi \rrbracket(c) :=
\begin{cases}
\uparrow & \text{if } \llbracket \varphi \rrbracket(c) = \uparrow \\
\{c\} & \text{if } \llbracket \varphi \rrbracket(c) = \emptyset \\
\emptyset & \text{otherwise}
\end{cases}
$$

$$
\llbracket \varphi_1 \; ; \; \varphi_2 \rrbracket(c) :=
\begin{cases}
\uparrow & \text{if } \llbracket \varphi_1 \rrbracket(c) = \uparrow \\
& \text{or } \exists c' \in \llbracket \varphi_1 \rrbracket(c) \text{ with } \llbracket \varphi_2 \rrbracket(c') = \uparrow \\
\bigcup \{ \llbracket \varphi_2 \rrbracket(c') \mid c' \in \llbracket \varphi_1 \rrbracket(c) \} & \text{otherwise.}
\end{cases}
$$

The definition of $\llbracket \varphi_1 \; ; \; \varphi_2 \rrbracket$ employs the fact that all contexts in $\llbracket \varphi \rrbracket(c)$ have the same length This property follows by an easy induction on formula structure from the definition of the relational semantics. Thus, if one element $c' \in \llbracket \varphi_1 \rrbracket(c)$ is such that $\llbracket \varphi_2 \rrbracket(c') = \uparrow$, then all $c' \in \llbracket \varphi_1 \rrbracket(c)$ have this property.

Dynamic implication $\varphi_1 \rightarrow \varphi_2$ is defined in terms of $\neg$ and $\;;\;$ by means of $\neg(\varphi_1 \; ; \; \neg\varphi_2)$. Universal quantification $\forall\varphi$ is defined in terms of $\exists, \neg$ and $\;;\;$ as $\neg(\exists \; ; \; \neg\varphi)$, or alternatively as $\exists \rightarrow \varphi$.

One advantage of the use of contexts is that indefinite NPs do not have to carry index information anymore. Thus a sentence such as 'Some man loved some woman' can be analysed as:

$$
\exists \; ; \; M i \; ; \; \exists \; ; \; W i + 1 \; ; \; L i (i + 1)
$$

where i denotes the length of the input context. On the empty input context, this gets interpreted as the set of all contexts $[e_0, e_1]$ that satisfy the relation 'love' in the model under consideration. The result of this is that a subsequent sentence 'He$_0$ kissed her$_1$.' can use this contextual discourse information to pick up the references. Thus we assume that pronouns carry index information. But if a procedure for reference resolution of pronouns in context is added we can do away with that assumption.

Extension to Type Logic

Compositionality has always been an important concern in the use of logical systems in natural language semantics. And it is through the use of higher order

logics (such as type theory) that a thoroughly compositional account of, e.g., the quantificational system of natural language could be achieved. The prime example of this development is that of classical Montague Grammar [Montague, 1974; 1970; 1973]. Cf., [Partee, 1997] for an overview. It is only natural, therefore that the dynamic approach was extended to higher order systems.

However, the various proposals that have been made, such as [Groenendijk and Stokhof, 1990; Chierchia, 1992; Jansche, 1998; Muskens, 1995; 1996; 1994; van Eijck, 1997; van Eijck and Kamp, 1997; Kohlhase *et al.*, 1996; Kuschert, 2000], all share a problem with the DPL-system, viz., that of making re-assignment destructive. Interestingly, DRT does not suffer from this problem: the discourse representation construction algorithms of [Kamp, 1981] and [Kamp and Reyle, 1993] are stated in terms of functions with finite domains, and carefully talk about 'taking a fresh discourse referent' to extend the domain of a verifying function, for each new noun phrase to be processed.

Here we present the extension to typed logic of incremental dynamics that is based on variable free indexing and that avoids the destructive assignment problem. The resulting system is called Incremental Type Logic (ITL) [van Eijck, 2000]. Exploiting techniques from polymorphic type theory [Hindley, 1997; Milner, 1978] it uses type specifications of contexts that carry information about the length of the context. E.g., the type of a context is given as $[e]_i$, where i is a type variable. Here, we will cavalierly use $[e]$ for the type of any context, and ι for the type of any index, thus relying on meta-context to make clear what the current constraints on context and indexing into context are. In types such as $\iota \to [e]$, we will tacitly assume that the index fits the size of the context. Thus, $\iota \to [e]$ is really a type scheme rather than a type, although the type polymorphism remains hidden from view. Since $\iota \to [e]$ generalises over the size of the context, it is shorthand for the types $0 \to [e]_0$, $1 \to [e]_1$, $2 \to [e]_2$, and so on.

Let us illustrate this by considering how this applies to the ordinary static higher order translation of an indefinite noun phrase. In extensional Montague grammar 'a man' translates as:

$$\lambda P \exists x (\text{man } x \wedge Px).$$

In ITL this becomes:

$$\lambda P \lambda c \lambda c'. \exists x (\text{man } x \wedge P|c|(c\hat{\ }x)c').$$

Here P is a variable of type $\iota \to [e] \to [e] \to t$, while c, c' are variables of type $[e]$ (variables ranging over contexts). The translation as a whole has type $(\iota \to [e] \to [e]) \to [e] \to [e] \to t$. The P variable marks the slot for the VP interpretation. $|c|$ gives the length of the input context, i.e. the position of the next available slot. Note that $c\hat{\ }x[|c|] = x$.

Note that the translation of the indefinite NP does not introduce an anaphoric index, as would be the case for example in DMG [Groenendijk and Stokhof, 1990]. Instead, an anaphoric index i is picked up from the input context. Also, the context is not reset but incremented: context update is not destructive, whereas it is in DPL and DMG.

In order to obtain a proper dynamic higher order system we first define the appropriate dynamic operations in typed logic. Assume φ and ψ have the type of context transitions, i.e. type $[e] \rightarrow [e] \rightarrow t$, and that c, c', c'' have type $[e]$. Note that $\hat{}$ is an operation of type $[e] \rightarrow e \rightarrow [e]$.

$$
\begin{array}{rcl}
\mathcal{E} & := & \lambda cc'.\exists x(c\hat{}x = c') \\
\sim\varphi & := & \lambda cc'.(c = c' \wedge \neg\exists c''\varphi cc'') \\
\varphi \,;\, \psi & := & \lambda cc'.\exists c''(\varphi cc'' \wedge \psi c''c')
\end{array}
$$

These operations encode the semantics for incremental quantification, dynamic incremental negation and dynamic incremental conjunction in typed logic. Dynamic implication, $\Rightarrow$, is defined in the usual way.

We have to assume that the lexical meanings of CNs, VPs are given as one-place predicates (type $e \rightarrow t$) and those of TVs as two place predicates (type $e \rightarrow e \rightarrow t$). We therefore define blow-up operations for lifting one-placed and two-placed predicates to the dynamic level. Let A be an expression of type $e \rightarrow t$, and B an expression of type $e \rightarrow e \rightarrow t$; we use c, c' as variables of type $[e]$, and j, j' as variables of type ι, and we employ postfix notation for the lifting operations:

$$
\begin{array}{rcl}
A^\circ & := & \lambda jcc'.(c = c' \wedge Ac[j]) \\
B^\bullet & := & \lambda jj'cc'.(c = c' \wedge Bc[j]c[j'])
\end{array}
$$

The encodings of the dynamic operations in typed logic and the blow-up operations for one- and two-placed predicates are employed in the semantic specification of the following simple fragment. The semantic specifications employ variables P, Q of type $\iota \rightarrow [e] \rightarrow [e] \rightarrow t$, variables j, j' of type ι, and variables c, c' of type $[e]$.

We also define an operation $! : (\iota \rightarrow [e] \rightarrow [e] \rightarrow t) \rightarrow [e] \rightarrow [e] \rightarrow t$ (from lifted one-place predicates to context transformers), to express that a lifted predicate applies to a single individual in a given context. Assuming P to be an expression of type $(\iota \rightarrow [e] \rightarrow [e] \rightarrow t)$ (a lifted predicate), and c, c' to be of type $[e]$ (contexts), we define $!$ as follows:

$$
!P \quad := \quad \lambda cc'.\exists x\forall y(P|c|(c\hat{}y)c' \leftrightarrow x = y).
$$

This expresses that P is the lift of a predicate that applies to a single individual.

As said above, we assume that pronouns are the only NPs that carry indices; pronoun reference resolution is not treated. Appropriate indices for proper names are extracted from the current context. In the rules, X refers to the semantics of the left-hand side of the syntax rule, to be defined in terms of the semantic translations of the members of the right-hand side of the syntax rule. X_i refers to the semantics of the i-th member of the right-hand side of the syntax rule.

S	::=	**NP VP**	X	::=	$(X_1 X_2)$
S	::=	*if* **S S**	X	::=	$X_2 \Rightarrow X_3$
S	::=	**S** . **S**	X	::=	$X_1 \; ; \; X_3$
NP	::=	*Mary*	X	::=	$\lambda Pcc'.\exists j(c[j] = m \wedge Pjcc')$
NP	::=	PRO^k	X	::=	$\lambda Pcc'.(Pkcc')$
NP	::=	**DET CN**	X	::=	$(X_1 X_2)$
NP	::=	**DET RCN**	X	::=	$(X_1 X_2)$
DET	::=	*every*	X	::=	$\lambda PQc.(\sim(\mathcal{E} \; ; \; P|c| \; ; \; \sim Q|c|))c$
DET	::=	*some*	X	::=	$\lambda PQc.(\mathcal{E} \; ; \; P|c| \; ; \; Q|c|)c$
DET	::=	*no*	X	::=	$\lambda PQc.(\sim(\mathcal{E} \; ; \; P|c| \; ; \; Q|c|))c$
DET	::=	*the*	X	::=	$\lambda PQc.(!P \; ; \; \mathcal{E} \; ; \; P|c| \; ; \; Q|c|)c$
CN	::=	*man*	X	::=	M°
CN	::=	*woman*	X	::=	W°
CN	::=	*boy*	X	::=	B°
RCN	::=	**CN** *that* **VP**	X	::=	$\lambda j.((X_1 \; j) \; ; \; (X_3 \; j))$
RCN	::=	**CN** *that* **NP TV**	X	::=	$\lambda j.((X_1 \; j) \; ; \; (X_3(\lambda j'.((X_4 \; j')j))))$
VP	::=	*laughed*	X	::=	L°
VP	::=	*smiled*	X	::=	S°
VP	::=	**TV NP**	X	::=	$\lambda j.(X_2 \; ; \; \lambda j'.((X_1 \; j')j))$
TV	::=	*loved*	X	::=	$L'^{\bullet}$
TV	::=	*respected*	X	::=	$R^{\bullet}$

Note that determiners do not carry indices, the appropriate index is provided by the length of the input context. It is assumed that all proper names are linked to anchored elements in context. In fact, the anchoring mechanism has been greatly improved by the switch to the incremental, non-destructive approach, for the incremental nature of the context update mechanism ensures that no anchored elements can ever be overwritten.

The following very simple example illustrates how the system deals with cross-sentential anaphora:

2. Some man smiled. He laughed.

The structures assigned to the sentences making up this sequence by the system are the following:

(2) a.   b.
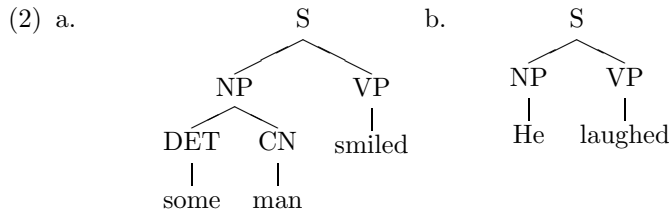
Note that the tree for the second sentence in sequence 2 actually can not be produced by the rules given above: those rules assume that surface pronouns are generated as indexed abstract PRO-elements, as in:

(2) b'.

$$
\begin{array}{c}
\text{S} \\
\diagup \quad \diagdown \\
\text{NP} \quad \text{VP} \\
| \quad\quad | \\
\text{PRO}^5 \quad \text{laughed}
\end{array}
$$
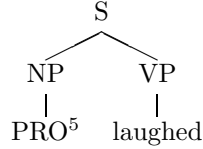
Translations of the two sentences are derived in a compositional fashion. For example, the NP 'Some man' translates as:

$$(\lambda PQc.(\mathcal{E} \ ; \ P|c| \ ; \ Q|c|)c)(M^\circ)$$

With S° as the translation of the VP 'smile', the sentence, 'Some man smiled' then receives the following translation:

$$\mathcal{E} \ ; \ M^\circ|c| \ ; \ S^\circ|c|$$

This is an expression of type $[e] \to [e] \to t$ and denotes a relation between contexts. It takes a context and extends it with an object that is both a man and that smiles, as is evident if we reduce it as follows, using the definitions of the dynamic existential quantifier, the dynamic conjunction and the lift operation.

We first rewrite $\mathcal{E}$:

$$(\lambda cc'.\exists x(c\hat{\ }x = c') \ ; \ M^\circ|c| \ ; \ S^\circ|c|$$

and next the lifted predicates:

$$(\lambda cc'.\exists x(c\hat{\ }x = c') \ ; \ (\lambda cc'(c = c' \wedge Mc[|c|]) \ ; \ (\lambda cc'(c = c' \wedge Sc[|c|])$$

The indefinite determiner extends the context with a new object. The other clauses test the last element of the current context for the properties M and S, respectively.

Rewriting the dynamic conjunction shows how the element introduced by the indefinite determiner is passed on to the other clauses. The first two clauses become:

$$\lambda cc'.\exists c''((\lambda cc'.\exists x(c\hat{\ }x = c')cc'' \wedge (\lambda cc'(c = c' \wedge Mc[|c|])c''c')$$

which after some reduction becomes:

$$\lambda cc'.\exists x(c\hat{\ }x = c' \wedge Mx)$$

Rewriting the second occurrence of the dynamic conjunction gives the following reduced translation for the first sentence:

$$\lambda cc'.\exists x(c\hat{\ }x = c' \wedge Mx \wedge Sx)$$

For the second sentence we get:

$$\lambda cc'.(L^\circ 5cc')$$

which reduces to

$$\lambda cc'.(c = c' \land Lc[5])$$

and for the sequence as a whole we get:

$$\lambda cc'.\exists x(\hat{c}x = c' \land Mx \land Sx) \; ; \; \lambda cc'.(c = c' \land Lc[5])$$

which reduces to:

$$\lambda cc'.\exists x(\hat{c}x = c' \land Mx \land Sx \land Lc[5])$$

Note that we obtain the reading in which the pronoun in the second sentence of 2 refers back to the man introduced in the first sentence only if the index of the PRO-element is suitably chosen. This means that this approach relies on a separate pronoun resolution component in the grammar.

7.3 Update Semantics

In section we illustrate the use of dynamic logic in another area of natural language semantics, one that is concerned with epistemic concerns, modal expressions and with the interaction between issues that are strictly semantic and phenomena that are of a pragmatic nature, i.e. that pertain to the use of language in information exchange.

The gist of the dynamic approach to natural language meaning is captured in the slogan 'Meaning is context-change potential'. In the case of a theory such as DPL, the context consists of assignments of objects satisfying certain properties to variables. In that case, context-change means change of assignments. Such changes are brought about typically by referring expressions such as proper names or temporal expressions and by quantificational expressions such as noun phrases or tense operators. All other expressions are tests. In the case of DRT a different notion of context is used, viz. that of a discourse representation that contains discourse referents satisfying certain properties that point to objects satisfying corresponding properties: here context change is change of the discourse representation. With respect to empirical coverage that does not make a difference, again it is referential and quantificational expressions that change the context, other expressions are treated as parts of conditions.

In epistemic systems, context is yet another type of object, viz., information, modelled by a set of possible worlds or possible situations or propositions. The pioneering work in this area is that of Stalnaker (cf., among others, [Stalnaker, 1979; 1974]). Stalnaker focused on the context as the 'common ground', i.e. the information that is available by all speech participants and that is maintained as it gets updated during a linguistic information exchange. This common ground can be characterised as a set of worlds, viz., those worlds which are compatible with the shared information, or, alternatively, as a set of propositions. A linguistic exchange then consists of utterances that shift the context, by updating the common ground, or that test whether something holds in the context. Each utterance represents

a particular way of updating or testing the common ground, and this update is conceived as the meaning of the utterance in question.

Within such an approach, sentences that are tests in DPL or conditions in DRT in most cases do have an effect on the context, and thus are treated dynamically. A simple subject-predicate sentence such as 'John is at home' updates the common ground with the information that John has the property of walking, and conjunctions are ordered updates. Examples of exceptions, i.e., sentences that do not update the context but test it, are modal sentences, such as 'John might be at home', and 'John must be at home'. These do not add new information, but check whether the existing common ground satisfies a requirement: that it is possible that John is at home, and that it not possible that John is not at home, respectively.

Another type of linguistic construction that can be treated in this fashion concerns presuppositions. A sentence carrying a presupposition typically tests the common ground for the presence of the presupposed information, besides updating it with new information And yet another example is presented by conditionals: the sentence 'If John is at home, Mary is there, too' tests the context by checking whether updating with the antecedent 'John is at home' leads to a context in which 'Mary is at home' holds.

Of particular interest is what consequences obtain if a test or an update fails. In the case of a presupposition failing because the information is not present, but is consistent with the common ground, the presupposition is often said to be 'accommodated', i.e. an implicit update takes place [Beaver, 1997]. In other cases, e.g., the failure of a test such as 'John might be at home', or of a straightforward update such as 'John is at home', the context needs to be down-dated, i.e. revised. This is the area of belief revision [Gärdenfors, 1984] another aspect of the dynamics of information exchange.

System

Update semantics was originally devised as a way of dealing with the semantics of modal expressions such as 'might' and 'must' [Veltman, 1984]. These expressions have a specifically epistemic meaning, which makes implicit reference to the information states of speaker and hearer. Other uses of update semantics are, among others, in accounts of conditionals [Veltman, 1986], defaults [Veltman, 1996], presuppositions [Beaver, 1997], [Zeevat, 1992], and other issues involving information exchange.

Here we present a core system that forms the basis of many variations in the literature.

Let P be a set of atomic formulae. The language is that of propositional logic, with an additional operator $\Diamond$. Assume p ranges over P.

$$\begin{aligned} \varphi \quad &::= \quad p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Diamond\varphi \\ \varphi' \quad &::= \quad \Diamond\varphi \end{aligned}$$

The other connectives are defined in the usual fashion.

A model $\mathbf{M}$ consists of a set of possible worlds W and in interpretation function $V : P \rightarrow \mathcal{P}(W)$. Information states s are subsets of W, with $\emptyset$ the absurd information state, W the state of no information, and singletons $\{w_i\}$ states of maximal information.

The semantics takes the form of a definition of '$s[\varphi]_{\mathbf{M}}$', i.e. the result of updating an information state s in $\mathbf{M}$ with (the information conveyed by) φ:

$$
\begin{aligned}
s[p]_{\mathbf{M}} &= s \cap \{s \in S \mid s \in V_{\mathbf{M}}(p)\} \\
s[\neg\varphi]_{\mathbf{M}} &= s \setminus s[\varphi]_{\mathbf{M}} \\
s[\varphi_1 \wedge \varphi_2]_{\mathbf{M}} &= s[\varphi_1]_{\mathbf{M}}[\varphi_2]_{\mathbf{M}} \\
s[\Diamond\varphi]_{\mathbf{M}} &= \begin{cases} s & \text{if } s[\varphi]_{\mathbf{M}} \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}
\end{aligned}
$$

An atomic formula updates s with the information it conveys; a negation $\neg\varphi$ deletes those worlds in which the information conveyed by φ holds from s; conjunction is a sequential update with the conjuncts. The modal $\Diamond\varphi$ is a test: it returns the original state if an update with φ is possible, the absurd state otherwise.

This system analyses a special case of public announcement logic [Plaza, 1989; Gerbrandy, 1999b], where the knowledge of a single agent is modelled. The model $\mathbf{M}$ above can be viewed as an S5 model with a universal accessibility relation [van Eijck and de Vries, 1995]. Updating with a propositional formula F has the effect of announcing F to the agent, i.e. updating with action model
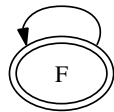
in the sense of [Baltag and Moss, 2004].Updating with a modal formula $\Diamond F$ boils down to updating with the following action model:
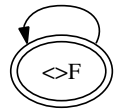
The notion of 'acceptance in $\mathbf{M}$, s' is defined as follows:

$$s \models_{\mathbf{M}} \varphi \text{ iff } s \subseteq s[\varphi]_{\mathbf{M}}$$

Validity can be defined in a number of ways; the most common one is as follows:

$$\varphi_1 \ldots \varphi_n \models \psi \text{ iff for all } \mathbf{M}, s \colon s[\varphi_1]_{\mathbf{M}} \psi \ldots [\varphi_n]_{\mathbf{M}} \models \psi$$

i.e. every state that accepts the premises, accepts the conclusion.

This system is eliminative ($s[\varphi]_{\mathbf{M}} \subseteq \varphi$); not distributive ($s \subseteq s' \not\Rightarrow s[\varphi] \subseteq s'[\varphi]$); neither right- nor left-monotone; and conjunction is not commutative. A complete sequent calculus can be found in [Groeneveld, 1995, chapter 3].

Characteristic examples

A characteristic example, that illustrates the non-commutativity of conjunction, involves the $\diamondsuit$-operator. If we read it as the formal counterpart of the modal expression 'might' (in its epistemic meaning), and represent discourse sequencing as conjunction, we can explain the difference between the following two sentences:

a. Somebody is knocking at the door ... It might be John ... It is Mary

b. Somebody is knocking at the door ... It is Mary ... *It might be John

In the first sequence the second sentence 'It might be John' tests the state (that contains the information that somebody is at the door, due to the update with the first sentence) for the possibility that the person knocking is John. If that succeeds, it is only confirmed that this is a possibility. The subsequent update with the information that in fact it is Mary, is consistent with that. In the second sequence the information that it is Mary is added before the test takes place, resulting in its failure, which explains the odd status of this sequence.

The failure of right- en left-monotonicity is also due to the $\diamondsuit$-operator:

$$\diamondsuit\neg\varphi \models \diamondsuit\neg\varphi \quad \text{but} \quad \diamondsuit\neg\varphi, \varphi \not\models \diamondsuit\neg\varphi$$
$$\models \diamondsuit\varphi \quad \text{but} \quad \neg\varphi \not\models \diamondsuit\varphi$$

Another instantiation of the ideas behind update semantics is provided by conditionals. Many aspects of conditionals in natural language can be captured in an update framework, by keeping in mind the 'modal' nature of the conditional construction:

$$s[\varphi_1 \rightarrow \varphi_2]_{\mathbf{M}} \quad = \quad \{i \in s \mid \text{ if } i \in s[\varphi_1]_{\mathbf{M}} \text{ then } i \in s[\varphi_1]_{\mathbf{M}}[\varphi_2]_{\mathbf{M}}\}$$

The update effect of a condition thus is to retain those possibilities in a given state s such that updating them with the antecedent allows a subsequent update with the consequent.

Applications of update semantics can be found in a variety of areas, such as deontic modality [van der Torre and Tan, 1998]; interrogatives [Groenendijk, 1999]; imperatives [Zarnic, 2002; Lascarides and Asher, 2003]; counterfactuals and other irrealis-constructions [Veltman, 2005].

7.4   Combining dynamic and update semantics

The dynamic semantics used in systems such as DPL and DRT can be combined with an update type of semantics as just defined. Various proposals exists (cf.,

e.g., [Groenendijk *et al.*, 1995; Dekker, 1996]). The idea is to put the semantics for quantified formulae in an update format. In [Groenendijk *et al.*, 1996] this is done as follows.

Existential quantifiers introduce new kind of objects, so-called 'pegs', modelled by the natural numbers. This notion was first introduced by Vermeulen, cf., [Vermeulen, 1995]. A referent system r is a function from a finite set of variables to pegs. An existential quantifier $\exists x$ add its variable x, introduces the next peg and associates x with that peg. So, if r is a referent system with domain v and range of pegs n, then $r[x/n]$ is the referent system r' which is like r except that its domain is $v \cup \{x\}$ its range is $N + 1$ and $r'(x) = n$. Let r and r' be two referent systems with domain v and v', and range n and n', respectively. Then we say that r' is an *extension* of r, $r \leq r'$, iff $v \subseteq v'$; $n \leq n'$; if $x \in v$ then $r(x) = r'(x)$ or $n \leq r'(x)$; if $x \notin v$ and $x \in v'$ then $n \leq r'(x)$.

States s are sets of triples i consisting of the same referent system r, an assignment g and a world w. So states contain information about both the world (via the possible world parameter) as well as the discourse (via the referent system). Growth of information is then twofold as well: via the elimination of possibilities, and via extension of the referent system. First we introduce:

$$
\begin{aligned}
i[x/d] &= \langle r[x/n], g[n/d], w \rangle \\
s[x/d] &= \{ i[x/d] \mid i \in s \}
\end{aligned}
$$

and then we define these two notions of information growth as follows. Let $i, i' \in I, i = \langle r, g, w \rangle$ and $i' = \langle r', g', w' \rangle$, and $s, s' \in S$:

$$
\begin{aligned}
i' \leq i' &\quad \text{iff} \quad r \leq r', g \subseteq g', w = w' \\
s \leq s' &\quad \text{iff} \quad \text{for all } i' \in s' \colon \text{ there exists an } i \in s \colon i \leq i'
\end{aligned}
$$

Finally, we define the update semantics for existentially quantified formulae $\exists x \varphi$ as follows (the other clauses are merely repetitions of the above):

$$
s[\exists x \varphi]_{\mathbf{M}} \quad = \quad \cup_{d \in D_{\mathbf{M}}} (s[x/d][\varphi]_{\mathbf{M}})
$$

This defines the update effect of $\exists x \varphi$ point-wise on the objects in the domain: the referent system of the state s is updated by adding a peg, the variable is associated with the peg, and an object d is selected and assigned to the peg; then the resulting state $s[x/d]$ is updated with φ; this procedure is repeated for every object in the domain; the results are collected and together make up the new state $s[\exists x \varphi]$.

The resulting system is capable of treating complex cases concerning the interaction of quantifiers and modalities. For example it can be used to show that whereas $\exists x P x \wedge \Diamond \forall y \neg P y$ is not consistent, $\exists x P x \wedge \forall y \Diamond \neg P y$ is: if we know that something has the property P this ipso facto rules out the possibility that no-one has that property, but it does not rule out the possibility that we are uninformed about the identity of this P. For other examples, involving also identity we refer the reader to [Groenendijk *et al.*, 1996] and [Aloni, 2002].

8   CONCLUDING REMARKS

The overview of dynamic logics and their applications presented in this paper has focused on a number of core systems (Floyd/Hoare logic, PDL, epistemic PDL, QDL, DPL), and a number of central applications: program analysis, tree description, analysis of communication, semantics of natural language. References to other applications were thrown in as an incentive to the reader for further exploration.

The field of dynamic logic, including its applications in various domains, is still developing. Dynamic logic started out as a way of studying various aspects of computation, mainly in traditional computational settings, with a focus on sequential transformational programs. When theoretical computer science broadened to encompass the theory of reactive systems and concurrency, dynamic logic evolved by developing systems that could handle these too (branching time logics and μ calculus). Thus, the core concepts of dynamic logic have proved to be applicable in a wide range of settings, allowing formalisation of a great diversity of concepts and phenomena.

In certain areas, such as natural language semantics, the use of dynamic concepts initially arose independently, and it was only subsequently that these notions were embedded in dynamic logic. This have given rise to interesting interactions, that are still being actively pursued.

The application to communicative action stays somewhat closer to the original motivation for the development of dynamic logic. Here the use of dynamic logic ties in with an existing tradition of using modal logic in the analysis of communication protocols [Halpern *et al.*, 1995]. Also in the analysis of various other phenomena that are concerned with interactions between individuals and with properties of the collectives (groups, societies) that they form, concepts of dynamic logic play a role, as is testified by work done on, for example, collective decisions (cf., [Pauly and Parikh, 2003] on game logic as an extension of propositional dynamic logic).

As more aspects of the ways in which human beings interact are brought into the picture, concepts like perception, causality, justification and intention appear. Here insights from the philosophy of action and from game theory must augment the tool set from dynamic logic, thus creating an exciting amalgam of logic, theoretical computer science, philosophy and game theory. Whatever the future holds in store for this area, it seems more than likely that concepts and results from dynamic logic will continue to play a major role in its development.

BIBLIOGRAPHY

[Aloni, 2002] M. Aloni. Conceptual covers in dynamic semantics. In D. Barker-Plummer, D. Beaver, J. van Benthem, and P. Scotto di Luzio, editors, *Words, Proofs and Diagrams*, pages 167–190. CSLI, Stanford, CA, 2002.

[Apt, 1981] K.R. Apt. Ten years of Hoare's logic: A survey — Part I. *ACM Transactions on Programming Languages and Systems*, 3(4): 431–483, 1981.

[Apt and Bezem, 1999] K.R. Apt and M. Bezem. Formulas as programs. In K.R. Apt, V. Marek, M. Truszczyski, and D.S. Warren, editors, *The Logic Programming Paradigm: A 25 Year Perspective*, pages 75–107. Springer-Verlag, Berlin, 1999. Paper available as `http://xxx.lanl.gov/abs/cs.LO/9811017`.

[Areces *et al.*, 2001] C. Areces, P. Blackburn, and M. Marx. Hybrid logics: Characterization, interpolation and complexity. *Journal of Symbolic Logic*, 66:977–1010, 2001.

[Baltag, 2002] A. Baltag. A logic for suspicious players: Epistemic action and belief-updates in games. *Bulletin of Economic Research*, 54(1): 1–45, 2002.

[Baltag and Moss, 2004] A. Baltag and L.S. Moss. Logics for epistemic programs. *Synthese*, 139(2): 165–224, 2004.

[Baltag *et al.*, 1999] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. Technical Report SEN-R9922, CWI, Amsterdam, 1999.

[Baltag *et al.*, 2003] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. Technical Report, Dept of Cognitive Science, Indiana University and the Department of Computing, Oxford University, 2003.

[Barwise, 1987] J. Barwise. Noun phrases, generalized quantifiers and anaphora. In P. Gärdenfors, editor, *Generalized Quantifiers: Linguistic and Logical Approaches*, pages 1–30. Reidel, Dordrecht, 1987.

[Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Cambridge, MA, 1983.

[Beaver, 1997] D. I. Beaver. Presupposition. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 939–1008. Elsevier–MIT Press, Amsterdam–Cambridge, MA, 1997.

[Belnap *et al.*, 2001] N. D. Belnap, M. Perloff, and M. Xu. *Facing the Future*. Oxford University Press, Oxford, 2001.

[Blackburn *et al.*, 1993] P. Blackburn, C. Gardent, and W. Meyer-Viol. Talking about trees. In S. Krauwer, M. Moortgat and L. des Tombe, editors, *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics, Utrecht*, pages 21–29. Association for Computational Linguistics, 1993.

[Blackburn *et al.*, 2001] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Vol. 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 2001.

[Boole, 1854] G. Boole. *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*. Dover, New York. First edition, 1854. Reprint edition.

[Butler and Mathieu, 2004] A. Butler and E. Mathieu. *The Syntax and Semantics of Split Constructions*. Palgrave, London, 2004.

[Chang and Keisler, 1973] C.C. Chang and H.J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.

[Chierchia, 1992] G. Chierchia. Anaphora and dynamic binding. *Linguistics and Philosophy*, 15(2): 111–183, 1992.

[Chierchia, 1995] G. Chierchia. *Dynamics of Meaning*. University of Chicago Press, Chicago, IL, 1995.

[Clarke and Emerson, 1982] E.E. Clarke and E.A. Emerson. Synthesis of synchronisation skeletons for branching time temporal logic. In D. Kozen, editor, *Logic of Programs Workshop*, number 131 in *LNCS*. Springer-Verlag, Berlin and New York, 1982.

[Clarke *et al.*, 1993] E.M. Clarke, O. Grumberg, and D. Long. Verification tools for finite-state concurrent systems. In J.W. de Bakker, W.-P. de Roever, and G. Rozenberg, editors, *A Decade of Concurrency*, pages 124–175. Number 803 of *LNCS*. Springer-Verlag, Berlin and New York, 1994.

[Cook, 1978] S.A. Cook. Soundness and completeness of an axiom system for program verification. *SIAM Journal of Computing*, 7(1): 70–90, 1978.

[Davey and Priestley, 2002] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order (Second Edition)*. Cambridge University Press, Cambridge, 2002. First edition, 1990.

[Dekker, 1994] P. Dekker. Predicate logic with anaphora (seven inch version). In L. Santelmann and M. Harvey, editors, *Proceedings of the Fourth SALT Conference, Cornell Working Papers in Linguistics*, Cornell University, 1988, pages 79–95. DMLL Publications, Ithaca, NY, 1994.

[Dekker, 1996] P. Dekker. The values of variables in dynamic semantics. *Linguistics and Philosophy*, 19:211–257, 1996.

[Dekker, 2002] P. Dekker. Meaning and use of indefinite expressions. *Journal of Logic, Language and Information*, 11:141–194, 2002.

[Fischer and Ladner, 1979] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2): 194–211, 1979.

[Floyd, 1967] R.W. Floyd. Assigning meanings to programs. In J.T. Schwartz, editor, *Proceedings AMS Symposium Applied Mathematics, Providence, RI, 1967*, volume 19, pages 19–31. American Mathematical Society, Providence, RI, 1967.

[Gärdenfors, 1984] P. Gärdenfors. The dynamics of belief as a basis for logic. *British Journal for the Philosophy of Science*, 35:1–10, 1984.

[Gerbrandy, 1999a] J. Gerbrandy. *Bisimulations on planet Kripke*. PhD thesis, ILLC, 1999.

[Gerbrandy, 1999b] J. Gerbrandy. Dynamic epistemic logic. In L.S. Moss, J. Ginzburg and M. de Rijke, editors, *Logic, Language and Information, Vol. 2*, pages 67–84. CSLI Publications, Stanford, CA, 1999.

[Gochet, 2002] P. Gochet. The dynamic turn in twentieth century logic. *Synthese*, 130:175–84, 2002.

[Gochet and Gribomont, 2005] P. Gochet and P. Gribomont. Epistemic logic. In Dov Gabbay and John Woods, editors, *Logic and the Modalities in the Twentieth Century*, pages 101–175. Volume 6 of *Handbook of the History of Logic*. Elsevier, Amsterdam, 2005. This volume.

[Goldblatt, 1982] R. Goldblatt. *Axiomatising the Logic of Computer Programming*. Springer-Verlag, Berlin and New York, 1982.

[Goldblatt, 1992/1987] R. Goldblatt. *Logics of Time and Computation, Second Edition, Revised and Expanded*. Volume 7 of *CSLI Lecture Notes*. CSLI, Stanford, CA, 1992. First edition 1987.

[Gordon, 1988] M.J.C. Gordon. *Programming Language Theory and Its Implementation: Applicative and Imperative Paradigms*. Prentice Hall, 1988.

[Gries, 1981] D. Gries. *The Science of Programming*. Springer, Berlin, 1981.

[Groenendijk, 1999] J. Groenendijk. The logic of interrogation: Classical version. In T. Matthews and D. Strolovitch, editors, *Proceedings of SALT IX, Santa Cruz*, pages 109–126. Cornell University Press, Ithaca, NY, 1999.

[Groenendijk and Stokhof, 1990] J. Groenendijk and M. Stokhof. Dynamic Montague grammar. In L. Kálmán and L. Pólos, editors, pages 3–48. *Papers from The Second Symposium on Logic and Language*. Akadémiai Kiadó, Budapest, 1990.

[Groenendijk and Stokhof, 1991] J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.

[Groenendijk and Stokhof, 2000] J. Groenendijk and M. Stokhof. Meaning in motion. In K. von Heusinger and U. Egli, editors, *Reference and Anaphoric Relations*, pages 47–76. Kluwer, Dordrecht, 2000.

[Groenendijk et al., 1995] J. Groenendijk, M. Stokhof, and F. Veltman. This might be it. In D. Westerståhl and J. Seligman editors, *Language, Logic and Computation: The 1994 Moraga Proceedings*, pages 255–270. CSLI, Stanford, CA, 1995.

[Groenendijk et al., 1996] J. Groenendijk, M. Stokhof, and F. Veltman. Coreference and modality. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–213. Blackwell, Oxford, 1996.

[Groeneveld, 1995] W. Groeneveld. *Logical Investigations into Dynamic Semantics*. PhD thesis, ILLC, Amsterdam, 1995.

[Guy, 1981] R. Guy. *Unsolved Problems in Number Theory*. Springer, New York, 1981.

[Halpern et al., 1995] J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, MA, 1995.

[Harel, 1979] D. Harel. *First-Order Dynamic Logic*. Number 68 of *Lecture Notes in Computer Science*. Springer, Berlin, 1979.

[Harel, 1984] D. Harel. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II*, pages 497–604. Reidel, Dordrecht, 1984.

[Harel et al., 2000] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic. Foundations of Computing*. MIT Press, Cambridge, MA, 2000.

[Heguiabehere, 2001] J. Heguiabehere. Pre- and postcondition reasoning in DFOL. Manuscript. ILLC, 2001.

[Heim, 1982] I. Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982.

[Heim, 1983] I. Heim. File-change semantics and the familiarity theory of definiteness. In R. Bäuerle, C. Schwarze, and A, von Stechow, editors, *Meaning, Use, and Interpretation of Language*, pages 164–189. de Gruyter, Berlin, 1983.

[Hindley, 1997] J. Roger Hindley. *Basic Simple Type Theory*. Cambridge University Press, 1997.

[Hintikka, 1962] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, NY, 1962.

[Hintikka, 1983] J. Hintikka. *The Game of Language*. Reidel, Dordrecht, 1983.

[Hoare, 1969] C.A.R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, 12(10): 567–580, 583, 1969.

[Hollenberg, 1996] J.J. Hollenberg. Equational axioms of test algebra. *Logic Group Preprint Series* 172, Department of Philosophy, Utrecht University, 1996.

[Hollenberg, 1997] M. Hollenberg. An equational axiomatisation of dynamic negation and relational composition. *Journal of Logic, Language and Information*, 6:381–401, 1997.

[Hollenberg, 1998] M. Hollenberg. *Logic and Bisimulation*. PhD thesis, Utrecht University, 1998.

[Hollenberg and Visser, 1997] M. Hollenberg and A. Visser. Dynamic negation, the one and only. *Logic Group Preprint Series* 179, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, 1997. To appear in *Journal of Logic, Language, and Information*.

[Huth and Ryan, 2000] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and Reasoning about Systems*. Cambridge University Press, Cambridge and New York, 2000.

[Janin and Walukiewicz, 1996] D. Janin and I. Walukiewicz. On the expressive completeness of the propositional mu-calculus with respect to monadic second order logic. In U. Montanari and V. Sassone, editors, *CONCUR '96: Concurrency Theory, Proceedings of the 7th International Conference on Concurrency Theory*, pages 263–277. Volume 1119 of LNCS. Springer, Berlin and New York, 1996.

[Jansche, 1998] M. Jansche. Dynamic Montague Grammar lite. Dept of Linguistics, Ohio State University, November 1998.

[Jones, 2003] S. Peyton Jones, editor. *Haskell 98 Language and Libraries; The Revised Report*. Cambridge University Press, 2003.

[Kamp, 1981] H. Kamp. A theory of truth and semantic representation. In J. Groenendijk et al., editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam, 1981.

[Kamp, 1990] H. Kamp. Comments on: J. Groenendijk & M. Stokhof, Dynamic Predicate Logic. In J. van Benthem, editor, *Partial and Dynamic Semantics, Part I. Dyana Deliverable 2.1.A*, pages 109–31. ILLC, Universiteit van Amsterdam, 1990.

[Kamp and Reyle, 1993] H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.

[Kanazawa, 1994] M. Kanazawa. Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and Philosophy*, 17:109–158, 1994.

[Karttunen, 1976] L. Karttunen. Discourse referents. In J. McCawley, editor, *Syntax and Semantics 7*, pages 363–385. Academic Press, 1976.

[Kohlhase et al., 1996] M. Kohlhase, S. Kuschert, and M. Pinkal. A type-theoretic semantics for λ-DRT. In P. Dekker and M. Stokhof, editors, *Proceedings of the Tenth Amsterdam Colloquium*, pages 479–498. ILLC, Amsterdam, 1996.

[Kooi and van Benthem, 2004] B. Kooi and J. van Benthem. Reduction axioms for epistemic actions. Manuscript. Groningen/Amsterdam, 2004.

[Kozen, 1983] D. Kozen. Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

[Kracht, 1995] M. Kracht. Syntactic codes and grammar refinement. *Journal of Logic, Language and Information*, 4:41–60, 1995.

[Kuschert, 2000] S. Kuschert. *Dynamic Meaning and Accommodation*. PhD thesis, Universität des Saarlandes, 2000. Thesis defended in 1999.

[Lascarides and Asher, 2003] A. Lascarides and N. Asher. Imperatives in dialogue. In K. Peter, editor, *Perspectives on Dialogue in the New Millenium*. John Benjamins, Amsterdam and Philadelphia, 2003.

[Marx, 2004] M. Marx. XCPath, the first order complete XPath dialect. *Proceedings of PODS*, pages 13–22. ACM SIGMOD, 2004.

[Milner, 1978] R. Milner. A theory of type polymorphism in programming. *Journal of Computer and System Sciences*, 17, 1978.

[Montague, 1970] R. Montague. Universal grammar. *Theoria*, 36:373–98, 1970.

[Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English. In J. Hintikka, editor, *Approaches to Natural Language*, pages 221–242. Reidel, Dordrecht, 1973.

[Montague, 1974] R. Montague. English as a formal language. In R.H. Thomason, editor, *Formal Philosophy; Selected Papers of Richard Montague*, pages 188–221. Yale University Press, New Haven, CN and London, 1974.

[Muskens, 1994] R. Muskens. A compositional discourse representation theory. In P. Dekker and M. Stokhof, editors, *Proceedings 9th Amsterdam Colloquium*, pages 467–486. ILLC, Amsterdam, 1994.

[Muskens, 1995] R. Muskens. Tense and the logic of change. In U. Egli, P. E. Pause, C. Schwarze, A. von Stechow, and G. Wienold, editors, *Lexical Knowledge in the Organization of Language*, pages 147–183. Benjamins, Amsterdam, 1995.

[Muskens, 1996] R. Muskens. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19:143–186, 1996.

[Nielson and Nielson, 1992] H.R. Nielson and F. Nielson. *Semantics with Applications: A Formal Introduction*. John Wiley, Chichester and New York, 1992.

[Parikh, 1978] R. Parikh. The completeness of propositional dynamic logic. In J. Winkowski, editor, *Proceedings of the Seventh Symposium on the Mathematical Foundations of Computer Science, 1978*, pages 403–415. Springer-Verlag, Berlin and New York, 1978.

[Partee, 1997] B. Partee. Montague grammar. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Linguistics*, pages 5–92. Elsevier, Amsterdam, 1997.

[Passy and Tinchev, 1991] S. Passy and T. Tinchev. An essay in combinatory dynamic logic. *Inf. Comput.*, 93(2): 263–332, 1991.

[Pauly and Parikh, 2003] M. Pauly and R. Parikh. Game logic: An overview. *Studia Logica*, 75(2): 165–182, 2003.

[Plaza, 1989] J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and W. Ras Zbigniew, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems held in Charlotte, NC*, pages 201–216. North-Holland, New York, 1989.

[Plotkin, 1981] G. Plotkin. *Structural Operational Semantics*. Aarhus University, Denmark, 1981. Lecture notes.

[Pnueli, 1981] A. Pnueli. A temporal logic of programs. *Theoretical Computer Science*, 13:45–60, 1981.

[Pratt, 1976] V. Pratt. Semantical considerations on Floyd–Hoare logic. In P. Abrahams, R. Lipton, and S. Bourne, editors, *Proceedings 17th IEEE Symposium on Foundations of Computer Science*, pages 109–121. IEEE Computer Science Society Press, Long Beach, CA, 1976.

[Pratt, 1978] V. Pratt. A practical decision method for propositional dynamic logic (Preliminary Report). In R. Lipton, W. Burkhard, W. Savitch, E. P. Friedman, and A. Aho, editors, *Proceedings 10th Annual ACM Symposium On Theory of Computation, May 01–03, 1978, San Diego, CA*, pages 326–337. ACM Press, New York, 1978.

[Pratt, 1980] V. Pratt. Application of modal logic to programming. *Studia Logica*, 39:257–274, 1980.

[Prior, 1957] A. Prior. *Time and Modality*. Oxford University Press, 1957.

[Prior, 1967] A. Prior. *Past, Present and Future*. Clarendon Press, Oxford, 1967.

[Ruan, 2004] J. Ruan. Exploring the update universe. Master's thesis, ILLC, Amsterdam, 2004.

[Schmidt, 1986] D.A. Schmidt. *Denotational Semantics: A Methodology for Language Development*. Allyn and Bacon, Boston, MA, 1986.

[Segerberg, 1982] K. Segerberg. A completeness theorem in the modal logic of programs. In T. Traczyck, editor, *Universal Algebra and Applications*, pages 36–46. Polish Science Publications, 1982.

[Segerberg, 1992] K. Segerberg. Getting started: Beginnings in the logic of action. *Studia Logica*, 51:347–378, 1992.

[Stalnaker, 1974] R. Stalnaker. Pragmatic presuppositions. In M.K. Munitz and P.K. Unger, editors, *Semantics and Philosophy*, pages 197–213. New York University Press, 1974.

[Stalnaker, 1979] R. Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics Vol.9: Pragmatics*, pages 315–332. Academic Press, New York, 1979.

[Stalnaker, 1998] R. Stalnaker. On the representation of context. *Journal of Logic, Language and Information*, 7:3–19, 1998.

[Staudacher, 1987] P. Staudacher. Zur Semantik indefiniter Terme. In B. Asbach-Schnitker and J. Roggenhofer, editors, *Neue Forschungen zur Wortbildung und Historiographie der Linguistik*, pages 239–258. Gunter Narr Verlag, Tübingen, 1987.

[Stoy, 1977] J.E. Stoy. *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory*. MIT Press, Cambridge, MA, 1977.

[Streett, 1982] R.S. Streett. Propositional dynamic logic of looping and converse is elementarily decidable. *Information and Control*, 54(1/2): 124–141, July/August 1982.

[Streett and E.A, 1989] R.S. Streett and E.A. Emerson. An automata theoretic decision procedure for the propositional mu-calculus. *Information and Computation*, 81(3): 249–264, 1989.

[Tarski, 1941] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6:73–89, 1941.

[ten Cate et al., 2001] B. ten Cate, J. van Eijck, and J. Heguiabehere. Expressivity of extensions of dynamic predicate logic. In R. van Rooy and M. Stokhof, editors, *Proceedings of the Thirteenth Amsterdam Colloquium, December 17–19, 2001*, pages 61–66. University of Amsterdam, Amsterdam, 2001.

[Turing, 1936] A.M. Turing. On computable real numbers, with an application to the Entscheidung problems. *Proceedings of the London Mathematical Society*, 2(42): 230–265, 1936.

[Van Benthem, 1976] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, University of Amsterdam, 1976.

[van Benthem, 1994] J. van Benthem. Programming operations that are safe for bisimulation. *Studia Logica*, 60(2): 311–330, 1994.

[van Benthem, 1996] J. van Benthem. *Exploring Logical Dynamics*. CSLI, Stanford, CA, 1996.

[van Benthem and Cepparello, 1994] J. van Benthem and G. Cepparello. Tarskian variations. dynamic parameters in classical semantics. Research Report CS-R9419, Centrum voor Wiskunde en Informatica, Amsterdam, 1994.

[van Benthem et al., 1997] J. van Benthem, R. Muskens, and A. Visser. Dynamics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Linguistics*, pages 587–648. Elsevier, Amsterdam, 1997.

[van der Torre and Tan, 1998] L. van der Torre and Y.-H. Tan. An update semantics for deontic reasoning. *Proceedings of DEON '98*, pages 409–426. 1998.

[van Eijck, 1997] J. van Eijck. Typed logics with states. *Logic Journal of the IGPL*, 5(5): 623–645, 1997.

[van Eijck, 1999] J. van Eijck. Axiomatising dynamic logics for anaphora. *Journal of Language and Computation*, 1:103–126, 1999.

[van Eijck, 2000] J. van Eijck. The proper treatment of context in NL. In P. Monachesi, editor, *Computational Linguistics in the Netherlands 1999; Selected Papers from the Tenth CLIN Meeting*, pages 41–51. Utrecht Institute of Linguistics OTS, 2000.

[van Eijck, 2001] J. van Eijck. Incremental dynamics. *Journal of Logic, Language and Information*, 10:319–351, 2001.

[van Eijck, 2004] J. van Eijck. Reducing dynamic epistemic logic to PDL by program transformation. CWI, Amsterdam, 2004. `www.cwi.nl:~/papers/04/delpdl/`

[van Eijck and Kamp, 1997] J. van Eijck and H. Kamp. Representing discourse in context. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 179–237. Elsevier, Amsterdam, 1997.

[van Eijck and de Vries, 1992] J. van Eijck and F.-J. de Vries. Dynamic interpretation and Hoare deduction. *Journal of Logic, Language and Information*, 1(1): 1–44, 1992.

[van Eijck and de Vries, 1995] J. van Eijck and F.-J. de Vries. Reasoning about update logic. *Journal of Philosophical Logic*, 24:19–45, 1995.

[van Eijck *et al.*, 2001] J. van Eijck, J. Heguiabehere, and B. Ó Nualláin. Tableau reasoning and programming with dynamic first order logic. *Logic Journal of the IGPL*, 9(3): 411–445, May 2001.

[Veltman, 1984] F. Veltman. Data semantics. In J. Groenendijk, T.M.V. Janssen, and M. Stokhof, editors, *Truth, Interpretation and Information*, pages 43–62. Foris, Dordrecht, 1984.

[Veltman, 1986] F Veltman. Data semantics and the pragmatics of indicative conditionals. In E. Traugott, J. Reilly, A. ter Meulen and C. Ferguson (eds.), *On Conditionals*. Cambridge University Press, Cambridge and New York, 1986.

[Veltman, 1996] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[Veltman, 2005] F. Veltman. Making counterfactual assumptions. *Journal of Semantics*, 22:159–180, May 2005.

[Vermeulen, 1993] C.F.M. Vermeulen. Sequence semantics for dynamic predicate logic. *Journal of Logic, Language and Information*, 2:217–254, 1993.

[Vermeulen, 1995] C.F.M. Vermeulen. Merging without mystery. Variables in dynamic semantics. *Journal of Philosophical Logic*, 24:405–450, 1995.

[Vermeulen, 2000] C.F.M. Vermeulen. Variables as stacks. *Journal of Logic, Language and Information*, 9:143–167, 2000.

[Vermeulen, 2001] C.F.M. Vermeulen. A calculus of substitutions for DPL. *Studia Logica*, 68:357–387, 2001.

[Visser, 1995] A. Visser. Relation validity and dynamic predicate logic. Logic Group Preprint Series 144, Department of Philosophy, Utrecht University, 1995.

[Visser, 1997] A. Visser. Dynamic Relation Logic is the logic of DPL-relations. *Journal of Logic, Language and Information*, 6(4): 441–452, 1997.

[Visser, 1998] A. Visser. Contexts in dynamic predicate logic. *Journal of Logic, Language and Information*, 7(1): 21–52, 1998.

[Wright, 1983] G.H. von Wright. *Practical Reason*. Blackwell, Oxford, 1983.

[Zarnic, 2002] B. Zarnic. Dynamic semantics, imperative logic and propositional attitudes. Preprint, Uppsala Universitet, 2002.

[Zeevat, 1989] H. Zeevat. A compositional approach to discourse representation theory. *Linguistics and Philosophy*, 12:95–131, 1989.

[Zeevat, 1992] H. Zeevat. Presupposition and accomodation in update semantics. *Journal of Semantics*, 98:379–412, 1992.

SITUATION THEORY AND
SITUATION SEMANTICS

Keith Devlin

1 INTRODUCTION

Situation semantics is a mathematically based theory of natural language semantics introduced by the mathematician Jon Barwise in 1980, and developed jointly by Barwise and the philosopher John Perry (and subsequently several others) throughout the 1980s. The first major treatment of the new theory was presented in Barwise and Perry's joint book *Situations and Attitudes* [1983].

Initially, situation semantics was conceived as essentially synthetic, with a mathematical ontology built up on set theory. Soon after the appearance of [Barwise and Perry, 1983], however, the authors changed their approach and decided to handle the topic in an analytic fashion, abstracting a mathematical ontology from analyses of natural language use. Situation theory is the name they gave to the underlying mathematics that arose in that manner. From the mid 1980s onward, therefore, situation semantics was an analysis of semantic issues of natural language based on situation theory.

Much of the initial development work in situation semantics was carried out at the Center for the Study of Language and Information (CSLI), an interdisciplinary research center established at Stanford University through a $23 million gift to Stanford from the System Development Foundation (a spin-off from RAND Corporation).

As originally conceived, situation semantics is an information-based theory, that seeks to understand linguistic utterances in tems of the information conveyed. (Although work carried out by Devlin and Rosenberg in the 1990s showed that situation theory could also be used to analyze language use from an action perspective [1996].) Barwise and Perry began with the assumption that people use language in limited parts of the world to talk about (i.e. exchange information about) other limited parts of the world. Call those limited parts of the world *situations*.

In their paper *The Situation Underground* [1980], the first published work on situation semantics, Barwise and Perry wrote of situations:

> The world consists not just of objects, or of objects, properties and
> relations, but of objects having properties and standing in relations

to one another. And there are parts of the world, clearly recognized (although not precisely individuated) in common sense and human language. These parts of the world are called situations. Events and episodes are situations in time, scenes are visually perceived situations, changes are sequences of situations, and facts are situations enriched (or polluted) by language.

The appearence of the word "parts" in the above quotation is significant. Situations are *parts* of the world and the information an agent has about a given situation at any moment will be just a *part* of all the information that is theoretically available. The emphasis on partiality contrasts situation semantics from what was regarded by many as its principal competitor as a semantic theory, possible worlds semantics.

It is important to realize that, the use of mathematical concepts notwithstanding, in situation theory and situation semantics, situations are taken to be real, actual parts of the world, and the basic properties and relations the situation semantics deals with are taken to be real uniformities across situations (and not bits of language, ideas, sets of n-tuples, functions, or some other mathematical abstractions).

Situation semantics provides a *relational theory of meaning*. In its simplest form, the meaning of an expression ϕ it taken to be a relation

$$d, c \| \phi \| e$$

between an utterance or discourse situation d, a speaker's connection function c, and a described situation e. These concepts will all be described in due course.

Although described as a "theory", situation theory is more profitably approached as a set of mathematically-based tools to analyze, in particular, the way context facilitates and influences the rise and flow of information. Similarly, situation semantics is best approached as a method for analyzing semantic phenomena. This perspective is reflected in the structure of this article. After providing a brief explanation of the key ideas of situation theory and situation semantics, we present a number of specific topics in situation semantics. It is not intended to be a comprehensive coverage. Rather the goal is to provide some indication of the manner in which the methods of situation semantics may be applied.

2   INFORMATION

Information is always taken to be information *about* some situation, and is assumed to be built up from discrete informational items known as *infons*. Infons are of the form

$$\langle\!\langle R, a_1, \ldots, a_n, 1 \rangle\!\rangle \quad , \quad \langle\!\langle R, a_1, \ldots, a_n, 0 \rangle\!\rangle$$

where R is an n-place relation and $a_1, \ldots, a_n$ are objects appropriate for R.

Infons are not things that in themselves are true or false. Rather a particular item of information may be true or false about a situation. Given a situation, s, and an infon σ, write

$$s \models \sigma$$

to indicate that the infon σ is made factual by the situation s. The offical terminology is that s *supports* σ. Thus,

$$s \models \langle\!\langle R, a_1, \ldots, a_n, 1 \rangle\!\rangle$$

means that, in the situation s, the objects $a_1, \ldots, a_n$ stand in the relation R, and

$$s \models \langle\!\langle R, a_1, \ldots, a_n, 0 \rangle\!\rangle$$

means that, in the situation s, the objects $a_1, \ldots, a_n$ do not stand in the relation R.

Infons may be combined, recursively, to form *compound infons*. The combinatory operations are conjunction, disjunction, and situation-bounded existential and universal quantification. This is discussed later.

Given a situation s and a compound infon σ,

$$s \models \sigma$$

is defined by recursion in the obvious way. The actuality $s \models \sigma$ is referred to as a *proposition*.

3   TYPES

From a formal viewpoint, situation theory is many sorted. The objects (called *uniformities*) in the ontology include the following:

- *individuals*, denoted by $a, b, c, \ldots$

- *relations*, denoted by $P, Q, R, \ldots$

- spatial *locations*, denoted by $l, l', l'', l_0, l_1, l_2, \ldots$

- *temporal locations*, denoted by $t, t', t_0, \ldots$

- *situations*, denoted by $s, s', s'', s_0, \ldots$

- *types*, denoted by $S, T, U, V, \ldots$

- *parameters*, denoted by $\dot{a}, \dot{s}, \dot{t}, \dot{l}$, etc.

These entities are assumed to be — or to correspond to — aspects of the agent's cognition of the world. That is, the agent has a *scheme of individuation* whereby it carves the world up into manageable pieces. This "carving up" may take the form of cognitive individuation or merely behavioral discrimination.

A particular feature of intelligent behavior is the recognition of types. The agent recognizes (either consciously or through its behavior) various types of object, various types of activity, etc.

The *basic types* of the formal theory are:

- TIM : the type of a temporal location

- LOC : the type of a spatial location

- IND : the type of an individual

- REL^n : the type of an n-place relation

- SIT : the type of a situation

- INF : the type of an infon

- TYP : the type of a type (see later)

- PAR : the type of a *parameter* (see later)

- POL : the type of a polarity (0 and 1)

Given an object, x, and a type, T, we write

$$x : T$$

to indicate that the object x is *of* type T.

4   PARAMETERS

During the development of situation theory and situation semantics, considerable discussion was devoted to the topic of parameters. The reason for this attention was that, uniquely in the ontology, parameters are not individuated (in any direct sense) by the agent; they are theoretical constructs. They do, however, correspond to, and capture within the theoretical framework, important aspects of the agent's cognitive behavior. It is the very essence of cognitive activity that the agent tracks various connections. For example, an agent aware of the connection between smoke and fire, who knows that smoke is an indication of fire, needs to be able to connect any specific instance of smoke to a specific instance of fire, one directly linked to the perceived smoke. Within situation semantics, parameters capture such linkages. It is through the mechanism of parameters that the general regularities that govern cognitive activity, reasoning, and information flow become applicable in actual circumstances.

For each basic type T other than PAR, there is an infinite collection $T_1, T_2, T_3, \ldots$ of *basic parameters*, used to denote arbitrary objects of type T.

The parameters T_i are sometimes referred to as T-*parameters*.

Notation: $\dot{l}$, $\dot{t}$, $\dot{a}$, $\dot{s}$, etc. to denote parameters (of type *LOC, TIM, IND, SIT*, etc.).

Parameters are place-holders for specific entities, which the theoretical framework uses to track crucial information links. Anchors for parameters provide a formal mechanism for linking parameters to actual entities. An *anchor* for a set, A, of basic parameters is a function defined on A, which assigns to each parameter T_n in A an object of type T.

If σ is a compound infon and f is an anchor for some of the parameters in σ, $\sigma[f]$ denotes the compound infon that results from replacing each parameter $\dot{a}$ in $\text{dom}(f)$ by $f(a)$.

In order to provide a more streamlined treatment of various linguistic and (other) cognitive phenomena, situation theory provides a mechanism for restricting the scope of parameters. *Restricted parameters* are constructed as follows.

Let v be a parameter. A *condition* on v is a finite conjunction of infons. (At least one conjunct should involve v, otherwise the definition is degenerate.)

Given a parameter, v, and a condition, C, on v, define a new parameter, $v \restriction C$, called a *restricted parameter*. $v \restriction C$ denotes an object of the same type as v, that satisfies the requirements imposed by C (*in any situation where this applies*). (If C consists of a single parametric infon σ, we write $v \restriction \sigma$ instead of $v \restriction \{\sigma\}$.)

> Let $r = v \restriction C$ be a parameter. Given a situation s, a function f is said to be an *anchor* for r in s if:
>
> 1. f is an anchor for v and for every parameter that occurs free in C ;
> 2. for each infon σ in C: $s \models \sigma[f]$;
> 3. $f(r) = f(v)$.

5   INFON LOGIC

Using parameters, the formal definition of the *conjunction* $\sigma \wedge \tau$ of two infons σ, τ is as follows.

> For any situation, s,

$$s \models \sigma \wedge \tau \ \text{ iff } \ s \models \sigma \text{ and } s \models \tau.$$

The conjunction is not itself an infon, but a *compound infon*.

The *disjunction* of two infons σ, τ is a compound infon $\sigma \vee \tau$ such that for any situation s,

$$s \models \sigma \vee \tau \ \text{ iff } \ s \models \sigma \ \text{ or } \ s \models \tau \ \text{(or both)}.$$

The above definitions are in fact clauses in a recursive definition of compound infons.

If σ is an infon (or compound infon) that involves the parameter $\dot{x}$ and u is some set, then

$$(\exists \dot{x} \in u)\sigma$$

is a compound infon.

For any situation, s, that contains (as constituents) all members of u:

$$s \models (\exists \dot{x} \in u)\sigma$$

iff there is an anchor, f, of $\dot{x}$ to an element of u, such that $s \models \sigma[f]$.

The anchor, f, here may involve some *resource situation* other than s. f must assign to $\dot{x}$ an appropriate object in some anchoring situation, e, that supports the various infons that figure in the structure of $\dot{x}$.

For example, let σ be the compound infon

$$\langle\!\langle \text{tired}, \dot{c}, t_0, 1 \rangle\!\rangle \wedge \langle\!\langle \text{hungry}, \dot{c}, t_0, 1 \rangle\!\rangle$$

where $\dot{c}$ is a parameter for a cat.

Let s be a room situation at time t_0 and u the set of individuals in s. Then:

$$s \models (\exists \dot{c} \in u)\sigma$$

iff there is an anchor, f, of $\dot{c}$ to some fixed object, c, in u (c necessarily a cat) such that $s \models \sigma[f]$, i.e. such that

$$s \models \langle\!\langle \text{tired}, c, t_0, 1 \rangle\!\rangle \wedge \langle\!\langle \text{hungry}, c, t_0, 1 \rangle\!\rangle.$$

That is to say, $s \models (\exists \dot{c} \in u)\sigma$ iff there is a cat, c, in u that at time t_0 is tired and hungry in s.

The existence of the anchor, f, entails the existence of an associated *anchoring* (or *resource*) *situation*, e, such that (in particular)

$$e \models \langle\!\langle \text{cat}, c, 1 \rangle\!\rangle.$$

In particular, c is a constituent of e.

Note that the object c has to be in the (room) situation, s, at time t_0 in order for the proposition

$$s \models \langle\!\langle \text{tired}, c, t_0, 1 \rangle\!\rangle \wedge \langle\!\langle \text{hungry}, c, t_0, 1 \rangle\!\rangle$$

to obtain.

If σ is an infon (or compound infon) that involves the parameter $\dot{x}$, and if u is some set, then

$$(\forall \dot{x} \in u)\sigma$$

is a compound infon.

For any situation, s, that contains (as constituents) all members of u:

$$s \models (\forall \dot{x} \in u)\sigma$$

iff, for all anchors, f, of $\dot{x}$ to an element of u, $s \models \sigma[f]$.

In the cases both of existential and universal quantification, the bounding set u may consist of all the objects of a certain kind that are in the situation s. Consequently, the definitions do provide a notion of 'unrestricted' quantification, but it is a notion of *situated* quantification.

For an example of situated quantification, when someone truthfully asserts

All citizens have equal rights.

they are presumably quantifying over some country such as the United States, not the entire world, for which such a claim is not true.

6 TYPE ABSTRACTION

Situation theory provides various mechanisms for defining types. The two most basic methods are type-abstraction procedures for the construction of two kinds of types: situation-types and object-types.

Situation-types. Given a *SIT*-parameter, $\dot{s}$, and a compound infon σ, there is a corresponding *situation-type*

$$[\dot{s} \mid \dot{s} \models \sigma],$$

the *type* of situation in which σ obtains.

This process of obtaining a type from a parameter, $\dot{s}$, and a compound infon, σ, is known as *(situation-) type abstraction*. The parameter $\dot{s}$ is called the *abstraction parameter* used in this type abstraction.

For example,

$$[SIT_1 \mid SIT_1 \models \langle\!\langle \text{running}, \dot{p}, LOC_1, TIM_1, 1 \rangle\!\rangle].$$

Object-types. These include the basic types *TIM, LOC, IND, REL^n, SIT, INF, TYP, PAR,* and *POL,* as well as the more fine-grained uniformities described below.

Object-types are determined over some initial situation. Let s be a given situation. If $\dot{x}$ is a parameter and σ is some compound infon (in general involving $\dot{x}$), then there is a type

$$[\dot{x} \mid s \models \sigma],$$

the *type* of all those objects x to which $\dot{x}$ may be anchored in the situation s, for which the conditions imposed by σ obtain.

This process of obtaining a type $[\dot{x} \mid s \models \sigma]$ from a parameter, $\dot{x}$, a situation, s, and a compound infon, σ, is called *(object-) type abstraction*.

The parameter $\dot{x}$, is known as the *abstraction parameter* used in this type abstraction.

The situation s is known as the *grounding* situation for the type. In many instances, the grounding situation, s, is the world or the environment we live in (generally denoted by w).

For example, the *type* of all people could be denoted by

$$[IND_1 \mid w \models \langle\langle \text{person}, IND_1, \dot{l}_w, \dot{t}_{now}, 1 \rangle\rangle].$$

Again, if s denotes Jon's environment (over a suitable time span), then

$$[\dot{e} \mid s \models \langle\langle \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \rangle\rangle]$$

denotes the type of all those situations Jon sees (within s). This is a case of an object-type that is a type of situation.

This example is not the same as a *situation-type*. Situation-types classify situations according to their internal structure, whereas in the type

$$[\dot{e} \mid s \models \langle\langle \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \rangle\rangle],$$

the situation is typed from the outside.

7   CONSTRAINTS

Types and the type abstraction procedures provide a mechanism for capturing the fundamental process whereby a cognitive agent classifies the world. *Constraints* provide the situation theoretic mechanism that captures the way that agents make inferences and act in a rational fashion. Constraints are linkages between situation types. They may be natural laws, conventions, logical (i.e. analytic) rules, linguistic rules, empirical, law-like correspondences, etc.

For example, humans and other agents are familiar with the constraint:

Smoke means fire.

If S is the type of situations where there is smoke present, and S' is the type of situations where there is a fire, then an agent (e.g. a person) can pick up the information that there is a fire by observing that there is smoke (a type S situation) and being aware of, or *attuned to*, the constraint that links the two types of situation. This constraint is denoted by

$$S \Rightarrow S'.$$

(This is read as "S *involves* S'.")

Another example is provided by the constraint

FIRE *means fire.*

This constraint is written

$$S'' \Rightarrow S'.$$

It links situations (of type S'') where someone yells the word FIRE to situations (of type S') where there is a fire.

Awareness of the constraint

$$\text{FIRE } means \text{ } fire$$

involves knowing the meaning of the word FIRE and being familiar with the rules that govern the use of language.

The three types that occur in the above examples may be defined as follows:

$$
\begin{aligned}
S &= [\dot{s} \mid \dot{s} \models \langle\!\langle \text{smokey}, \dot{t}, 1 \rangle\!\rangle] \\
S' &= [\dot{s} \mid \dot{s} \models \langle\!\langle \text{firey}, \dot{t}, 1 \rangle\!\rangle] \\
S'' &= [\dot{u} \mid \dot{u} \models \langle\!\langle \text{speaking}, \dot{a}, \dot{t}, 1 \rangle\!\rangle \wedge \langle\!\langle \text{utters}, \dot{a}, \mathit{fire}, \dot{t}, 1 \rangle\!\rangle].
\end{aligned}
$$

Notice that constraints link types, not situations. However, any particular instance where a constraint is utilized to make an inference or to govern/influence behavior will involve specific situations (of the relevant types). Constraints function by capturing various regularities across actual situations.

A constraint

$$C = [S \Rightarrow S']$$

allows an agent to make a logical inference, and hence faciliates information flow, as follows. First the agent must be able to discriminate the two types S and S'. (This use of the word 'discriminate' is not intended to convey more than the most basic of cognitive activities.) Second, the agent must be aware of, or behaviorally attuned to, the constraint. Then, when the agent finds itself in a situation s of type S, it knows that there must be a situation s' of type S'. We may depict this diagrammatically as follows:

$$
\begin{array}{ccc}
S & \xRightarrow{\;C\;} & S' \\
s : S \uparrow & & \uparrow s' : S' \\
s & \xrightarrow{\;\exists\;} & s'.
\end{array}
$$

For example, suppose $S \Rightarrow S'$ represents the constraint *smoke means fire*. Agent $\mathcal{A}$ sees a situation s of type S. The constraint then enables $\mathcal{A}$ to conclude correctly that there must in fact be a fire, that is, there must be a situation s' of type S'. (For this example, the constraint $S \Rightarrow S'$ is most likely reflexive, in that the situation s' will be the same as the encountered situation s.)

A particularly important feature of this analysis is that it separates clearly the two very different kinds of entity that are crucial to the creation and transmission of information: one the one hand the abstract types and the constraints that link

them, and on the other hand the actual situations in the world that the agent either encounters or whose existence it infers.

It should be noted that the ontology of situation theory has no bottom layer; every individual or situation can be subdivided into constituents, if desired. This implies that it is possible to represent and analyze a domain at any degree of granularity, to move smoothly up and down the granularity scale during an analysis, and to "zoom" the granularity to investigate specific issues in an analysis, while keeping the remainder of the representation fixed. This feature can play a major role in applications; for example, the analysis of engineer repair reports from a large computer manufacturer, described in [Devlin and Rosenberg, 1996].

8   SITUATION SEMANTICS: THE BASIC IDEA

The object of study in situation semantics is the utterance. In the simplest version, situation semantics analyzes utterances in terms of three situations:

- Utterance situation,

- Resource situation,

- Focal situation.

The utterance situation. This is the context in which the utterance is made and received.

If Melissa says to Naomi

$$A \ man \ is \ at \ the \ door$$

the utterance situation, u, is the immediate context in which Melissa utters these words and Naomi hears them.

The situation u includes both Melissa and Naomi (for the duration of the utterance), and should be sufficiently rich to identify various salient factors about this utterance, such as the door that Melissa is referring to..

This is probably the one in her immediate environment, but not necessarily. For instance, if Melissa utters the sentence $A \ man \ is \ at \ the \ door$ as part of a larger discourse, the situation u could provide an alternative door.

The connections between the utterance and the various objects referred to, are known as just that: *connections* (or *speaker's connections*). Thus

$$u \models \ \langle\!\langle \text{utters, Melissa}, \Phi, l, t, 1 \rangle\!\rangle \wedge \langle\!\langle \text{refers-to, Melissa}, the \ door, D, l, t, 1 \rangle\!\rangle$$

where Φ is the sentence $A \ man \ is \ at \ the \ door$ and D is a door that is fixed by u.

The speaker's connections link the utterance (as part of u) of the phrase *the door* to the object D.

Resource situations. If Melissa says

The man I saw running yesterday is at the door,

she is making use of a situation that she witnessed the day before, the one in which a certain man was running, in order to identify the man at the door.

There is another situation, r, a situation that occurred the day before the utterance, and which Melissa witnessed, such that

$$u \quad \models \quad \langle\!\langle \text{utters, Melissa, } \Phi, l, t, 1\rangle\!\rangle \wedge$$

$$\langle\!\langle \text{refers-to, Melissa, } \textit{the man}, M, l, t, 1\rangle\!\rangle \wedge$$

$$\langle\!\langle \text{refers-to, Melissa, } \textit{the door}, D, l, t, 1\rangle\!\rangle$$

where Φ is the sentence

The man I saw running yesterday is at the door

and where Melissa is making use of r and the fact that M is the unique man such that (for some appropriate values of l', t')

$$r \models \langle\!\langle \text{runs, } M, l', t', 1\rangle\!\rangle.$$

Resource situations can become available for exploitation in various ways, such as:

1. by being perceived by the speaker;

2. by being the objects of some common knowledge about the world;

3. by being the way the world is;

4. by being built up by previous discourse.

The focal situation. Also known as the *described situation*, the focal situation is that part of the world the utterance is about.

Features of the utterance situation serve to identify the focal situation. For instance, suppose Melissa makes her utterance while peering out of the upstairs window at the house across the street. Then her utterance refers to the situation, s, that she sees, the situation at the house across the street, and we have

$$s \models \langle\!\langle \text{present, } M, l, t, 1\rangle\!\rangle$$

where l is the location of the door and t is the time of the utterance.

9   PROPOSITIONAL CONTENT

By adopting an ontology that includes items of information (infons), we are able to capture the notion of *the information* encoded by a representation, and can account for the fact that the same information can be encoded by two quite different representations, using quite different representation schemas.

There are then three notions that are often treated as if they were somewhat interchangeable, but which situation theory regards as quite distinct (though related):

- information
- representations
- propositions.

In the case of a linguistic utterance, say Jon's utterance of the assertive sentence

Mary is running

the representation is the utterance itself, which we regard as a situation, call it u.

The *propositional content* of the utterance u is the proposition

$$e \models \sigma$$

where e is the focal situation, σ is the infon $\langle\!\langle \text{runs}, M, t_u, 1 \rangle\!\rangle$, M denotes the individual Mary to whom Jon refers, t_u is the time of the utterance, and e is determined by various features of the utterance.

For example, e could be determined by Jon and the listener being part of some larger situation in which this individual Mary is running, or more generally by means of some other form of previously established context of utterance.

The propositional content is what might normally be referred to as the "information conveyed by the utterance".

10   LINGUISTIC MEANING

As we have seen already, the *meaning* of an assertive sentence, Φ, is a constraint, an abstract link that connects the type of an utterance of Φ with the type of the described situation. More generally, we can describe the meaning of other kinds of sentence, and of a word or phrase, α, and in these cases too the meaning will be a link between appropriate types.

In the case where a speaker utters the word, phrase, or sentence, α, to a single listener, we shall use u to denote the utterance situation, e the (larger) embedding situation, r any resource situation, and s the described situation. We denote the speaker in u by a_u, and the listener by b_u. The time and location of the utterance are denoted by t_u, l_u, respectively.

$U(\alpha)$ denotes the situation-type of an utterance of α, namely:

$$U(\alpha) = [\dot{u} \mid \dot{u} \models \quad \ll \text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll \text{utters}, \dot{a_u}, \alpha, \dot{l_u}, \dot{t_u}, 1 \gg].$$

Situation semantics distinguishes two different kinds of meaning. The *abstract meaning* supplies the answer to the question "What does this word/phrase/sentence mean (*in general*)?", where the word/phrase/sentence is taken out of any context; the *meaning-in-use* answers the question "What does this word/phrase/

sentence mean (*as it is being used in this instance*)?", where the word/phrase/sent-ence is uttered in a particular context. The meaning-in-use is *induced* by the abstract meaning, with the former a particular instantiation of the latter. In the case of an utterance of a sentence, the meaning-in-use is closely related to the propositional content. The *abstract meaning* is represented as an abstract link between two types; the *meaning-in-use* as a relation between pairs of objects, in general not types.

The abstract meaning of a part of speech, α, will be denoted by $\mathcal{M}(\alpha)$; the meaning-in-use of α will be denoted by $\|\alpha\|$.

In the case of individual words, the meaning-in-use provides a link between the utterance situation and the object (possibly an abstract object, such as a relation) in the world that the word denotes.

It should be born in mind that the brief account that follows provides a fairly crude notion of word meaning. In practice, when a word is uttered as part of a sentence or an extended discourse, the overall context of utterance can contribute features to the meaning of that word (in that context).

11   THE MEANING OF 'I'

In any utterance, u, 'I' denotes the speaker, a_u, of u. The meaning-in-use, $\|\mathrm{I}\|$, of 'I' is the relation that connects u to a_u for any utterance u. So, for given objects u and a,

$$u\|\mathrm{I}\|a \ \text{ if and only if } \ u : U(\mathrm{I}) \ \text{ and } \ a = a_u.$$

Thus the meaning-in-use of the pronoun 'I' is a relation linking situations to indi-viduals.

The abstract meaning of 'I', $\mathcal{M}(\mathrm{I})$, is the link between the situation-type

$$U(\mathrm{I}) = [\dot{u} \mid \dot{u} \models \quad \ll\text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll\text{utters}, \dot{a_u}, \mathrm{I}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the object-type

$$E = [\dot{a} \mid \dot{u} \models \ll =, \dot{a}, \dot{a_u}, \dot{l_u}, \dot{t_u}, 1 \gg].$$

Notice that there is exactly one type E such that $U(\mathrm{I})[\mathcal{M}(\mathrm{I})]E$ here.

The abstract link $\mathcal{M}(\mathrm{I})$ induces the relation $\|\mathrm{I}\|$ in the fashion:

$$\|\mathrm{I}\| = \{(u, a) \mid u : U(\mathrm{I}) \ \& \ a : E \text{ where } U(\mathrm{I})[\mathcal{M}(\mathrm{I})]E\}.$$

12   THE MEANING OF 'YOU'

In any utterance situation, 'YOU' denotes the listener. Thus the meaning-in-use of the word 'YOU' is such that

$$u\|\text{YOU}\|b \ \text{ if and only if } \ u : U(\text{YOU}) \ \text{ and } \ b = b_u$$

and the abstract meaning, $\mathcal{M}(\text{YOU})$, is the link between the situation-type

$$U(\text{YOU}) = [\dot{u} \mid \dot{u} \models \ \ \ll\text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll\text{utters}, \dot{a_u}, \text{YOU}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the object-type

$$E = [\dot{b} \mid \dot{u} \models \ll =, \dot{b}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg].$$

13   THE MEANING OF 'HE', 'SHE', 'IT'

Taking the case 'HE' for definiteness, the significant feature of the pronoun 'HE', when considered out of context, is that it is used to denote a male individual. The appropriate type then to figure in the abstract meaning is the type of any male individual:

$$F = [\dot{b} \mid w \models \ll\text{male}, \dot{b}, 1 \gg]$$

where $\dot{b}$ is an *IND*-parameter and where w denotes the world.

The abstract meaning, $\mathcal{M}(\text{HE})$, will be the link between the situation-type

$$U(\text{HE}) = [\dot{u} \mid \dot{u} \models \ \ \ll\text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll\text{utters}, \dot{a_u}, \text{HE}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the object-type F.

Of course, in this case, the abstract meaning does not really capture the main feature of a pronoun, which is to refer to a particular individual of the appropriate gender. Rather, pronouns really acquire meaning when used in a specific context, and accordingly it is the meaning-in-use that is the more important of the two forms of meaning in this case.

Turning to that meaning-in-use, there are two main ways a pronoun can pick up its referent: either through the speaker or else by having some other noun phrase as an antecedent. Consider, for instance, the sentence:

Jon thought he was wrong.

Uttered one way, 'HE' refers to Jon himself; that is to say, the pronoun picks up its referent *anaphorically* from a previous part of the utterance. Alternatively, the speaker could be using 'HE' *dictically*, to refer to some other person, say Jerry. This referent could be provided by the speaker pointing to Jerry, or could be supplied by some previous utterance as part of a discourse, such as:

Jerry said there was a language of thought. Jon thought he was wrong.

Thus the interpretation of an utterance of the pronoun 'HE' requires the provision of a referent by means of the utterance situation. That is to say, the utterance situation, u, must supply some individual $h = i_u(\text{HE})$ (or $h = i_u(\text{HIM})$) such that for some resource situation, r,

$$r \models \ll\text{male}, h, 1 \gg$$

and then, for any a,

$$u\|\text{HE}\|a \text{ if and only if } u : U(\text{HE}) \text{ and } a = i_u(\text{HE}).$$

Notice that the individual $h = i_u(\text{HE})$ need not be a constituent of the utterance situation. Rather the speaker uses, or relies upon, some resource situation, r, and it is that resource situation, r, that has h as a constituent. Similarly for the other pronouns, 'SHE', 'IT', etc.

14   THE MEANING OF PROPER NAMES

Used correctly, a proper name should designate a particular individual. Since many individuals often share the same name, this means that the context should somehow identify the requisite individual the speaker has in mind. Thus for a proper use of the name 'JAN', the utterance situation, u, should provide an individual $p = i_u(\text{JAN})$ such that for some resource situation, r,

$$r \models \ll\text{named}, p, \text{JAN}, 1 \gg$$

and then, for any a,

$$u\|\text{JAN}\|a \text{ if and only if } u : U(\text{JAN}) \text{ and } a = i_u(\text{JAN}).$$

As with the case of third-person pronouns above, there is no requirement that the person Jan be present in the utterance situation. Rather Jan is a constituent of the resource situation, r, which the speaker makes use of when he makes his utterance.

Also as with third-person pronouns, the abstract meaning of a proper name does not really capture what names are about in the way that the meaning-in-use does. For example, $\mathcal{M}(\text{JAN})$ is the link between the situation-type

$$U(\text{JAN}) = [\dot{u} \mid \dot{u} \models \quad \ll\text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll\text{utters}, \dot{a_u}, \text{JAN}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the object-type

$$E = [\dot{b} \mid w \models \ll\text{named}, \dot{b}, \text{JAN}, 1 \gg].$$

To point out one particular manner in which the abstract meaning of proper names is simply at too high a level of abstraction to really capture the way names are used, notice that, if a is an individual of type E, then we shall have

$$w \models \ll \text{named}, a, \text{Jan}, 1 \gg$$

so for some temporal location t we will have

$$w \models \ll \text{named}, a, \text{Jan}, t, 1 \gg .$$

So all this tells us is that, at *some* time, this individual a is named 'Jan'. But of course, people can and do change their names, whereas correct usage of proper names requires using the name that prevails at the appropriate time. And indeed this may be reflected in the meaning-in-use. In the present framework this could result from the resource situation having the appropriate temporal duration. But there are other possibilities.

For instance, if the word 'Jan' were uttered as part of a complete sentence, then features of the utterance as a whole could provide an appropriate temporal location t_0 so that in the meaning-in-use of the proper name 'Jan' (on this occasion) we have

$$r \models \ll \text{named}, a, \text{Jan}, t_0, 1 \gg$$

where r is a resource situation.

15   THE MEANING OF NOUNS

The abstract meaning of a noun, α, is the link between the type, $U(\alpha)$, of an utterance of α, and the type of the object denoted by α. For example, the abstract meaning of the noun 'APPLE' is the link between the situation-type

$$U(\text{APPLE}) = [\dot{u} \mid \dot{u} \models \quad \ll \text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge$$
$$\ll \text{utters}, \dot{a_u}, \text{APPLE}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the object-type of all apples:

$$[\dot{b} \mid w \models \ll \text{apple}, \dot{b}, 1 \gg]$$

where 'apple' here denotes the property of being an apple.

As for meaning-in-use, this concept applies not so much to nouns as to noun phrases. The normal usage of a noun is as part of a noun phrase, and even on those occasions where a noun is uttered in naked fashion, such as when a small child looks at her plate and says "Apple", this can be regarded, for our purposes, as an abbreviation for the noun phrase 'An apple'.

16   THE MEANING OF VERBS

The meaning-in-use of any verb is the link between the verb and the relation it denotes. For example, the verb 'RUNS' corresponds to the relation, R, of running, and for any utterance situation, u,

$$u\|\text{RUNS}\|R.$$

(We deal with the issue of tense presently.)

To be consistent with the development so far, the abstract meaning of a verb, say 'runs', should be taken to be the link between the type of an utterance of the word 'RUNS' and the type of all relations of running. However, in this summary account we do not have parameters for relations and do not form relation-types, hence we cannot accommodate such a notion of abstract meaning of verbs. A more complete development, in which relation-types abstraction was allowed, would be able to handle this issue in the manner suggested.

17   SPEAKER'S CONNECTIONS

Notice that, in each case so far, the meaning-in-use of a word, α, is a relation, $\|\alpha\|$, that links an utterance situation, u, with a certain object, a, either an individual in the case where α is a pronoun or name, or a relation in the case of a verb. The relation $u\|\alpha\|a$ places a constraint on the utterance situation, u, to supply or contain a suitable object.

Given different utterance situations, the same word can be linked to different objects. Around CSLI at the time situation semantics was being developed, the name 'John' was very much dependent on the utterance situation: did the speaker mean John Perry, John Etchemendy, or John Nerbonne (or even Jon Barwise in the case of a *spoken* utterance)?

The notation used to denote the object that the utterance situation, u, provides to correspond to a word, α, via its meaning, is $c_u(\alpha)$. Thus, in the case of a third-person pronoun or a proper name, c_u is the same as the function i_u introduced a short while ago.

In case an utterance of a word or phrase, α, in an utterance, u, makes use of a resource situation, r, this resource situation is denoted by $c_u^{res}(\alpha)$.

In the case where u is an utterance of a sentence, Φ, there will also be a *described situation*, that part of the world the utterance of Φ is about. Denote this situation by $s_u(\Phi)$.

The term *speaker's connections* refers to any or all of the functions c_u, c_u^{res}, and s_u.

Thus the speaker's connections are the functional links between the words the speaker utters and those parts of, or objects in, the world she uses these words to refer to. They thus provide a mathematical realization of the intentionality of speech, the fact that agents use language to talk *about* the world.

Notice that effective communication requires that, in general, the listener is aware of the identity of the described situation, $s_u(\Phi)$, and the values of the speaker's connection function, c_u, and the onus is on the speaker to ensure that the listener is so aware. In general there is, however, no need for the listener to know the values of the resource-situation function, c_u^{res}. The role played by resource situations is simply that of a supporting background.

For instance, if, in the course of a conversation, a speaker uses the noun 'APPLE', then there must be some resource situation that supports the fact that the object referred to is indeed an apple, and if challenged the listener might well agree that there will be such a situation, but the identity of that resource situation is not in general important.

18   SPEAKER'S CONNECTIONS AND TENSED VERBS

Consider the following sentences.

> *Mary is running.*
> *Mary was running.*
> *Mary will run.*

In each case, the meaning of the word 'run' (ignoring the morphological differences between 'run', 'runs', 'running') connects this word to the same relation, R, the relation of running. In using a particular tense of this verb, the speaker is providing a reference to a particular time, the time at which the running takes/took place. Situation semantics accounts for this by means of the speaker's connections function. Thus,

- $c_u(\text{IS}) = t_u$

- $c_u(\text{WAS}) = t$ where $t \prec t_u$

- $c_u(\text{WILL}) = t$ where $t_u \prec t$.

The last two often occur in the context of an existential quantification over t.

19   THE MEANING OF SINGULAR NOUN PHRASES

We shall restrict attention to meaning-in-use, and leave it to the reader to supply the more general notion of abstract meaning (the link between the utterance type and an appropriate object-type, that induces the meaning-in-use).

We commence with definite descriptions. For example:

  (I)   THE MAN IN A BLACK HAT.
 (II)   THE PRESIDENT OF THE UNITED STATES.
(III)   THE KING OF FRANCE.

Each of these can be used to denote, or refer to, a specific individual. Such usage of a definite description is known as the *referential* use, which we consider first.

In each of the above three examples then, if we assume the phrase is used to refer to a particular individual, the question arises: where is that individual, i.e. what situation(s) is the individual a constituent of? Clearly, he need not necessarily be

a constituent of the utterance situation, or even the larger, embedding situation. In the case of example (I), an utterance of this phrase could well have the relevant individual present in the embedding situation, but most utterances of (II) will not be made in the presence of the US President. And of course no contemporary situation can include an individual that fits the description in (III), since there is no current King of France.

Rather, in making (referential) use of a definite description

$$\alpha = \text{THE } \pi$$

in the utterance situation, u, the speaker is making use of some resource situation, $r = c_u^{res}(\alpha)$, of which the requisite individual is a constituent.

So the meaning-in-use of α, $\|\alpha\|$, links u to an individual $a = c_u(\alpha)$ such that:

1. $r \models \ll \Pi, a, l_\Pi, t_\Pi, 1 \gg$; and

2. a is the unique individual in r with property (i),

where Π is the property (possibly complex) that corresponds to π, namely the property of *being a π*, and where l_Π and t_Π are the location and time associated with Π if this is location or time dependent.

That is to say, for any given situation u and individual a,

$u\|\text{THE } \pi\|a$ if and only if

$u : U(\text{THE } \pi)$ and a satisfies (i) and (ii), where $r = c_u^{res}(\text{THE } \pi)$.

Thus, in the case of example (I), suppose this sentence is uttered at a party, and it is this party (or maybe some time interval within this event) that we take to be the utterance situation, u. Then the legitimate utterance of this phrase, with reference to the situation u itself as resource situation, will require that there is a man in u wearing a black hat, and moreover there is only one such man.

On the other hand, if we take u to be some conversation that is going on at the party, say a conversation about the rock group playing at the other end of the room, then the phrase (I) may be legitimately uttered provided that precisely one man in the rock group is wearing a black hat, *even though at the party as a whole there may be many men wearing black hats*. This is because the conversation itself determines an appropriate resource situation, namely the situation comprising the rock group.

In either case, the entire party as a resource situation or the rock group as a resource situation, the speaker's connections provide a resource situation, r, in which there is exactly one man wearing a black hat (i.e. possessing the complex property associated with the phrase 'MAN IN A BLACK HAT', that is to say, *being a man in a black hat*), and then the meaning of the definite description (I) links the utterance situation u to this particular individual.

Returning now to example (II), this differs from (I) only in that the resource situation will in general be quite distinct from the utterance situation. In fact,

for most (referential) utterances of (II), the 'default' resource situation will be the entire USA over some period of time, a situation that may include the utterance situation or be quite disjoint from it.

Finally, sentence (III) is different from the other two in that there is, currently, no individual in the world that fits this description: there is no King of France. Thus a legitimate referential utterance of this phrase can only be made with reference to a resource situation located in the past, at a time when there was such a person

The meaning-in-use of an indefinite description (used referentially) such as

<div align="center">

A BLACK CAT

</div>

or

<div align="center">

A SMALL TOWN IN GERMANY

</div>

is defined in a similar way to that of a definite description, the only difference being that the uniqueness condition (clause (ii) in the above) is not required.

Other singular noun phrases are handled similarly. For instance, when used referentially by an individual KD, a phrase such as

<div align="center">

MY DOG

</div>

functions very much like a definite description, in that there must be a resource situation, r, in which there is one dog, d, that, at the appropriate time t, belongs to KD, that is to say

$$r \models \ll \text{dog}, d, t, 1 \gg \wedge \ll \text{owns}, KD, d, t, 1 \gg$$

and the meaning of this phrase links the utterance situation with that dog.

20 SENTENCE MEANING

Consider an *utterance situation*, u, in which a *speaker*, a_u, utters a sentence, Φ, to a single *listener*, b_u, at a time t_u and a location l_u. The situation u may be part of a larger, *discourse situation*, d. (Otherwise we take $d = u$.) The situation d is part of some (possibly larger) *embedding situation*, e, that part of the world of direct relevance to the utterance. During the utterance, the speaker may refer to one of several *resource situations*. The utterance u will determine a *described situation*, $s_u = s_u(\Phi)$.

For definiteness, take the utterance of the single assertive sentence

$$\Phi : \quad \textit{Keith bought a dog.}$$

Factors about the utterance situation, u, should, if this utterence is to succeed in imparting to the listener the information Jan wants to convey, determine a unique individual $k = c_u(\text{KEITH})$ such that for some resource situation $r_k = c_u^{res}(\text{KEITH})$:

(1) $r_k \models \ll \text{person}, k, t_k, 1 \gg \wedge \ll \text{named}, k, \text{KEITH}, t_k, 1 \gg$

(2) k is the only such individual in r_k

where, according to the overall context, either t_k includes t_u or else t_k includes the time t introduced below.

The meaning of the word 'BOUGHT' relates Jan's usage of this word to a relation 'buys', and the usage of the past tense determines that for some time, t, preceding t_u:

(3) $s_u \models \ll \text{buys}, k, p, t, 1 \gg$

where p is as below.

Finally, for the utterance to be true, there must be an individual p and a resource situation $r_p = c_u^{res}(\text{A DOG})$ such that

(4) $r_p \models \ll \text{dog}, p, t, 1 \gg$

(5) $s_u \models \ll \text{buys}, k, p, t, 1 \gg .$

Let's examine the various components of this analysis, beginning with the resource situation r_k. In making her utterance the way she does, Jan presumably assumes that the listener has some (possibly quite miminal) information about r_k, in particular the information that there is an individual k' such that:

(6) $r_k \models \ll \text{person}, k', t_k, 1 \gg \wedge \ll \text{named}, k', \text{KEITH}, t_k, 1 \gg$

(7) k' is the only such individual in r_k.

It is not necessary that the listener can identify the k' here with the individual, k, Jan is referring to, though Jan might well be assuming the listener has such knowledge.

The *assumption* by Jan of a certain shared knowledge about the resource situation, r_k, is what enables her to use the name 'KEITH' the way she does. Though she herself may well have a very extensive stock of information about r_k, the listener's knowledge could be quite meager. It might only amount to the two items (6) and (7) above. More likely, the listener's knowledge of the rules governing English proper names would allow him to conclude in addition that

(8) $r_k \models \ll \text{male}, k', t_k, 1 \gg .$

A fairly cursory knowledge of Jan's family circumstances might also provide the listener with the further information

(9) $r_k \models \,\ll\text{husband-of}, k', a_u, t_k, 1 \gg .$

The listener then, requires only quite minimal knowledge about r_k in order for Jan's usage of the word 'KEITH' to be informational. But notice that Jan too actually needs to draw on very little information about r_k in order to make this utterance.

Though more traditional, AI-oriented approaches to this issue might refer to r_k as a 'Keith-*file*', this would be misleading, in that use of the word 'file' suggests a list of facts about Keith, a list to which the speaker and listener may each add new information, and through which they each search for information. This is not at all what is meant here. Rather, associated to this guy Keith is a certain situation r_k, and as the occasion demands, different people can draw on various items of information about r_k (in terms of our ontology, we might say they can utilize various compound infons, σ, such that $r_k \models \sigma$). The situation r_k remains constant here, a fixed situation, part physical and part abstract, intimately associated with Keith. We could, if we wished, refer to the collection of infons that the speaker and listener each know to be supported by r_k, as the speaker's 'Keith-file' and the listener's 'Keith-file', respectively. In which case *these files* are dynamic entities that change with time. But the *situation* r_k remains fixed.

Turning next to Jan's utterance of the word BOUGHT, in keeping with our overall treatment of relations in this study, assume that both the speaker and the listener associate with this word the same relation, *buys*, a complex, structured object relating a number of arguments.

Now look at Jan's usage of the phrase A DOG. This is likewise linked to a certain situation r_p, a situation associated with the dog Keith bought, a situation that supports, among other things, the fact of that dog being a dog.

Notice that Jan may or may not have any direct knowledge of just *which* dog Keith bought. All we can say as theorists is that there must be such a p and an associated resource situation r_p. The use of the indefinite article leaves aside all questions as to the identity of the dog.

Thus, Jan's utterance refers to a situation in which there are two individuals, k and p. The individual k is referred to directly in the utterance, and facts about the resource situation r_k are required in order for the utterance to convey the information Jan intends of it (assuming the obvious intent, discussed below). The individual p is not referred to in the utterance, nor is the resource situation r_p. There must of course *be* such an individual, and associated with that individual there will *be* a resource situation, r_p. But Jan's utterance does not identify them the way it does the individual k and the situation r_k. This distinction will be highlighted in the following discussion about the informational content of the utterance.

Turning now to that informational content, in the most straightforward case, the item of information that Jan wants to convey by means of her utterance is what

is referred to as the *propositional content* of the utterance. This is the proposition

$$s_u \models \exists \dot{p} \exists \dot{t} \ll \text{buys}, k, \dot{p}, \dot{t}, 1 \gg$$

where $\dot{p}$ is a parameter for a dog and $\dot{t}$ is a parameter for a time period prior to t_u, for example $\dot{t} = \mathit{TIM}_{56} \restriction \ll \prec, \mathit{TIM}_{56}, t_u, 1 \gg$.

Notice that this content has as constituents the described situation, s_u, the individual k, and the relation *buys*. The speaker makes explicit reference both to the individual k and the relation *buys*. The described situation, s_u, is not referred to in the utterance. Rather the speaker's connections put s_u into the propositional content. Neither the actual time of the buying nor the actual dog bought get into the propositional content.

Contrast this with an utterance of the sentence

$$\Psi : \textit{Keith bought the dog.}$$

Here the propositional content is

$$s_u \models \exists \dot{t} \ll \text{buys}, k, p, \dot{t}, 1 \gg .$$

This time the particular dog, p, gets into the propositional content as an articulated constitutent of the utterance. But where does this individual come from? The utterance of this one sentence alone does not serve to identify p. Rather some previous utterance, or some embedding circumstance, has to pick out the particular dog Jan refers to. Normal language use requires that an utterance of sentence Ψ is indeed either preceded by an utterance that supplies the individual, p, referred to in Ψ by the phrase 'THE DOG', or else the utterance is made in a circumstance where other factors serve to make this identification, such as the utterance being made while the speaker and listener are jointly viewing a scene in which there is exactly one dog.

Notice that the fact that the person, k, referred to in any veridical utterance of Φ, is named 'Keith,' does not contribute directly to the meaning of Φ, nor does the fact that the individual bought, p, is a dog, although these are part of the *meanings* of the two words concerned. Rather these facts are reflected in our framework by virtue of the way parameters operate. Any veridical utterance of Φ is constrained to have the word 'KEITH' refer to a person named 'Keith' and the word 'DOG' refer to a dog.

The propositional content of the utterance of an assertive sentence is our theory's way of getting at the principal item of information that, under normal circumstances, the speaker intends to convey by the utterance. As such it is closely related to the meaning of the sentence, which we turn to next.

The abstract meaning of a sentence is an extrinsic feature of the sentence, independent of any particular context of utterance. For the present example, the *abstract meaning* of the sentence Φ is an abstract link, $\mathcal{M}(\Phi)$, that connects the situation-type

$$U = [\dot{u} \mid \dot{u} \models \ll \text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \land$$
$$\ll \text{utters}, \dot{a_u}, \Phi, \dot{l_u}, \dot{t_u}, 1 \gg \land$$
$$\ll \text{refers-to}, \dot{a_u}, \text{KEITH}, \dot{k}, \dot{l_u}, \dot{t_u}, 1 \gg]$$

and the situation-type

$$E = [\dot{s} \mid \dot{s} \models \exists \dot{p} \exists \dot{t} \ll \text{buys}, \dot{k}, \dot{p}, \dot{t}, 1 \gg]$$

where $\dot{k}$ is a parameter for a person named 'Keith', $\dot{p}$ is a parameter for a dog, and $\dot{t}$ is a parameter for a time period preceding $\dot{t_u}$, say $\dot{t} = TIM_5 \lceil \ll \prec, TIM_5, \dot{t_u}, 1 \gg$.

The meaning-in-use of Φ, $\|\Phi\|$, should link any particular utterance of Φ with the fact of the world (or relevant part thereof) being the way Φ says it should be. That is to say it is the relation between situations u and v, induced by $\mathcal{M}(\Phi)$, such that:

$$u\|\Phi\|v \quad \text{if and only if} \quad [u : U] \And [s_u(\Phi) \subseteq v] \And [v : E]$$

where $U[\mathcal{M}(\Phi)]E$.

The parametric, compound infon that determines the type E above is known as the *descriptive content* of Φ, denoted by $\mathcal{C}(U)$. That is:

$$\mathcal{C}(U) = \exists \dot{p} \exists \dot{t} \ll \text{buys}, \dot{k}, \dot{p}, \dot{t}, 1 \gg .$$

It is denoted by $\mathcal{C}(U)$ rather than $\mathcal{C}(\Phi)$, since the descriptive content is really a function of the type of an utterance of Φ, rather than the sentence Φ. In particular, it is U that provides the link between the word 'KEITH' in Φ and the parameter $\dot{k}$ in $\mathcal{C}(U)$. In practice, however, this distinction is often blurred: $\mathcal{C}(\Phi)$ being understood to mean the descriptive content of Φ *with respect to the type of an utterance of* Φ.

The descriptive content captures the 'information template' that produces the principal item of information conveyed by any veridical utterance of the sentence (that is to say, the information about the described situation that consitutes the propositional content of the utterance) when the various parameters are anchored to the appropriate objects.

Thus the descriptive content provides an intermediate layer between the syntactic unit Φ and the propositional content of an actual utterance of Φ. It allows us to account for Barwise and Perry's *efficiency of language*; in this case the fact that the same sentence Φ can be used over and over again, by different speakers, referring to different Keiths and different dogs, to convey the 'same' item of information each time, namely that the particular Keith referred to bought some dog. The descriptive content is thus a uniformity across all propositional contents of all veridical utterances of Φ.

Notice that the descriptive content transcends the actual syntax of Φ. Rather it gets at something deeper than syntax. For example, translations of Φ into different languages will all have the same descriptive content. The sentence is a string of symbols, constructed in accordance with certain rules; the descriptive

content is a parametric, compound infon, a genuine object in our ontology. A veridical utterance of the sentence provides anchors for the various parameters in the descriptive content, and the result is that item of information about the described situation that constitutes the propositional content of the utterance.

In other words, if $\sigma = \mathcal{C}(\Theta)$ is the descriptive content of an assertive sentence Θ, then for any utterance, u, of Θ, if f_u denotes the anchor that u provides for the parameters in σ, then the propositional content of this utterance is

$$s \models \sigma[f_u]$$

where $s = s_u(\Theta)$ (the described situation).

The anchors for the parameters in $\mathcal{C}(\Theta)$ are clearly related to what we have called the speaker's connections for some of the words that go to make up Θ. If α is a word or phrase in Θ and if the speaker's connections link α to the individual $c_u(\alpha)$, and if $\dot{a}$ is the parameter in $\mathcal{C}(\Theta)$ that corresponds to α, then

$$f_u(\dot{a}) = c_u(\alpha).$$

The descriptive content of a sentence is essentially a parametric object. According to the convention adopted in this article that there are no parameters for relations, any descriptive content will involve relations, but by and large all other constituents will be parameters. Exceptions would be where a word or phrase has a fixed meaning, independent of context of utterance, such as 'Earth' or 'Mars' or '*Principia Mathematicae*'. (Though it is possible to argue for the context dependency of each of these.)

Further discussion of sentence meaning requires the concept of 'impact' of an utterance, introduced later.

21   ATTRIBUTIVE USES OF DEFINITE AND INDEFINITE DESCRIPTIONS

Hitherto our discussion of both definite and indefinite descriptions has been in terms of what is generally known as the *referential* use, where the description is used to refer to a particular individual — a uniquely specified individual in the case of a definite description, not uniquely identified in the case of an indefinite description. There are, however, other uses of noun phrases.

Starting with definite descriptions, consider the following sentences, all involving one of our original examples of a definite description:

1. *The President of the United States lives in Washington.*

2. *George Bush is the President of the United States.*

3. *George Bush, the President of the United States, lives in Washington.*

Sentence 1 has two quite distinct readings. When the noun phrase is used referentially, to refer to the particular individual who happens to be the President of the United States at the relevant time, the propositional content of the utterance (u) is of the form

$$s_u \models \,\ll\text{lives-in}, p, c, t_u, 1 \gg$$

where

$$p = c_u(\text{THE PRESIDENT OF THE UNITED STATES})$$

and

$$c = c_u(\text{WASHINGTON}).$$

[In fact c is the city of Washington DC (a situation in our ontology) and, if the utterance is made at the time of writing this article, in 2004, p is President George Bush (an individual in our ontology).]

In using the phrase 'THE PRESIDENT OF THE UNITED STATES', the speaker makes use of a resource situation r, possibly the whole of the United States, to identify the particular individual p, that is to say, to determine the value of the function c_u for this particular noun phrase.

The second reading of sentence 1 is the *attributive* reading, where the sentence has a meaning roughly the same as:

> *The President of the United States, whoever it is, always lives in Washington.*

Under this reading, the phrase 'THE PRESIDENT OF THE UNITED STATES' does not refer to a particular individual, but rather to the general property of *being* a President of the United States. Under this reading, an utterance, u, of sentence 1 expresses a constraint, and the propositional content of u is:

$$s_u \models (S \Rightarrow T)$$

where

$$S = [\dot{s} \mid \dot{s} \models \,\ll\text{US-President}, \dot{p}, \dot{t}, 1 \gg]$$

$$T = [\dot{s} \mid \dot{s} \models \,\ll\text{lives-in}, \dot{p}, c, \dot{t}, 1 \gg]$$

where s_u, the described situation, is probably the entire United States, and where c is the city of Washington DC, as before.

Turning now to sentence 2, there is clearly no meaningful reading of this sentence in which the definite description 'THE PRESIDENT OF THE UNITED STATES' is used referentially, since that would just amount to the triviality

George Bush is George Bush.

Under the attributive reading, the phrase 'THE PRESIDENT OF THE UNITED STATES' determines a predicate, the property of being the President of the United States, and the propositional content of an utterance, u, of sentence 2 is:

$$s_u \models \ll \text{US-President}, p, t_u, 1 \gg$$

where $p = c_u(\text{GEORGE BUSH})$ is the individual (President) George Bush.

Finally, sentence 3 provides an example of an *appositive* use of a definite description. Uttering the phrase 'THE PRESIDENT OF THE UNITED STATES' as part of sentence 3 provides additional information about the individual named 'GEORGE BUSH' referred to by the subject of the sentence. Among other things it serves to specify precisely *which* George Bush the speaker has in mind.

The propositional content of an utterance, u, of sentence 3 will be:

$$s_u \models \ll \text{lives-in}, p, c, t_u, 1 \gg \wedge \ll \text{US-President}, p, t_u, 1 \gg$$

where $p = c_u(\text{GEORGE BUSH})$ is the individual (President) George Bush and $c = c_u(\text{WASHINGTON})$ is the city of Washington DC

Notice that, in the case of the attributive reading of sentence 1, the definite description picks out a *function*, $\mathcal{P}$, the function that associates with each time t the current President of the United States at time t, and the propositional content amounts to the claim that for any time t:

$$s_u \models \ll \text{lives-in}, \mathcal{P}(t), c, t, 1 \gg .$$

A particularly striking example of such a *functional* use of a definite description arises in connection with the so-called Partee Puzzle. This purports to show that it is not always possible to substitute equals for equals, by considering the pair of sentences:

- *The temperature is ninety.*

- *The temperature is increasing.*

A naive substitution of equals for equals in this pair of sentences produces the absurdity

- *Ninety is increasing.*

Of course, such a substitution is not possible, and the question then is "Why not?"

The answer is that in the first sentence, the definite description 'THE TEMPERATURE' is used referentially to refer to the actual temperature at the time of utterance, whereas in the second sentence the same definite description is used functionally to refer to the function that links the time to the temperature at that time.

Broadly similar remarks to all the above can be made about indefinite descriptions. For example, paralleling the three examples of sentences involving definite descriptions, the following exhibit the same overall features:

1. *A Scotsman wears a kilt.*

2. *Angus is a Scotsman.*

3. *Angus, a Scotsman, lives in Oxford.*

22   IMPACT

Another feature of sentence utterance considered in situation semantics is the *impact*. Every sentence utterance has an impact, regardless of whether that sentence is assertive or not.

As before, u is an *utterance situation*, in which a *speaker*, a_u, utters a sentence, Φ, to a single *listener*, b_u, at a time t_u and a location l_u. In general, u is part of a larger, *discourse situation*, d. The discourse, d, is part of a (possibly larger) *embedding situation*, e, that part of the world of direct relevance to the discourse. The sentence Φ is not necessarily an assertive sentence.

Denote by t_u^+ some time following the utterance. At the current level of generality, it is not possible to say exactly how much later than t_u this time t_u^+ is, nor what its duration is. It depends very much on context. In the case of a command that should be obeyed immediately, t_u^+ could be an interval immediately following the utterance, the time when the command should be obeyed. In the case of the utterance, u, made as part of an ongoing discourse, d, a common value for t_u^+ will be t_v, where v is the next sentence utterance in the discourse.

The *impact* of u, $\mathcal{I}(u)$, consists of compound infons, σ, built up from basic infons of the form $\ll R, \ldots, t, i \gg$, where $t \preceq t_u^+$, such that:

- $e \models \sigma$

- $u \triangleright [e \models \sigma]$ (more precisely, $u \triangleright \{\ll \models, e, \sigma, 1 \gg\}$).

Intuitively, the impact of an utterance is the (relevant) change in the embedding situation that the utterance brings about. (The parenthetic use of the word 'relevant' here is to exclude such 'irrelevant' changes as the movement of molecules in the air caused by the utterance, etc.). For example, in the case where Φ is an assertive sentence, where the speaker (a_u) has the straightforward intention of conveying to the listener (b_u) the information comprising the propositional content, p, of u, and where this intention is fulfilled (i.e. the listener does acquire that information), $\mathcal{I}(u)$ contains the infon

$$\ll \text{has-information}, b_u, p, t_u^+, 1 \gg .$$

Notice that the speaker's intention here is in terms of the listener having certain information. We do not refer to the belief or knowledge of the listener. To do so would be quite inappropriate. There are many cases where information is conveyed without the listener, or indeed the speaker, either knowing or believing that information. For example, the speaker or listener might be a computer,

which can acquire and dispense vast amounts of information but which neither believes nor knows anything. Or again, one suspects that a great many television newsreaders neither know nor believe all the information they read to camera. Conveying information does not require belief or knowledge of that information, though it does of course require that the speaker *has* that information.

One obvious property of the impact is that it serves to distinguish between certain of Searle's five illocutionary acts.

In the case of a directive, one might imagine that the impact will include the listener's act of compliance or non-compliance to the command.

For example, if Naomi says to Melissa

$$Close\ the\ door$$

then in the case where Melissa obeys the command, the impact of this utterance, u, could include the infon

$$\ll\text{closes, Melissa, } D, l_D, t_u^+, 1 \gg$$

where $D = c_u(\text{THE DOOR})$, or, if Melissa does not obey the command, it could include the infon

$$\ll\text{closes, Melissa, } D, l_D, t_u^+, 0 \gg.$$

However, this is not quite right. For as far as the act of communication is concerned, the utterance of a directive has succeeded if, as a result of the utterance, the listener forms the intention to perform the requisite action. Some other factor(s) might frustrate the fulfillment of this intention, but that is independent of the success or failure of the speech act.

Accordingly, what the impact of Naomi's utterance, u, *will* contain is either the infon

$$\ll\text{of-type}, Melissa, I(D), t_u^+, 1 \gg,$$

or the infon

$$\ll\text{of-type}, Melissa, I(D), t_u^+, 0 \gg$$

where $I(D)$ is the object-type of having an intention to close the door D.

Whether the directive is in fact obeyed or not is not reflected in the impact. The impact is concerned exclusively with the effects of the utterance *as a speech act*. But notice that it is the nature of a directive that exactly one of the above two intentional-state infons must be in the impact. There is no 'neutral' position, whereby the impact is void of any infon pertaining to Melissa's intention regarding the closing of the door.

That is to say, one feature of a directive is that if u is an utterance of a command 'Do $\mathcal{K}$' then precisely one of

$$\ll\text{of-type}, b_u, I(\mathcal{K}), t_u^+, 1 \gg$$

or the infon

$$\ll \text{of-type}, b_u, I(\mathcal{K}), t_u^+, 0 \gg$$

is in $\mathcal{I}(u)$, where $I(\mathcal{K})$ is the object-type of having an intention to perform the action $\mathcal{K}$.

For a commisive, the impact will be the formation by the speaker of the intention to perform some future action. Thus if Melissa says to Naomi

I will close the door

then the impact of this utterance, u, will include the infon

$$\ll \text{of-type, Melissa, } I, t_u^+, 1 \gg$$

where I is the object-type of having an intention to close the door.

The impact of a declarator will be that act brought about by the utterance. Thus, if Keith says to Dale:

You are now in charge of the department

then the impact of this utterance, u, includes the infon

$$\ll \text{in-charge-of, Dale, D, } t_u^+, 1 \gg$$

where $D = c_u(\text{THE DEPARTMENT})$.

The above examples illustrate the prominent and characteristic role played by the impact in an utterance of a directive, commisive, or declarator. The impact is not such a prominent feature of the utterance of an assertive or an expressive.

Indeed, at the present level of treatment, the impact does not distinguish between assertives and expressives. Both assertives and expressives are considered purely in terms of the information conveyed, in the sense of propositional content.

But this does not mean that the utterance of assertive or expressive sentences does not have an impact, as the following discussion indicates.

From the point of view of discourse analysis, one important feature of the impact is that it enables us to handle the way that, as a discourse proceeds, referents are supplied for subsequently used pronouns and otherwise ambiguous proper names.

For instance, consider the example mentioned earlier, where a speaker says:

The farmer bought a donkey. He beat it.

The discourse, d, here comprises two sentences. Let u_1 be the utterance of the first sentence, u_2 that of the second. The embedding situation, e, extends the discourse and includes the farmer and a donkey. The utterance u_1 introduces the two objects

$$F = c_{u_1}(\text{THE FARMER}) \quad \text{and} \quad D = c_{u_1}(\text{A DONKEY})$$

into the discourse situation. Then, the utterance u_2 may take

$$c_{u_2}(\text{He}) = F \quad \text{and} \quad c_{u_2}(\text{It}) = D.$$

In this case, the impact of u_1, $\mathcal{I}(u_1)$, includes the infons

$$\ll\text{salient-in}, F, d, t_{u_1}^+, 1 \gg$$

$$\ll\text{salient-in}, D, d, t_{u_1}^+, 1 \gg.$$

In general, if u is an utterance of a word/phrase/sentence, α, such that one or more of $c_u(\alpha), c_u^{res}(\alpha)$, or (in the case where α is a sentence) $s_u(\alpha)$ is defined, then if a is any one of these objects, we have

$$\ll\text{salient-in}, a, d, t_u^+, 1 \gg \in \mathcal{I}(u)$$

which implies that

$$e \models \ll\text{salient-in}, a, d, t_u^+, 1 \gg.$$

Moreover:

- if $a = c_u(\alpha)$ is an individual that is referred to by α in u, then

$$\ll\text{refers-to}, a_u, \alpha, a, t_u, 1 \gg \in \mathcal{I}(u)$$

- if $a = c_u(\alpha)$ and $r = c_u^{res}(\alpha)$, then

$$\ll\text{resource-for}, r, a, t_u, 1 \gg \in \mathcal{I}(u)$$

- if α is a sentence and $s = s_u(\alpha)$, then

$$\ll\text{speaking-about}, a_u, \alpha, s, t_u, 1 \gg \in \mathcal{I}(u).$$

The function $\mathcal{I}$ is such that, if u_1 is a subutterance of u_2, then $\mathcal{I}(u_1) \subseteq \mathcal{I}(u_2)$, whenever both these sets are defined.

Consider now the following discourse (set in the late 1980s):

Ed : *Did you see the 49ers game yesterday?*

Jan: *Yes, I think Montana is wonderful.*

Ed : *Yes, his last pass to Rice was amazing.*

Let u_1 be the first utterance, that of Ed, let u_2 be the second, Jan's, and let u_3 be Ed's final utterance. Let t_1, t_2, t_3 be the time intervals corresponding to each of these utterances, respectively, and let Φ_1, Φ_2, Φ_3 be the three sentences uttered.

The impact of u_1 includes the introduction into the discourse situation of the San Fransisco 49ers, SFO, as the resource situation, and

$$G = \text{yesterday's 49ers game}$$

as the described situation, the focus of the ensuing discourse.

Thus, $\mathcal{I}(u_1)$ includes the following infons:

$$\ll \text{salient-in, SFO, } d, t_1^+, 1 \gg$$

$$\ll \text{salient-in, } G, d, t_1^+, 1 \gg$$

$$\ll \text{refers-to, Ed, THE 49ERS GAME, } G, t_1, 1 \gg$$

$$\ll \text{resource-for, SFO, } G, t_1, 1 \gg$$

where in this case t_1^+ denotes the time interval comprising both t_2 and t_3.

In asking the question he does, Ed is assuming that Jan is familiar with the 49ers, that she has access to the situation SFO. In making the initial 'Yes' response she does, Jan confirms that she does indeed have such access. Otherwise, a more appropriate response would have been "Who?" Likewise, her initial "Yes" shows that she is also familiar with the situation G, since she would otherwise have responded "No".

Now, among the facts that Jan knows about the situation SFO is that the quarterback is named Joe Montana. Thus, in making her response, u_2, Jan can take

$$c_{u_2}(\text{MONTANA}) = M \quad \text{and} \quad c_{u_2}^{res}(\text{MONTANA}) = \text{SFO}$$

where M is the individual Joe Montana.

In turn now, the impact of u_2 includes the introduction of the individual M into the discourse situation. That is to say, $\mathcal{I}(u_2)$ includes the infon

$$\ll \text{salient-in, } M, d, t_2^+, 1 \gg$$

where t_2^+ denotes the time interval t_3.

So, in making the utterance u_3, Ed can take

$$c_{u_3}(\text{HIS}) = M$$

in order to make his comment on the pass made by Montana to wide-receiver Jerry Rice.

In the absence of Ed's first utterance however, Jan's remark could equally well have been about the State of Montana. It was the utterance of u_1, with its impact including the introduction of the situation SFO into the embedding situation, that prevented any such breakdown in communication due to the ambiguity of the word 'Montana'.

Likewise, Ed's knowledge of the situation SFO included the fact that its star wide-receiver is a man, R say, called 'Rice', and thereby allowed him to take

$$c_{u_3}(\text{RICE}) = R.$$

The success of u_3 (in terms of the conveyance of information) depends upon Jan, the listener, also knowing that the 49ers have a player called 'Rice'. Otherwise, she might have taken the referent of the word 'rice' to be the white, granular substance found on the supermarket shelves, and not the person R that Ed was talking about.

(Well, this is conceivable — the word is ambiguous.) More likely though, Jan's background knowledge of ball games would have forced her to conclude that Ed's use of the word 'rice' must refer to some person by that name, even if she had never heard of that person before. Situation semantics can handle this possibility as well.)

It would be easy to pursue the above investigation to far greater depths. But the intention here is not to carry out a linguistic analysis, rather to indicate how the formal tools of situation theory, including the impact of an utterance, can be used to perform such an analysis.

23   SITUATION SEMANTICS AND SEARLE'S CLASSIFICATION OF SPEECH ACTS

The meaning of an assertive sentence has already been defined and investigated. But what is the meaning of other forms of sentence in the Searle classification, the directives, commisives, declarators, and expressives? The machinery we now have available is not only adequate for dealing with utterances of each of these types, it also provides features that distinguish utterances of one category from those of another.

As before, u is an *utterance situation* in which a *speaker*, a_u, utters a sentence, Φ, to a single *listener*, b_u, at a time t_u and a location l_u.

Let U be the type of an utterance of Φ by a_u to b_u, namely:

$$U = [\dot{u} \mid \dot{u} \models \ \ll \text{speaking-to}, \dot{a_u}, \dot{b_u}, \dot{l_u}, \dot{t_u}, 1 \gg \wedge \\ \ll \text{utters}, \dot{a_u}, \Phi, \dot{l_u}, \dot{t_u}, 1 \gg].$$

Start with the expressives, since from the standpoint of our situation semantics these turn out to be very similar to the assertives.

Suppose that the sentence Φ is an expressive:

$$\text{'I am } \Pi\text{'}$$

where Π is some psychological state, such as sorrow or anger. Let E be the situation-type

$$E = [\dot{s} \mid \dot{s} \models \ll \text{of-type}, \dot{a_u}, B(\Pi), \dot{t_u}, 1 \gg]$$

where $B(\Pi)$ denotes the object-type of being in the state Π. Then $\mathcal{M}(\Phi)$, the abstract meaning of Φ, is the link between the types U and E.

Turning to the meaning-in-use of Φ, this will be a relation linking utterances of Φ (i.e. situations of type U) to situations extending the described situation that are of type E. So one question to answer is what are the possible described situations? The answer is implicit in the nature of an expressive. In uttering an expressive, the speaker, a_u, describes her own state, so that will be the described situation, $s_u(\Phi)$. Then, given situations u and v we shall have

$u\|\text{I AM }\Pi\|v$ if and only if

$$[u : U] \ \& \ [s_u(\Phi) \subseteq v] \ \& \ [v \models \ll\text{of-type}, a_u, B(\Pi), t_u, 1 \gg].$$

In the three remaining categories of utterance, the directives, commisives, and declarators, the main function is not the conveyance of information, as was the case with the assertives and expressives; rather it is the regulatory effect the utterance has on action, either of the speaker or the listener. For such sentences, the impact of the utterance is the most significant feature, not the propositional content.

Consider first the case where the sentence Φ is a directive:

$$\text{`Do } \mathcal{K}\text{'}.$$

Let E be the type

$$E = [\dot{s} \mid \dot{s} \models \quad \ll\text{of-type}, \dot{b}_u, I(\mathcal{K}), \dot{t}_u^+, 1 \gg \wedge \\ \ll \rhd, \dot{u}, (\dot{s} \models \ll\text{of-type}, \dot{b}_u, I(\mathcal{K}), \dot{t}_u^+, 1 \gg), 1 \gg]$$

where $I(\mathcal{K})$ is the object-type of having an intention to perform the action $\mathcal{K}$.

Then the abstract meaning of the sentence Φ, $\mathcal{M}(\Phi)$, is defined to be the link between the two types U and E. The intention here is that the meaning of a directive is that link which, for a given utterance of the directive, connects the utterance with its compliance (in the sense of forming the intention to do as instructed). This explains the second component in the definition of the type E, which we have expressed in an abbreviated fashion for clarity. The meaning must reflect the fact that the intention to perform the action $\mathcal{K}$ that figures in Φ has to arise by way of complying with the directive.

The meaning-in-use of Φ, induced by $\mathcal{M}(\Phi)$, is a relation, $\|\Phi\|$, between utterances, u, of Φ and certain situations v that extend the described situation, $s_u(\Phi)$. Now the situation $s_u(\Phi)$ is identified by features of the utterance itself. For assertives it can be any situation whatever. For expressives the described situation is constrained to be the speaker's state. In the case of a directive, the described situation must be the listener's state. Then for any two situations u and v:

$u\| \text{ Do } \mathcal{K}\|v$ if and only if

$$[u : U]\&[s_u(\Phi) \subseteq v]\&[v \models \quad \ll \text{of-type}, b_u, I(\mathcal{K}), t_u^+, 1 \gg \wedge \\ \ll \rhd, u, (v \models \ll \text{of-type}, b_u, I(\mathcal{K}), t_u^+, 1 \gg), 1 \gg].$$

Suppose now that Φ is a commisive:

$$\text{`I will } \mathcal{K}\text{.'}$$

Let E be the type

$$E = [\dot{s} \mid \dot{s} \models \ll\text{of-type}, \dot{a_u}, I(\mathcal{K}), \dot{t}_u^+, 1 \gg]$$

where again $I(\mathcal{K})$ is the object-type of having an intention to perform the action $\mathcal{K}$.

The abstract meaning of Φ is again defined to be the link between the two types U and E.

Turning to $\|\Phi\|$, if we are given a particular utterance, u, of the commisive Φ, the described situation, $s_u(\Phi)$, will be the speaker's state, and the meaning-in-use of Φ relates the situation u to those situations v extending $s_u(\Phi)$ in which the speaker forms the intention to do as promised in Φ:

$u\|\text{I WILL } \mathcal{K}\|v$ if and only if

$$[u : U] \And [s_u(\Phi) \subseteq v] \And [v \models \ll\text{of-type}, a_u, I(\mathcal{K}), t_u^+, 1 \gg].$$

Finally, suppose Φ is a declarator:

'I declare $\mathcal{K}$'.

Let E be the type

$$E = [\dot{s} \mid \dot{s} \models \ll T(\mathcal{K}), t_u^+, 1 \gg]$$

where $T(\mathcal{K})$ expresses that fact that things are as the utterance of Φ declares them to be. For example, if

$$\mathcal{K} = \text{'You are in charge'}$$

then

$$T(\mathcal{K}) = \text{in-charge}, b_u.$$

Then $\mathcal{M}(\Phi)$ is the link between U and E.

For $\|\Phi\|$, if we are given an utterance u of Φ, then there is no general rule as to what is the described situation, $s_u(\Phi)$. It depends very much on $\mathcal{K}$. In the case of the example just given, $s_u(\Phi)$ will be whatever it is the listener is put in charge of, say, the department. Then, given situations u and v, we have:

$u\|\text{YOU ARE IN CHARGE}\|v$ if and only if

$$[u : U] \And [s_u(\Phi) \subseteq v] \And [v \models \ll\text{in-charge}, b_u, t_u^+, 1 \gg].$$

24   COMPOSITIONALITY

This brief article does not sent out to provide a full-blown account of the way that the meaning of a composite sentence or utterance is built up from the meanings of the various components. Certainly the high degree of context dependency of this process would seem to render as a hopeless dream any kind of development analogous to Tarski's semantics of predicate logic. But the tools described are adequate for an analysis of particular instances of compositionality, so it will be a

useful exercise to investigate two of the simplest, and most basic kinds of example: conjunction and disjunction. We restrict attention to meaning-in-use.

Start with conjunction. Let u be an utterance situation, in which a speaker a_u utters a conjunctive sentence $[\Phi \text{ AND } \Psi]$ to a single listener b_u at a time t_u and a location l_u. In general, u is part of a larger, discourse situation d. The discourse d is part of a (possibly larger) embedding situation e, that part of the world of direct relevance to the discourse. Let u_1 be the utterance situation in which the clause Φ is uttered, u_2 that pertaining to Ψ.

Naively, one might expect that, given assertives Φ and Ψ, the meaning-in-use of the sentence $[\Phi \text{ AND } \Psi]$ is given by

$$u \| \Phi \text{ AND } \Psi \| v \quad \text{if and only if} \quad u_1 \| \Phi \| v \text{ and } u_2 \| \Psi \| v.$$

This is indeed the case, but the superficial resemblance this has to the analogous Tarskian rule obscures some considerable complexity.

Suppose for instance the sentence uttered is:

Sid loves Nancy and she loves him.

Then the above reduction gives

$(*)$ $u \| \text{SID LOVES NANCY AND SHE LOVES HIM} \| v$ if and only if
$u_1 \| \text{SID LOVES NANCY} \| v$ and $u_2 \| \text{SHE LOVES HIM} \| v$.

The first conjunct here is straightforward enough. The speaker's connections should fix two individuals, $S = c_{u_1}(\text{SID})$ and $N = c_{u_1}(\text{NANCY})$, such that (in particular)

$$u_1 \models \quad \ll\text{refers-to}, a_{u_1}, \text{SID}, S, l_{u_1}, t_{u_1}, 1 \gg \wedge$$
$$\ll\text{refers-to}, a_{u_1}, \text{NANCY}, N, l_{u_1}, t_{u_1}, 1 \gg$$

and

$$v \models \quad \ll\text{loves}, S, N, t_{u_1}, 1 \gg .$$

The second clause involves two pronouns, 'SHE' and 'HIM'. The referents for these pronouns must be supplied by the utterance. The most natural case would be where

$$c_{u_2}(\text{SHE}) = N \quad \text{and} \quad c_{u_2}(\text{HIM}) = S$$

and then part of the requirement on v imposed by $(*)$ is

$$v \models \ll\text{loves}, N, S, t_{u_1}, 1 \gg .$$

In this case the impact of the utterance u_1 provides the relevant individuals to act as referents for the pronouns used in u_2. But there are other possibilities. The utterance could pick out other individuals to be referents for these pronouns.

The meaning of disjunctive sentences, $[\Phi \text{ OR } \Psi]$, is similar to conjunctions. Thus:

$$u \| \Phi \text{ OR } \Psi \| v \quad \text{if and only if} \quad u_1 \| \Phi \| v \text{ or } u_2 \| \Psi \| v.$$

Remarks analogous to those made in the case of conjunction apply here as well.

25   QUANTIFICATION

One of the most significant uses of parameters in situation theory arises in the semantics of natural language quantification. For example, let Φ be the sentence

Every logician admires Quine.

Let u be an utterance of Φ. The first question I ask is what is the described situation, $e = s_u(\Phi)$? Well, in the absence of any previously established context this will surely be the world, w, or at least some part of the world that pertains to, and in particular includes, all logicians — say the academic world. In any event, the propositional content of the utterance u will be of the form

$$e \models <compound\ infon>.$$

The question is, just what compound infon occurs here?

The first approach takes as the propositional content of the utterance, u, the proposition:

$$e \models (\forall \dot{p})\ll admires,\ \dot{p}, Q, t, 1 \gg$$

or, more precisely (recall the convention regarding quantification in compound infons):

$$e \models (\forall \dot{p} \in e)\ll admires,\ \dot{p}, Q, t, 1 \gg$$

where $\dot{p}$ is a partameter for a logician, Q is the individual W.V.O. Quine, and t is the present time. (Taking t to be the time of utterance, t_u, would be inappropriately restrictive in this connection. The time interval t will include t_u but have considerably longer duration. The utterance makes no specific reference to time, though it is clearly intended to be about 'the present time' or perhaps 'the present epoch'.)

By virtue of the manner in which quantifiers operate on infons, this means that for any anchor f for the parameter $\dot{p}$ to an object p in e, it must be the case that

$$e \models \ll admires,\ p, Q, t, 1 \gg .$$

In order for f to be an anchor for $\dot{p}$, there must be a resource situation, r, such that:

$$r \models \ll logician,\ p, t, 1 \gg .$$

But there is no requirement that r should be the same situation as e, or indeed bear any particular relation to e. (Though if e is the world, then r will be a subsituation of e, of course.) Indeed, all that is required is that to each p in e to which $\dot{p}$ can be anchored, there will be *some* such resource situation $r = r_p$ that depends on p.

Consider now the sentences

Φ_1 : *Every player touched the ball.*

Φ_2 : *Every player ate a cookie.*

Let u_1 be an utterance of Φ_1, u_2 an utterance of Φ_2.

Starting with Φ_1, the described situation, $s_{u_1}(\Phi_1)$, will be some ball game, say e, and the propositional content of u_1 will be

$$e \models (\forall \dot{p})(\exists \dot{t}) \ll \text{touch}, \dot{p}, b, \dot{t}, 1 \gg$$

where $\dot{p}$ is a parameter for a player, $b = c_u(\text{THE BALL})$, and $\dot{t}$ is a parameter for a time preceding t_{u_1}. The game situation e will provide the resource situation for all the individuals p to which the parameter $\dot{p}$ can be anchored. That is to say, for any anchor f of $\dot{p}$ to an individual p in e, it will be the case that for some time t within the time-span of e:

- $e \models \ll \text{player-in}, p, e, t, 1 \gg$

- $e \models \ll \text{touch}, p, b, t, 1 \gg$.

The resource situation for the fact that t precedes t_{u_1} is, as always, the world:

$$w \models \ll \prec, t, t_{u_1}, 1 \gg$$

since this is the nature of the basic type $\prec$.

Turning now to the second sentence, Φ_2, assuming the players eat the cookies during the game, the described situation, $s_{u_2}(\Phi_2)$, will be the game e, as before, and the propositional content of u_2 will be

$$e \models (\forall \dot{p})(\exists \dot{c})(\exists \dot{t}) \ll \text{eats}, \dot{p}, \dot{c}, \dot{t}, 1 \gg$$

where $\dot{p}$ is a parameter for a person and $\dot{t}$ is a parameter for a time preceding t_{u_2}, much as before, and where $\dot{c}$ is a parameter for a cookie. (The reading of Φ_2 whereby every player eats the *same* cookie is too implausible to consider; rather, assume that to each player there corresponds a cookie which that player, and only that player, eats.)

Clearly, there is no reason to suppose the game situation e supports the facticity of any particular individual being a cookie. Nor is it necessarily the case that every cookie eaten by some player is of the same variety, with its cookieness being supported by one and the same resource situation. Rather, for each individual p in e to which $\dot{p}$ may be anchored and each corresponding time t to which $\dot{t}$ is anchored, and for which, therefore

$$e \models \ll \text{player-in}, p, e, t, 1 \gg$$

there will be an individual c and a resource situation r_c, such that

$$r_c \models \ll \text{cookie}, c, t, 1 \gg .$$

Given the assumption that the players eat the cookies during the game e, then the cookie c will be a constituent of e. But this is not necessarily the case. The cookies could be eaten at some other time. For instance, they could be eaten in the locker-room after the game is over at some time t' preceding t_{u_2} . To be definite, consider the case where a previous utterance has established, by way of its impact, a speaker's connection to a time t' when the cookies were eaten. Then the described situation e' will be a situation different from the game e, and the propositional content of u_2 will be:

$$e' \models (\forall \dot{p})(\exists \dot{c}) \ll \text{eats}, \dot{p}, \dot{c}, t', 1 \gg .$$

Whatever the described situation turns out to be, the two points to notice are, firstly, that the described situation may or may not provide the scope and resource situation for the quantified parameters, and secondly, the resource situation for an instance of the quantifier $(\exists \dot{c})$ is not necessarily the same as that for the instance of $(\forall \dot{p})$ to which it corresponds.

In the case where the cookies are eaten during the game, then the described situation provides the scope of the quantifier $(\forall \dot{p})$ and the resource situation for each anchor of $\dot{p}$ being a player in e. The described situation also provides the scopes for the quantifiers $(\exists \dot{c})$ and $(\exists \dot{t})$, but for neither of these quantifiers does it provide the appropriate resource situation.

If, on the other hand, the cookies are eaten at some other time determined by the speaker's connections associated with some prior utterance, then the described situation provides the scope for the quantifier $(\forall \dot{p})$ but not the resource situation for any anchor of $\dot{p}$ being a player in the game.

Thus, the theory places no restrictions on the possible scope of quantifiers or on the situations that can provide a resource for the anchor of a particular parameter. It is up to the speaker to ensure that the context of utterance provides the right connections to the scope of any quantifier and to the appropriate resource situations, where relevant. In the case of a cookie, this is clearly of little importance, at least in the majority of cases. But establishing the relevant game situation e and whether the cookies were eaten during the game or at some other time is critical to the success of the utterance as a conveyance of information. Situation semantics allows for, and reflects, all possibilities, but leaves the responsibility for effective communication where it belongs — with the speaker.

So far we have considered just two kinds of quantifiers, *for all* and *there exists*. In order to handle other quantifiers, some further development of the situation-theoretic framework is necessary.

One solution is to enlarge the collection of compound infons by introducing various generalized quantifiers. For example, we could allow the following constructions to figure as compound infons:

$$(M \dot{x} \in u)\sigma \quad \text{and} \quad (F \dot{x} \in u)\sigma$$

where σ is a compound infon, where $(M\dot{x} \in u)$ denotes 'for most $\dot{x}$ in u', and where $(F\dot{x} \in u)$ denotes 'for few $\dot{x}$ in u'. Some form of definition of what these quantifiers actually mean would then be necessary of course.

An alternative approach is to regard quantifiers (at least those that arise explicitly in natural language) not as operators acting on infons, but rather as relations within the theory's ontology; in particular, as relations between types. Thus, for example, among the relations we might have the basic five-place relations $\forall, \exists, M, F$, and then the following would be infons:

$$\ll \forall, u, S, T, l, t, i \gg \quad \ll \exists, u, S, T, l, t, i \gg$$
$$\ll M, u, S, T, l, t, i \gg \quad \ll F, u, S, T, l, t, i \gg$$

where u is a set or situation and S and T are one-place types.

The first of these is the informational item that: if $i = 1$ then *all* objects in u of type S are of type T, and if $i = 0$ then it is not the case that all objects in u of type S are of type T (at location l and time t).

The second is the informational item that: if $i = 1$ then *there is* an object in u of type S that is of type T, and if $i = 0$ then there is no such object (at l, t).

The third is the information that: if $i = 1$ then *most* objects in u of type S are of type T, and if $i = 0$ then this is not the case (at l, t).

Finally, the fourth infon is the informational item that: if $i = 1$ then *few* objects in u of type S are of type T, and if $i = 0$ then this is not the case (at l, t).

Since quantification is now of the infonic form

$$\ll Q, u, S, T, l, t, i \gg$$

a situation is required in order to obtain a proposition

$$e \models \ll Q, u, S, T, l, t, i \gg$$

so the quantification is situated in, and hence restricted to, e.

Using this new framework, let's take a second look at the two previous examples. The first of these is an utterance u_1 of the sentence

Φ_1 : *Every player touched the ball.*

Under the new framework, the analysis of this utterance goes as follows. As before, the described situation, $s_{u_1}(\Phi_1)$, is the game, say g. Let S, T be the following object-types:

$$S = [\dot{p} \mid g \models \quad \ll \text{player-in}, \dot{p}, g, 1 \gg]$$
$$T = [\dot{p} \mid g \models \quad (\exists \dot{t}) \ll \text{touches}, \dot{p}, b, \dot{t}, 1 \gg]$$

where $\dot{p}$ is a parameter for a person, $\dot{t}$ is a parameter for a time prior to t_{u_1} and $b = c_{u_1}(\text{THE BALL})$. Then the propositional content of the utterance u_1 is:

$$g \models \ll \forall, g, S, T, 1 \gg .$$

Notice that the use of the same parameter $\dot{p}$ in the two type-abstractions was in order to help the reader. In practice, since the abstraction parameter in a type-abstraction becomes 'absorbed', leaving solely an 'argument role', it does not matter which parameter is used in each abstraction. Rather, it is the nature of the relation $\forall$ that it links the argument roles of the two types.

One further remark that needs to be made at this juncture concerns the quantification of the time parameter $\dot{t}$ in the definition of the type T. This was done using the quantification mechanism for forming compound infons, rather than in terms of our new quantifier framework. This reflects the fact that an unarticulated quantification over time that arises by virtue of verb tense, is what might be called a 'structural' quantification. That is to say, verb tense mechanisms are part of the basic structure of language that our ontological framework is intended to handle: our ontology includes temporal locations and quantification over temporal locations in compound infons, and verb tense relates directly to this temporal aspect of our framework. Such implicit quantification is not at all the same as an articulated quantification, even one over time, such as an utterance u_1' of the sentence

$$\Phi_1' \; : \; \textit{Every player touched the ball many times.}$$

In this case, the analysis would be as follows.
Let M be a 'many' quantifier. Let T_b be the type

$$T_b = [\dot{t} \mid g \models \ll \text{touch}, \dot{p}, b, \dot{t}, 1 \gg]$$

where $\dot{t}$ is a parameter for a time prior to $t_{u_1'}$ and $\dot{p}$ is a parameter for a person. T_b is the parametric type of all instances at which some person touches b ($= c_{u_1'}(\text{THE BALL})$) during the course of the game g.

T_b is a parametric type with parameter $\dot{p}$, so we can form the type

$$T = [\dot{p} \mid g \models \ll M, g, TIM_1, T_b, 1 \gg]$$

the type of all persons for which there are many instances in g at which that person touches b. Then the propositional content of u_1' is:

$$g \models \ll \forall, g, S, T, 1 \gg$$

where S is as before.

Notice that the present framework allows for a quantifier such as 'for many' to be defined locally. In the case of the above example, the 'many' quantifier M could be specially tailored to ball games. This is a strong argument in favor of treating quantification as a relation within the ontology, rather than as part of the underlying framework. Indeed, we may use our framework to *investigate* such quantifiers.

The second of our two original examples is an utterance u_2 of

$$\Phi_2 \; : \; \textit{Every player ate a cookie.}$$

Let e be the described situation, $s_{u_2}(\Phi_2)$, whether this is the game g or some other situation. Let T_d be the type

$$T_d = [\dot{c} \mid e \models (\exists \dot{t}) \ll \text{eats}, \dot{p}, \dot{c}, \dot{t}, 1 \gg]$$

where $\dot{t}$ is a parameter for a time preceding t_{u_2}, $\dot{c}$ is a parameter for an edible individual, and $\dot{p}$ is a parameter for a person. Thus T_d is the type of all edible individuals that, in the situation e, are eaten at some time prior to t_{u_2} by some person. Noting that T_d is a parametric-type with parameter $\dot{p}$, let T_p be the type

$$T_p = [\dot{p} \mid e \models \ll \exists, e, T_c, T_d, 1 \gg]$$

where T_c is the type of a cookie. Thus T_p is the type of all those persons for which, in the situation e, there is a cookie eaten by that person at some time preceding t_{u_2}.

With S as before, the propositional content of u_2 is:

$$e \models \ll \forall, e, S, T_p, 1 \gg .$$

The only question that remains to be answered is what is the described situation, e? The naive answer is that e is simply the situation in which the cookies were eaten. But this does not work here, since the infon in the propositional content of the utterance involves the type S, which is an object-type with grounding situation g, and there is no reason to suppose that the situation in which the cookies were eaten supports an infon that concerns the game situation g. (If e and g coincide there is no problem. This is what happened with the previous example concerning the players touching the ball many times.) So we must look further for our answer.

In fact, the resolution to the problem involves a shift in the way we regard quantification, since the approach we have adopted provides us with a view of quantification that more traditional definitions do not. Given that quantification is essentially a relation (indeed, a quantitative comparison) between two types, the utterance of any sentence involving a quantifier must be *about* those two types, among other things. That is to say, the described situation must include those two types.

Thus in the present example, the described situation, e, must include both the game situation, g, and the situation in which the cookies are eaten, say h. Then what the utterance does is describe a relation between the two situations g and h, namely the quantitative comparison between the individuals in g that are players and the individuals in h that ate a cookie. In this case, the fact that all individuals of the former type are of the latter type.

Notice that, although this was not the original aim, our investigation has led to an alternative conception of the nature of quantification: it is simply a particular kind of relation between types. Indeed, we can apply this to the 'traditional-style' quantifiers we allow in the formation of compound infons. Although our theory treats these quantifiers as logical operators on compound infons, we may apply our 'quantifiers-as-relations' conceptualization at a meta-theoretic level in order to regard these quantifiers as relations too.

26 NEGATION

There are a number of ways that a sentence can involve negation. The most straightforward of these is verb phrase negation. This is easily handled in situation semantics by means of a polarity change and a possible quantifier switch. For example, let u_1 be an utterance of the sentence

Φ_1 : *John did not see Mary.*

Let e be the described situation, $e = s_{u_1}(\Phi_1)$. Then the propositional content of u is

$$e \models (\forall \dot{t}) \ll \text{sees}, J, M, \dot{t}, 0 \gg$$

where $\dot{t}$ is a parameter for a time prior to t_{u_1}, $J = c_{u_1}(\text{JOHN})$, and $M = c_{u_1}(\text{MARY})$.

There is, however, one question that needs to be answered. What is the described situation e? In the case of an utterance of the positive sentence '*John saw Mary*' there is no problem. In the absence of any context that determined otherwise, the described situation will be the act of John seeing Mary, the situation in which the seeing takes place. In other words, for a positive utterance, in the absence of any other contextual features, the utterance itself determines the described situation. But for a negative utterance this is not the case. There will be a great many situations in which John did not see Mary. Just which one is the speaker referring to?

The answer is that it is up to the speaker to fix the described situation. At least, this is what the speaker's obligation amounts to in our theory's terms. In everyday language, what the speaker must do is ensure that the listener is aware just *what* the utterance is about. To make the utterance u_1 without having set the relevant context results in a failure to communicate. Uttered on its own, without there being either a predetermined described situation or else an obvious 'default' situation, the sentence Φ_1 does not convey information, at least not the information that would be captured by the propositional content. (Most obvious scenarios for such an utterance do in fact supply an obvious default described situation.)

Since there will be a great many situations in which John did not see Mary, in order for the utterance u_1 to convey the right information, the speaker must ensure that some aspect of the context of utterance determines the described situation e. The utterance should convey the same information, in the sense of propositional content, as an utterance of the 'sentence'

♯ *John did not see Mary in e*

where the ♯ indicates a sentence that is not part of normal English (in that one does not normally mention a situation).

The above remarks apply to a great many negative utterances. Of course, in the vast majority of cases the utterance of a positive sentence too is made with reference to a predetermined described situation. Speakers generally speak about

some part of the world. Indeed, this is one of the main motivating factors behind situation theory.

Negated quantifiers are also handled quite easily. For example, let u_2 be an utterance of the sentence

Φ_2 : *Not every student passed the quiz.*

Let $q = c_{u_2}(\text{QUIZ})$, let t_q be the time of taking the quiz q, and let e be the situation comprising the taking of the quiz.

Presumably the speaker is referring to some particular class, c, the class that took the quiz q. Let $\dot{p}$ be a parameter for a person, and let

$$S = [\dot{p} \mid c \models \quad \ll \text{student-in}, \dot{p}, c, t_q, 1 \gg]$$
$$T = [\dot{p} \mid e \models \quad \ll \text{passes}, \dot{p}, q, t_q, 1 \gg].$$

Then the propositional content of u_2 is:

$$d \models \ll \forall, c, S, T, 0 \gg$$

where d is the described situation.

Recalling the discussion of the previous section concerning quantifiers, note that the utterance states a relationship between the type of all students in c and the type of all persons who passed the quiz q, and accordingly the described situation d will extend both c, the grounding situation for type S, and e, the grounding situation for type T.

A seemingly more problematical form of negation is exemplified by an utterance, u_3, of the sentence

Φ_3 : *No sailors were there.*

Assuming u_3 is part of a discourse about a particular dinner party, say d, the natural assumption is that d is the described situation. In which case, how can a proposition of the form

$$d \models \sigma$$

have anything to say about sailors? There are no sailors at the party!

Clearly, it cannot. But a few moments reflection should indicate that this issue has nothing to do with negation. Consider an utterance, u_4, of the positive sentence

Φ_4 : *There is a sailor that was there.*

Though on this occasion a sailor will be a constituent of the party, it is unlikely that this situation will have anything to say about this particular person being a sailor, and so once again the propositional content cannot be of the form

$$d \models \sigma.$$

So what has gone wrong?

The answer is that nothing is wrong, except for the assumption that d is the described situation for an utterance of Φ_3 or Φ_4. For both sentences involve quantifiers, and as we observed in the previous section, an utterance of a quantifier sentence states a relationship between two types, so the described situation must include the grounding situations of those two types.

Both u_3 and u_4 are about sailors: they describe a relation that connects the collection of all sailors and the dinner party d. The grounding situation for the type of all sailors is the world, or at least enough of the world to ground this type. So, if $\dot{t}$ is a parameter for a time preceding the utterance in each case, and if

$$S = [\dot{p} \mid w \models\ \ll\text{sailor}, \dot{p}, \dot{t}, 1 \gg]$$
$$T = [\dot{p} \mid d \models\ \ll\text{present-in}, \dot{p}, d, \dot{t}, 1 \gg]$$

then the propositional content of u_4 is

$$w \models (\exists \dot{t}) \ll \exists, d, S, T, \dot{t}, 1 \gg$$

and the propositional content of u_3 is

$$w \models (\forall \dot{t}) \ll \exists, d, S, T, \dot{t}, 0 \gg,$$

(or possibly

$$w \models (\forall \dot{t}) \ll \text{No}, d, S, T, \dot{t}, 1 \gg,$$

if the quantifier 'No' is regarded as a basic relation in the ontology).

Given our present conception of quantifiers then, even though u_3 or u_4 could be uttered as part of a discourse that until then had concerned the party situation exclusively, once the property of being a sailor is introduced, the so-called described situation is extended to include the grounding situation for being a sailor. Of course, you might object to my calling the resulting situation the *described situation* in this case, and look for another name. On the other hand, given a framework in which a quantifier is interpreted as a relation between two types, rather than some form of logical operator on the second of those types, which is the case in classical logic, then it really is the case that a quantifier utterance *describes* (some feature of) both those types (and hence their grounding situations in the case of object-types): indeed, it *compares* the two types.

It should be noted that the semantics assigned to u_4 is different from the semantics that would be assigned to an utterance u_4' (under the same circumstances and with reference to the same dinner party situation d) of the sentence:

A sailor was there.

In this case, the described situation is indeed the party, d, and the propositional content of the utterance is:

$$d \models \exists \dot{p} \exists \dot{t} \ll \text{present-in}, \dot{p}, d, \dot{t}, 1 \gg$$

where $\dot{p}$ is a parameter for a sailor and $\dot{t}$ is a parameter for a time prior to the time of utterance.

The distinction between u_4 and u_4' amounts to a difference in focus. Uttering the sentence

There is a sailor that was there

makes a definite claim *about the collection of sailors* (namely that at least one of them was at the party). On the other hand, uttering the sentence

A sailor was there

makes a claim *about the party* (namely that among the guests there was at least one sailor).

Of course, none of the above examples involves a *negation* in the sense of classical logic, where negation is a logical operator that acts on well-formed formulas. Rather they are simply utterances of sentences that involve a negative component. As we have seen, this generally requires more emphasis on the specification of the described situation than is the case for utterances where there is no such negative component, but apart from that there was no real difference between positive and negative assertions as far as the above analysis was concerned.

Far more reminiscent of the negation operator of classical logic is sentence *denial*, where a positive assertive sentence is prefixed by a phrase such as 'It is not the case that ...' For example, let u_5 be an utterance of the sentence

Φ_5 : *It is not the case that John saw Mary.*

The starting point of most discussions is to take the phrase *'It is not the case that'* as determining a denial operator that acts on the sentence *'John saw Mary'*. Situation semantics takes a different tack, regarding Φ_5 as a negative version of the sentence

Φ_6 : *It is the case that John saw Mary.*

In both cases, let J be the referent for the name JOHN, M the referent for the name MARY, $\dot{t}$ a parameter for a time prior to the time of utterance.

Let $e_5 = s_{u_5}(\Phi_5)$, $e_6 = s_{u_6}(\Phi_6)$. The propositional content of u_6 is:

$$w \models \ll\models, e_6, (\exists \dot{t}) \ll \text{sees}, J, M, \dot{t}, 1 \gg, 1 \gg .$$

That is to say, the effect of the prefix *'It is the case that'* in an utterance of a sentence *'It is the case that Φ'* is to make the propositional content of the sub-utterance of Φ the infon part of a proposition about the world.

Turning now to u_5, the most natural choice of the propositional content would seem to be:

$$w \models \ll\models, e_5, (\exists \dot{t}) \ll \text{sees}, J, M, \dot{t}, 1 \gg, 0 \gg$$

where the polarity of the world proposition has changed from a 1 in the case of u_6 to a 0 in the case of u_5. Does this accord with our intuitions?

Unravelling the notation a bit, what this proposition says is that

$$(*) \qquad e_5 \not\models (\exists \dot{t}) \ll \text{sees}, J, M, \dot{t}, 1 \gg .$$

Now, in order for a negative utterance to be informational (in the intended manner), the speaker should ensure that the described situation is adequately identified. That is to say, the speaker should make sure that the listener knows what the utterance is *about*. In the present case, e_5 is the John and Mary situation, or something extending it. Since John's seeing Mary is a relevant feature (the speaker talks about it), it ought to be the case that the situation e_5 that constitutes the described situation completely determines whether or not John actually did see Mary or not. That is to say, it should be the case that: either

$$e_5 \models (\exists \dot{t}) \ll \text{sees}, J, M, \dot{t}, 1 \gg,$$

or else

$$e_5 \models (\forall \dot{t}) \ll \text{sees}, J, M, \dot{t}, 0 \gg .$$

Assuming this is the case, then by $(*)$, the propositional content of u_5 should entail the second of these two propositions. This is what we would have expected.

Notice that the above places a restriction on the possible described situation for utterances involving denials. The requirement we have stipulated is considerably stronger than the universally true fact that for any situation s and any infon σ, either $s \models \sigma$ or else $s \not\models \sigma$. A cooperative use of a negative utterance such as u_5 places on the speaker an obligation to ensure that the described situation as understood by the listener (i.e. *what* the listener thinks the utterance is *about*) is sufficiently rich to *decide* the relevant issue, in this case whether John saw Mary or not, one way or the other.

A natural question to ask in connection with sentence denial is how it affects conjunctive and disjunctive sentences. The natural expectation is that there is some form of duality between the two, as occurs in classical logic. And indeed this is the case, given that certain requirements are met.

For example, imagine a discourse between Jan and Ed about last week's 49ers game, g, in which Jan makes the following utterance, u:

> *It is not the case that Joe threw the ball and Roger carried the ball.*

This has a propositional content of the form

$$w \models \ll \models, g, \sigma, 0 \gg$$

where σ is the compound infon

$$(\exists \dot{t}_1) \ll \text{throws}, J, b, \dot{t}_1, 1 \gg \wedge (\exists \dot{t}_2) \ll \text{carries}, R, b, \dot{t}_2, 1 \gg$$

and where $J = c_u(\text{JOE})$, $R = c_u(\text{ROGER})$, and $b = c_u(\text{THE BALL})$.

Unravelling the notation a little, this says the following:

$$(*) \quad g \not\models (\exists \dot{t}_1) \ll \text{throws}, J, b, \dot{t}_1, 1 \gg \wedge (\exists \dot{t}_2) \ll \text{carries}, R, b, \dot{t}_2, 1 \gg .$$

Now, since g is the actual game, either

$$g \models (\exists \dot{t}_1) \ll \text{throws}, J, b, t_1, 1 \gg,$$

or else

$$g \models (\forall \dot{t}_1) \ll \text{throws}, J, b, t_1, 0 \gg,$$

and again either

$$g \models (\exists \dot{t}_2) \ll \text{carries}, R, b, t_2, 1 \gg,$$

or else

$$g \models (\forall \dot{t}_2) \ll \text{carries}, R, b, t_2, 0 \gg.$$

So by $(*)$ it must be the case that at least one of

$$g \models (\forall \dot{t}_1) \ll \text{throws}, J, b, t_1, 0 \gg$$

and

$$g \models (\forall \dot{t}_2) \ll \text{carries}, R, b, t_2, 0 \gg.$$

Hence

$$g \models (\forall \dot{t}_1) \ll \text{throws}, J, b, t_1, 0 \gg \vee (\forall \dot{t}_2) \ll \text{carries}, R, b, t_2, 0 \gg.$$

Reverting back to infon notation, this becomes

$$w \models \ll \models, g, \overline{\sigma}, 1 \gg,$$

where $\overline{\sigma}$ is the compound infon

$$(\forall \dot{t}_1) \ll \text{throws}, J, b, t_1, 0 \gg \vee (\forall \dot{t}_2) \ll \text{carries}, R, b, t_2, 0 \gg.$$

In words:

> It is the case that Joe did not throw the ball or Roger did not carry the
> ball.

Which seems right.

The above example is related to the following notion of *infon duality*, which is important in studies of compositionality.

The *dual*, $\overline{\sigma}$, of a compound infon, σ, is defined by recursion as follows.

- If σ is a basic infon of the form $\ll R, a_1, \ldots, a_n, i \gg$ then
  $\overline{\sigma} = \ll R, a_1, \ldots, a_n, 1 - i \gg.$

- If $\sigma = \sigma_1 \wedge \sigma_2$, then $\overline{\sigma} = \overline{\sigma_1} \vee \overline{\sigma_2}$.

- If $\sigma = \sigma_1 \vee \sigma_2$, then $\overline{\sigma} = \overline{\sigma_1} \wedge \overline{\sigma_2}$.

- If $\sigma = (\forall \dot{x} \in u)\tau$, then $\overline{\sigma} = (\exists \dot{x} \in u)\overline{\tau}$.

- If $\sigma = (\exists \dot{x} \in u)\tau$, then $\overline{\sigma} = (\forall \dot{x} \in u)\overline{\tau}$.

We say a situation e is *complete* relative to the compound infon σ if at least (and hence exactly) one of the propositions

$$e \models \sigma \quad , \quad e \models \overline{\sigma}$$

is valid.

A generalization of the above argument shows that if u is an utterance of a denial

It is not the case that Φ

and if the sub-utterance of the sentence Φ has the propositional content

$$e \models \sigma$$

and if e is complete relative to σ, then the propositional content of u is

$$w \models \ll \models, e, \overline{\sigma}, 1 \gg$$

which is 'equivalent' to

$$e \models \overline{\sigma}.$$

The point made earlier is that, for an utterance of a denial to be suitably informational, the speaker should ensure that the listener is sufficiently aware of the context. In the theory's terms, what this amounts to is that the described situation as understood by both speaker and listener should be complete relative to the requisite infon.

27   CONDITIONALS

Conditionals, or *if–then* statements, are the bedrock of rational argument and as such are central not only to such overtly 'logical' pursuits as mathematics, computer science, the sciences in general, philosophy, and the legal system, but to large parts of our everyday life. And yet for all their ubiquity, conditionals resisted the attempts of generations of philosophers to understand just what the devil they *are*? What exactly does a conditional say about the world? There is a great deal that can, and has been, said. Here we shall simply pursue the matter sufficiently to indicate the role that situation theory can play.

In our current terminology, the issue to be investigated is this. If u is an utterance of a conditional of the form

If Φ *then* Ψ

then what is the propositional content of u (and hence what is the meaning of the sentence uttered)?

We consider four examples that, though having some similarities, lead to quite different, but in many ways paradigmatic analyses:

1. *If it freezes, Ovett wears a hat.*

2. *If it freezes, Ovett will not run.*

3. *If it freezes, Ovett will be cold.*

4. *If it had frozen, Ovett would not have run.*

Again as a homage to the era when situation semantics was being developed, all four examples will be understood to refer to cross-country races and the British athlete Steve Ovett, who dominated middle distance running in the 1980s.

Sentence 1 appears first because its analysis turns out to be different from the others. Indeed, although all four sentences have an *if–then* form, an utterance of any of sentences 2, 3, or 4 will refer to a specific, single event, a cross-country race in this case, whereas sentence 1 can only be used to refer to such events *in general*.

In fact, an utterance of sentence 1 does not express a *conditional* at all, but rather is a statement of the validity of a certain constraint, a general connection the obtains between all those events when it freezes and all those events when Ovett wears a hat. (Actually, there is a reading of sentence 2 that also serves to express a general link. We shall not consider this alternative reading, and the analysis presented below will exclude this possibility. As always, the main concern is with utterances of sentences, and by concentrating on utterances we avoid alternative readings of sentences.)

The remaining three sentences all do express genuine conditionals of one form or another. Sentences 2 and 3 are syntactically similar. Each may be used to predict some form of link between two specific future events. Sentence 4 is different in that a speaker would normally only utter sentence 4 after the race in question had taken place, and moreover only if, counter to the antecedent of the utterance, it had in fact not frozen. Statements made with sentences such as 4, where the antecedent is false, are known as *counterfactuals*. Non-counterfactual, predictive-type conditionals such as examples 2 and 3 are often referred to as *indicative* conditionals.

Let u_1 be an utterance of sentence 1:

If it freezes, Ovett wears a hat.

This does not refer to any particular pair of events. Rather the utterance states that there is a connection between two *types* of event, the type of race situation where it is freezing and the type of race situation where Ovett wears a hat. In other words, what u_1 does is state a certain constraint. We make this precise below.

Let
$$S = [\dot{e} \mid \dot{e} \models \quad \ll \text{present-in}, SO, \dot{e}, \dot{t}_r, 1 \gg \wedge$$
$$\ll \text{registered-in}, SO, \dot{r}, \dot{t}_r, 1 \gg \wedge$$
$$\ll \text{freezing}, \dot{t}_r, 1 \gg]$$

$$T = [\dot{e} \mid \dot{e} \models \quad \ll \text{wears-hat}, SO, \dot{t}_r, 1 \gg],$$

where $\dot{r}$ is a parameter for a race, $\dot{e}$ is a parameter for the situation surrounding $\dot{r}$ (that is to say, the race itself, the race organization, and the environment local to the race), $\dot{t}_r$ is a parameter for the time of $\dot{r}$, and $SO = c_{u_1}(\text{OVETT})$.

Then the propositional content of u_1 is

$$w \models (S \Rightarrow T),$$

or at least

$$d \models (S \Rightarrow T),$$

for a suitably large part of the world d (enough to include all the race situations involving Steve Ovett).

The remaining three examples all have in common the fact that they are used to refer to specific events. (At least, this is true in the case of their normal uses, the ones considered here.) Nevertheless they all exhibit quite distinctive features that make it difficult to come up with any kind of unified treatment that seems appropriate for all examples.

We shall present two alternative treatments, both of which have some appeal as well as some shortcomings.

One approach to handling conditionals in logic is the material conditional. The first treatment of the semantics of sentences 2 through 4 develops a version of this approach within the framework of situation semantics. (This treatment adopts an extreme form of the material conditional that expresses nothing more than the contingent prohibition of two particular eventualities. Other treatments of the conditional can be developed within the framework of situation theory that could also be described as a material conditional — for example, taking the relationship to link *types* rather than specific propositions as below.)

Let u_2 be an utterance of sentence 2:

If it freezes, Ovett will not run.

A situation-theoretic analysis of this utterance along the lines of the material conditional goes as follows.

The utterance u_2 refers to some particular circumstance, an upcoming race and how the weather will affect the participation of Ovett. The described situation, d, therefore, comprises the organization of the race and the meteorological environment local to the race.

Note that the race is not an existing situation, nor an event that has taken place in the past, but rather is some planned, future event: indeed an event that might eventually be cancelled, and not take place at all. Thus r has an objective existence purely as a result of the intentionality network of planning agents, *to whit* Man. But this does not prevent r being a perfectly well-defined situation in our ontology. People discuss future events all the time, and frequently plan their activities around future events.

What claim does the utterance make about the situation d? It does not state some kind of constraint, as does an utterance of 1. Nor is there a constraint of

which this is an instance, as is the case in example 3, which we consider presently. There is no generally prevailing causal link between the local temperature and Ovett running or not running. Runners can, and do, run in freezing conditions, Ovett among them. The freezing conditions might well be the *reason* Ovett decides not to run on this particular occasion, but that is Ovett's personal decision. There is no general rule, no constraint as there was in example 1.

Rather what the utterance does is claim that a certain event will not occur, namely the event of it freezing and Ovett running in the race. That is to say, if we let r be the race, l_r the location of r, t_r the time of r, and e the environment local to l_r, then the propositional content of u_2 is:

$$d \models \ll \text{precluded}, P \wedge Q, t_u, 1 \gg,$$

where P is the proposition

$$e \models \ll \text{freezing}, l_r, t_r, 1 \gg,$$

and Q is the proposition

$$r \models \ll \text{runs} - \text{in}, SO, r, t_r, 1 \gg,$$

and where $SO = c_{u_2}(\text{OVETT})$.

Turning now to sentence 3, let u_3 be an utterance of:

If it freezes, Ovett will be cold.

Again we develop a situation-theoretic analysis analogous to the material conditional of classical logic.

In this case the utterance u_3 expresses an instance of a general *constraint*, the constraint that if it is freezing then a person will be cold. There is a definite, generally prevailing, causal link between the antecedent *'it freezes'* and the consequent *'Ovett will be cold'*. However, it is arguable (see momentarily) that although 2 and 3 differ as to the reason for the validity of the expressed conditional, this difference does not affect the meaning of the sentence, and the propositional content in the case of example 3 will be just as in 2. Thus, if d is the described situation and t_d is the requisite time (so $t_d = c_{u_3}(\text{WILL})$ and $t_{u_3} \prec t_d$), then the propositional content of the sub-utterance of *'it freezes'* is

$$d \models \ll \text{freezing}, t_d, 1 \gg .$$

and the propositional content of the sub-utterance of *'Ovett will be cold'* is

$$d \models \ll \text{cold}, SO, t_d, 1 \gg .$$

Then the propositional content of u_3 is

$$d \models \ll \text{precluded}, P \wedge Q, t_{u_2}, 1 \gg$$

where P is the proposition

$$d \models \ll \text{freezing}, t_d, 1 \gg$$

and Q is the proposition

$$d \models \ll \text{cold}, SO, t_d, 0 \gg .$$

According to the above analysis then, the reason why the semantics of 3 works out the same as for 2 is that, although the utterance of 3 states an instance of a general constraint, it is not part of the utterance that it is such an instance. Rather the utterance asserts a simple conditional that expresses, as a matter of fact, that a particular pair of events cannot occur in conjunction. The distinction between 2 and 3 is part of the general background knowledge of the world that both the speaker and listener will be aware of. The constraint of which 3 states a particular instance is not part of the propositional content of the utterance u_3, since the utterance makes no reference to the constraint.

So far then, a material-conditional style analysis seemed to work for examples 2 and 3. What about the final example? Let u_4 be an utterance of the sentence 4:

If it had frozen, Ovett would not have run.

Presumably u_4 refers to a specific event, a past race r, run at a location l_r at a time t_r where $t_r \prec t_u$, in an environment e. The utterance refers to properties of each of the situations r and e, the property of it freezing in e and the property of Ovett running in r. This was also the case in example 2. If we attempt an analysis using the material-conditional approach as in example 2, we obtain the following propositional content for u_4:

$$d \models \ll \text{precluded}, P \wedge Q, t_u, 1 \gg$$

where P is the proposition

$$e \models \ll \text{freezing}, l_r, t_r, 1 \gg$$

and Q is the proposition

$$r \models \ll \text{runs-in}, SO, r, t_r, 1 \gg,$$

and where d is the described situation.

But what is the described situation? In the case of example 2, d comprised both r and e, that is to say, both the race and the (meteorological) environment local to the race. But this cannot be right in this case. Why? Well, the use of the subjunctive in 4 is only appropriate if in fact

$$e \models \ll \text{freezing}, l_r, t_r, 0 \gg$$

and if this is the case and we take d to extend e, then our proposed propositional content is degenerate and essentially non-informational: it would be a valid proposition regardless of whether or not Ovett ran in the race.

This is, of course, why the material conditional fails so miserably to handle counterfactuals in classical logic. The material conditional renders a proposition

$$P \rightarrow Q$$

as true whenever P is false, and consequently is unable to handle counterfactuals, which by their very nature have a false antecedent.

But a situation-theoretic framework saves us from falling into this trap, and in a way that squares with our everyday intuitions about counterfactuals. In making an utterance of 4 with the sincere intention of conveying information, the speaker is not referring to the situation as it was, but to some hypothetical variant thereof, a variant that resembles the actual situation in almost every way except for differing as to the fact of it freezing or not.

In other words, the described situation d is not a part of the world extending the actual race-environment situation e. It is some abstract situation postulated by the speaker. If d_a denotes the actual race organization and environment local to the race, what was the described situation in example 2, then d and d_a will have the same constituents and the same spatial and temporal extent, and for almost all infons σ it will be the case that

$$d \models \sigma \quad \text{if and only if} \quad d_a \models \sigma,$$

but

$$d \models \, \ll\text{freezing}, l_r, t_r, 1 \gg \quad \text{and} \quad d_a \models \, \ll\text{freezing}, l_r, t_r, 0 \gg .$$

What justification is there for allowing a situation such as d into the ontology? Well, people do indeed use conditionals such as the above all the time, and if you accept the two premises that (a) when two people are engaged in a successful exchange of information, they must be talking about *something*, and (b) we use situations to represent these 'somethings', then it follows that hypothetical entities such as the d above will figure as situations.

To summarize the above account, suppose u is an utterance of a conditional sentence of the form

$$\textit{If } \Phi \textit{ then } \Psi,$$

(or equivalent) and that

$$e_1 \models \sigma_1$$

is the propositional content of the sub-utterance of Φ and

$$e_2 \models \sigma_2$$

is the propositional content of the sub-utterance of Ψ. Then the propositional content of u is:

$$d \models \ll \text{precluded}, (e_1 \models \sigma_1) \wedge (e_2 \models \overline{\sigma_2}), t_u, 1 \gg$$

where $d = s_u(\text{IF } \Phi \text{ THEN } \Psi)$ is the described situation.

In the case of an indicative conditional, the described situation, d, will include both e_1 and e_2. In the case of a counterfactual, where in fact

$$e_1 \models \overline{\sigma_1}$$

then d will be a hypothetical situation that differs minimally from what actually occurred (i.e. from a situation including both e_1 and e_2) in that:

$$d \models \sigma_1.$$

The alternative approach to the semantics of conditionals is not only uniform across examples of forms 2, 3, and 4, as was the case with the first treatment, but in fact includes example 1 as well, in that an utterance of *any 'if–then'* statement is taken to refer to a constraint (in one way or another).

We commence with sentence 3. As before, u_3 is an utterance of the sentence

If it freezes, Ovett will be cold.

This utterance expresses an instance of the constraint that, if a person's environment is freezing, and that person is scantily clad (such as a runner), then that person will be cold. More precisely, let S and T be the situation-types

$$\begin{aligned}
S = [\dot{e} \mid \dot{e} \models \quad &\ll \text{freezing}, \dot{t}, 1 \gg \wedge \\
&\ll \text{present-in}, \dot{p}, \dot{e}, \dot{t}, 1 \gg \wedge \\
&\ll \text{scantily-clad}, \dot{p}, \dot{t}, 1 \gg] \\
T = [\dot{e} \mid \dot{e} \models \quad &\ll \text{cold}, \dot{p}, \dot{t}, 1 \gg]
\end{aligned}$$

where $\dot{e}$ is a situation parameter, $\dot{t}$ is a temporal parameter, and $\dot{p}$ is a parameter for a person.

Then the described situation for u_3 is the world and the propositional content is:

$$w \models (S \Rightarrow T)[f]$$

where f anchors $\dot{p}$ to $SO = c_{u_3}(\text{OVETT})$.

Turning now to example 2, let u_2 be an utterance of the sentence:

If it freezes, Ovett will not run.

As noted earlier, u_2 differs from u_3 in that it does not express an instance of a general constraint. And yet it does make a *prediction* of a future event. Assuming this prediction has an informational basis, and is not just a random guess, how can this be? Surely the only informational basis on which to make such a prediction is knowledge of some uniformity that systematically links the eventuality of it freezing and Ovett's deciding not to run; in other words, a constraint.

But what constraint? As observed earlier, runners can and do run in freezing conditions. Indeed, Ovett himself has run in freezing conditions, though as a matter of fact he prefers not to. Whether or not Ovett runs in the race referred to in the utterance u_2 is purely up to Ovett to *decide*. So where is the constraint?

The answer is that human beings are *planning* creatures. They form plans or intentions as to their future courses of action. And part of this plan-formation process will involve establishing what we might call *personal constraints*, constraints that govern their own action in accordance with their own desires and intentions.

Thus, Ovett, having found as a result of past experience that running in freezing conditions is unpleasant, and indeed can lead to illness and injury, might well decide that in future he will not run if it is freezing. Or it may be even more specific than this. Maybe he has just recovered from a cold and decides that, as far as next Saturday's race is concerned, the one referred to in u_2, he will not run if it is freezing. Beyond next Saturday he forms no intentions either way as far as running in cold weather is concerned. But for this one occasion he forms a personal constraint that will guide his future actions.

Knowing of this constraint, a speaker may then confidently utter sentence 2. That is to say, it is the knowledge of the constraint that provides the speaker with an informational basis for the utterance. In effect, what the utterance of sentence 2 conveys to the listener is that 'this guy Ovett has formed the intention that *if* it is freezing on the day of this particular race, *then* he will not run'. Indeed, we may adopt the position that it is precisely this constraint that provides the propositional content of u_2.

More precisely, let S and T be the situation-types

$$S = [\dot{e} \mid \dot{e} \models \quad \ll \text{environment-of}, \dot{e}, r, t_r, 1 \gg \wedge$$
$$\ll \text{freezing}, l_r, t_r, 1 \gg]$$

$$T = [\dot{e} \mid \dot{e} \models \quad \ll \text{run-in}, SO, r, l_r, t_r, 0 \gg]$$

where r is the race in question, l_r is its location, t_r is its time, and $\dot{e}$ is a situation parameter.

Taking the described situation, d, to be Ovett's state at the temporal interval t_{u_2}) then the propositional content of u_2 is:

$$d \models (S \Rightarrow T).$$

At which point a not unnatural question would be: why does the same treatment not work in the case of sentence 3? Though in the case of 3 there was a prevailing general constraint, the actual utterance only referred to an instance of that constraint involving Ovett. So why in case 3 did we take the described situation to be the world, and the propositional content to be

$$w \models (S \Rightarrow T)[f]$$

where $(S \Rightarrow T)$ is a general constraint and f an anchor to Ovett? Why not particularize the constraint to Ovett in the first place, as in example 2?

The answer is this. In case 2, the utterance has nothing to do with Ovett's state of mind, with his desires and his intentions. There is no personal constraint of this nature. For all the speaker or listener knows, Ovett has not given a thought to it being cold on race day and his getting cold then. Moreover, there is no reason to assume that the situation d will support the general constraint that if it is freezing a person will get cold, or even that if it is freezing Ovett will get cold. Nevertheless, if it does freeze on race day, Ovett certainly will get cold. Not because of any plan of intention he has formed. Simply because there is a prevailing general constraint to the effect that scantily clad people get cold if the temperature falls below freezing. The propositional content of u_3 has a structure that accords with this observation.

In example 2, on the other hand, there is no prevailing general constraint, only the personal constraint (or 'contingency plan') formulated by Ovett.

In neither case does the speaker explicitly mention the constraint. But, according to the present account, the constraint is nevertheless the *content* of the utterance: the propositional content captures what it is the speaker claims to be the case.

Finally, what about the counterfactual case, example 4? Let u_4 be an utterance of the sentence

If it had frozen, Ovett would not have run.

The grammatical structure of the sentence makes it clear that the utterance is made after the race has taken place, and that in fact it had not frozen. The speaker is describing the personal constraint Ovett had formed prior to the race. As it happens, the conditions that would have brought that constraint into play, and resulted in Ovett's not running, did not prevail — it did not freeze. But Ovett nevertheless *had formed* that constraint, and would have acted in accordance with it. This is what the utterance claims. Accordingly, the propositional content of the utterance is almost the same as in the previous case.

What distinguishes these two cases are the circumstances of utterance. In example 2, at the time of the utterance, the race has not yet taken place ($t_u \prec t_r$) and the utterance describes a constraint that prevails at the time of utterance; in example 4, the race has already taken place ($t_r \prec t_u$) and moreover it did not freeze, and the utterance describes a constraint that prevailed at the time of the

race. Thus with the types S and T as before, the propositional content of u_4 is

$$d \models (S \Rightarrow T)$$

where in this case the described situation, d, is Ovett's state *at the time of the race*.

We finish this section by examining a famous pair of examples due to Quine. The traditional question is what is the status of the following two sentences?

(1) *If Bizet and Verdi had been compatriots, Bizet would have been Italian.*

(2) *If Bizet and Verdi had been compatriots, Verdi would have been French.*

A lot of the considerable discussion generated by these examples has concentrated on their counterfactual nature. But similar problems arise if we consider the following two indicative sentences involving the contemporary American philosopher John Perry and the British linguist Robin Cooper:

(3) *If Perry and Cooper are compatriots, then Perry is English.*

(4) *If Perry and Cooper are compatriots, then Cooper is American.*

We investigate both pairs of sentences first using the material conditional framework and then in terms of the constraint-based approach. The conclusion we shall draw is that the material conditional works moderately well in the case of sentences (1) and (2), but fails hopelessly when presented with (3) and (4), whereas the treatment in terms of constaints handles both pairs with ease. Indeed, my examination of these examples provides strong evidence to suggest that the constraint-based approach is the right way to handle conditionals, be they counterfactual or indicative.

Of course, unlike many of the discussions that have taken place concerning sentences (1) and (2), our approach will be in terms of utterances of these sentences. So, starting with the material conditional treatment of the first pair of sentences, let u_1 be an utterance of sentence (1). Let $B = c_{u_1}(\text{BIZET})$, $V = c_{u_1}(\text{VERDI})$, and let t be the time to which the utterance implicitly refers, i.e. the time when both Bizet and Verdi were alive. Let d denote the described situation.

According to the framework developed above, the propositional content works out to be:

(i) $d \models \;\ll\text{precluded}, P \wedge Q, t_{u_1}, 1 \gg$

where P is the proposition

(ii) $d \models \;\ll\text{compatriots}, B, V, t, 1 \gg$

and Q is the proposition

(iii) $d \models \;\ll\text{Italian}, B, t, 0 \gg,$

and where d differs from reality, d_a, in a minimal fashion such that (ii) is valid.

Now,

(iv) $d_a \models \,\ll\text{Italian}, V, t, 1 \gg$

and

(v) $d_a \models \,\ll\text{French}, B, t, 1 \gg .$

So if d is to differ from d_a minimally it must, by (i), be the case that

(vi) $d \models \,\ll\text{Italian}, V, t, 1 \gg$

and

(vii) $d \models \,\ll\text{Italian}, B, t, 1 \gg .$

Thus in this case d is a hypothetical situation in which both Bizet and Verdi are Italian.

Starting with an utterance u_2 of sentence (2) we likewise end up with a hypothetical situation d' such that

(viii) $d' \models \,\ll\text{French}, V, t, 1 \gg$

and

(ix) $d' \models \,\ll\text{French}, B, t, 1 \gg .$

These are the only possible outcomes if the described situation is to differ *minimally* from reality.

Is this a reasonable account? Although it does provide a consistent semantics of utterances of sentences (1) and (2), you may not find it particularly convincing. But still, it is a solution.

On the other hand, as far as the second pair of examples is concerned, utterances of sentences (3) and (4), the material conditional approach simply does not get off the ground. An utterance of either (3) or (4) certainly does not postulate a hypothetical, alternative world the way that the subjunctive in (1) and (2) does. Rather the described situation must be (part of) the real world. But then the falsity of the antecedent renders the entire semantics degenerate.

Ultimately, it is this example, and others like it, that persuade many to opt for the second of my two treatments, the one in which conditionals are taken to refer to constraints, even though the material conditional does provide a good semantics for the future-directed, predictive type of indicative conditional and an acceptable, if not wholly convincing, semantics for counterfactuals.

The constraint-based semantics for conditionals provides a uniform treatment for all four sentences, as well as clarifying the issues involved in these examples.

An utterance of any one of the four sentences refers to a generally prevailing constraint of the form:

> If person A and person B are compatriots and person A has nationality
> N, then person B has nationality N.

for some nationality N.

Let u_1 be an utterance of sentence (1), and let B, V, t denote Bizet, Verdi, and the time they were both alive, as before. Let $\dot{a}$, $\dot{b}$ be parameters for people, and let S_1, T_1 be the situation-types

$$S_1 = [\dot{e} \mid \dot{e} \models \;\ll\text{compatriots}, \dot{a}, \dot{b}, \dot{t}, 1 \gg \wedge \ll\text{Italian}, \dot{a}, \dot{t}, 1 \gg]$$

$$T_1 = [\dot{e} \mid \dot{e} \models \;\ll\text{Italian}, \dot{b}, \dot{t}, 1 \gg].$$

Then the described situation for u_1 is the world and the propositional content of u_1 is:
$$w \models (S_1 \Rightarrow T_1)[f]$$
where $f(\dot{a}) = V, f(\dot{b}) = B$.

Similarly, the propositional content of an utterance, u_2, of sentence (2) is:

$$w \models (S_2 \Rightarrow T_2)[f]$$

where

$$S_2 = [\dot{e} \mid \dot{e} \models \;\ll\text{compatriots}, \dot{a}, \dot{b}, \dot{t}, 1 \gg \wedge \ll\text{French}, \dot{b}, \dot{t}, 1 \gg]$$

$$T_2 = [\dot{e} \mid \dot{e} \models \;\ll\text{French}, \dot{a}, \dot{t}, 1 \gg]$$

and where f is as before.

Notice that this semantics for utterances u_1 and u_2 resolves the confusion that can arise between (1) and (2). Given the constraint that figures in its propositional content, an utterance of sentence (1) will be appropriate — that is to say it will be informational — if the listener and speaker know that Verdi was Italian. Then the utterance makes a valid assertion that describes this particular instance of that constraint. Likewise, an utterance of u_2 will be appropriate given the knowledge that Bizet was French.

Of course, an anchor, f, that assigns Verdi to the parameter $\dot{a}$ and Bizet to the parameter $\dot{b}$ is not possible for any situation that includes both of these individuals and is of type S_1, so there can be no actual situation to which the constraint

$$(S_1 \Rightarrow T_1)[f]$$

applies.

If you accept the existence of hypothetical situations, then this is not an obstacle. Since the constraint is reflexive, it simply guarantees that in any hypothetical situation e in which Bizet and Verdi are compatriots and Verdi is Italian, then Bizet is Italian.

However, even if you reject hypothetical situations, the propositional content is still informational, in that it describes a valid constraint: the constraint itself is not invalid, it is just that it does not apply to the pair Bizet, Verdi.

Similar remarks apply in the case of the second sentence.

Turning now to the pair (3), (4), all we need to do now is observe that the above analysis works equally well in this case. Indeed, the temporal location plays no external role in the above discussion, and hence there is no distinction between the first pair and the second as far as our analysis is concerned. It applies equally to sentences that refer to past events and sentences that apply to the present, and indeed to sentences that refer to the future.

28   THE LIAR PARADOX

To finish, we see how situation semantics resolves the famous semantic paradox, the Liar. This is most often given as a query for the truth or falsity of the sentence:

(1)   *This sentence is false.*

In this form, the problem does not arise for us, since sentences are not the kinds of object that are either true or false. Instead, they are objects that can be *used* to convey information. Since a sentence is not an appropriate argument for the property 'true/false', from a situation semantic perspective (1) simply has no meaning.

Suppose we modify the question to the truth or falsity of an *utterance* of the sentence:

(2)   *This utterance is false.*

Again there is no paradox: this sentence too has no meaning. As with sentences, utterances are not the kinds of things that are true or false. However, we are getting closer, since an utterance of an assertive sentence will determine a *proposition*, and in situation semantics it is the propositions that are the bearers of truth.

The correct formulation of the Liar paradox in situation semantics is in terms of the proposition determined by an utterance, u, of the sentence:

(3)   *The proposition expressed by this utterance is false.*

or, more simply but less precise, an utterance of the sentence

(Φ)   *This proposition is false.*

Let s be the described situation, $s = s_u(\Phi)$. Then, taking the basic property here to be *'true'*, with *'false'* being identified with *'not true'*, the propositional content, p, of u is:

$$s \models \ll \text{true}, p, 0 \gg.$$

Notice that p is itself the proposition referred to by the phrase THIS PROPOSITION in the utterance, that is:

$$p = c_u(\text{THIS PROPOSITION}).$$

The utterance claims that p is false. In other words, p claims that p is false. This is starting to look like a paradox. But we need to make sure we know exactly what is meant by 'false' here. Since we are taking *'false'* to mean *'not true'*, this amounts to clarifying what we mean by 'truth'.

In situation semantics, every proposition

$$e \models \sigma$$

is either *true* or *false* (i.e. *not true*). Truth means that e does indeed support σ, which is a strong condition to place on the situation s. Falsity, on the other hand, means simply that s fails to support σ, which is fairly weak — unlike the proposition $e \models \overline{\sigma}$, which is strong, but in general quite different. Let's examine the proposition p above with this notion firmly in mind.

Suppose first that p is true. Thus

$$s \models \ll \text{true}, p, 0 \gg$$

is a valid proposition. Then, since s is part of the entire world, w, it follows that

$$w \models \ll \text{true}, p, 0 \gg$$

is a valid proposition. In other words, p is false. This is a contradiction.

Hence p must be false. In other words,

$$s \not\models \ll \text{true}, p, 0 \gg .$$

But this is not necessarily a paradox. All it says is that the situation s does not support the infon

$$\ll \text{true}, p, 0 \gg .$$

Now since p is false, we certainly have

$$w \models \ll \text{true}, p, 0 \gg$$

but again this is not necessarily paradoxical unless $s = w$. So what our investigation amounts to is not a paradox but a straightforward proof of a theorem:

Theorem 1: $\quad s_u(\Phi) \neq w$

In words, the described situation in an utterance of the Liar sentence Φ cannot be the world.

Moreover, since we have shown that p is false, we have also established another theorem:

Theorem 2: Any utterance of the Liar sentence Φ expresses a false proposition.

In short, the Liar paradox has been resolved. Or has it? Can't we simply re-introduce the paradox by modifying Φ to read:

(Φ')   *This proposition is false in the world.*

Surely in this case the described situation, $s_u(\Phi')$, will have to be $w = c_u(\text{THE WORLD})$, won't it?

In fact it will not. Analogs of Theorems 1 and 2 go through for the modified sentence Φ', so the same argument as before shows that w cannot be the described situation.

The conclusion has to be that w is *not* a situation. And so we have a third theorem:

Theorem 3: w is not a situation.

This is analogous to the result in set theory that the class of all sets is not a set. But notice that this does not prevent set-theorists from discussing the so-called *universe* of sets, V, all the time, and often treating it as if it were a set. The trick is simply to develop enough sophistication to do this with safety.

Similarly, in situation theory we often handle *the world* much as if it were a situation. We just have to bear in mind that it is not in fact a situation, and make sure we do not use it inappropriately.

29 FURTHER READING

For a historical introduction to the early development of situation semantics, see Barwise and Perry's initial [1980] paper and their [1983] book.

For a comprehensive coverage of situation theory and situation semantics as it eventually settled down, see [Devlin, 1991] book. Much of this article is based on that treatment.

For a complete survey of all of Barwise's papers on situation theory and situation semantics, see the [Devlin, 2004] article.

For an extended discussion of the application of situation semantics to the resolution of the classical Liar Paradox, see Barwise and Etchemendy's [1987] book.

An excellent compilation of many of Barwise's papers on situation theory and situation semantics is provided by his [1989] monograph.

The [1997] book by Barwise and Seligman provides an exploration of the situation theoretic notion of a constraint in a very general setting.

For an application of situation theory to decision making using an action- oriented approach, see Devlin and Rosenberg's [1996] monograph. A less technical coverage of roughly the same material, aimed at the business world, is provided by Devlin's [1999] book.

See also the series of conference volumes *Situation Theory and Its Applications*, Volumes I, II, and III, all published in the CSLI Lecture Notes series (1991–1993).

BIBLIOGRAPHY

[Barwise, 1989] J. Barwise. *The Situation in Logic*. Volume 17, *CSLI Lecture Notes*. CSLI Publications, 1989.

[Barwise and Etchemendy, 1987] J. Barwise and J. Etchemendy. *The Liar: An Essay on Truth and Circularity.* Oxford University Press, 1987.

[Barwise and Perry, 1980] J. Barwise and J. Perry. The Situation Underground. In J. Barwise and I. Sag (eds.), *Stanford Working Papers in Semantics*, volume 1, section D, pages 1–55. Stanford Cognitive Science Group, 1980.

[Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes.* Bradford Books– MIT Press, 1983.

[Barwise and Seligman, 1997] J. Barwise and J. Seligman. *Information Flow: The Logic of Distributed Systems.* Cambridge University Press, 1997.

[Devlin, 1991] K. Devlin. *Logic and Information.* Cambridge University Press, 1991.

[Devlin, 1999] K. Devlin. *Infosense: Turning Information into Knowledge.* W.H. Freeman, 1999.

[Devlin, 2004] K. Devlin. Jon Barwise's Papers on Natural Language Semantics. *Bulletin of Symbolic Logic*, 10:54–85, 2004.

[Devlin and Rosenberg, 1996] K. Devlin and D. Rosenberg. *Language at Work: Analyzing Communication Breakdown in the Workplace to Inform Systems Design.* CSLI Publications– Cambridge University Press, Standford University, 1996.

DIALOGUE LOGIC

Erik C.W. Krabbe

1 INTRODUCTION

In 1961 a PhD-ceremony took place in Kiel (Germany), where the new doctor defended a dissertation with a title that must have sounded a bit peculiar at the time: *Arithmetik und Logik als Spiele* [Arithmetic and Logic as Games]. The supervisor of this dissertation was Paul Lorenzen (1915–1994), who was known as a contributor to research into the foundations of mathematics. The new doctor was Kuno Lorenz. The innocent spectator was bound to wonder what serious sciences such as arithmetic and logic could have to do with games. And what kind of game could that be? Glancing through the dissertation one would discover that these were *dialogue games*, systems for discussion in which assertions are attacked and defended according to meticulously formulated *rules of dialogue*. The dissertation expands on ideas and suggestions, then recently proposed by Paul Lorenzen, that were primarily motivated by certain problems in the foundation of mathematics.

The theory of dialogue games based on logical rules can be seen as a technical elaboration of Wittgenstein's notion of 'language games'. However, given its emphasis on formal rules, it admittedly displays no more than a weak affinity to Wittgenstein's later philosophy [1953]. After 1961, this theory — dialogical logic or dialogue logic, for short — was further developed, both by Paul Lorenzen himself [1962; 1969; 1987; Kamlah and Lorenzen, 1967; 1973; Lorenzen and Schwemmer, 1973; 1975] and by Kuno Lorenz [1968; 1973] as well as by other authors.[1]

Related ideas were proposed, independently, by Jaakko Hintikka [1968], who developed a game-theoretic semantics, and by Charles Hamblin (1922–1985) [1970], who reshaped the theory of fallacies.

Nowadays, the influence of dialogue logic is clearly traceable in the field of theory of argumentation (including informal logic).[2] The more formal studies in artificial intelligence that pertain to defeasible reasoning constitute another area where dialogue logic had some influence.[3]

---

[1] [Stegmüller, 1964; Drieschner, 1966, Kindt, 1970; 1972; 1980; Van Dun, 1972, Thiel, 1978; Barth and Krabbe, 1982; Haas, 1980; 1984; Mayer, 1981; Stegmüller and Varga von Kibéd, 1984; Felscher, 1985; 1986; Krabbe, 1985; 1985a; 1986; 1988; Hoepelman and Van Hoof, 1993; Van Hoof, 1995; Valerius, 1990; Rahman and Rückert, 1999; 2001b].

[2] See, for instance, Van Eemeren and Grootendorst [1982; 1984; 1992; 2004], Walton [1984; 1998], and Walton and Krabbe [1995].

[3] See, for instance, [Vreeswijk, 1993; Rahman and Rückert, 2001b].

Along these two streams, argumentation theory and artificial intelligence, the dialogical or dialectical point of view has become almost a commonplace and nowadays very few researchers will be plagued by associations with frivolity when they hear that some serious intellectual activities are analyzed as games. But around 1961 the climate was still thoroughly 'monolectical'. Logic was supposed to be about formal proofs, with 'deduction' and 'derivation' as key terms. The prevailing picture consisted of one mind diligently deriving propositions from a list of axioms or assumptions. Another picture, based on semantic analysis and the search for counterexamples was still relatively new, but gaining in importance. The picture of logic as an interpersonal affair, however, seemed almost forgotten. By these lights, the step taken by Lorenzen was no less than revolutionary, as it set out to replace the prevailing monolectism of a solitary reasoner by an interactive dialectism of discussants, thus returning to logic's dialectical roots.

In this essay, I am concerned with the history of dialogue logic in the second half of the 20$^{\text{th}}$ century. By dialogue logic I mean the study of systems of rules (called 'dialogue games', 'dialectic systems', or 'systems of formal dialectic') that regulate discussions between two (or more) parties, where some of the rules are *logical*, which is to say that they prescribe attacks and defenses on the basis of the logical form of the sentences attacked or defended. I shall mainly, but not exclusively, focus on the systems proposed by Paul Lorenzen and by those logicians that were inspired by him.

The discussion of 20$^{\text{th}}$ century developments will be preceded by a brief glance at the dialectic character of logic at the time of its birth (section 2). Next we shall study Lorenzen's earliest papers (section 3). We must then take up the discussion about structural rules, i.e. the rules that are independent of the logical forms of sentences. In this respect Lorenz's systems will be seen to differ considerably from those of Lorenzen (section 4).[4] One section will be devoted to comparisons between Lorenzen systems and those introduced by Hintikka and by Hamblin (section 5). At the end we shall discuss further developments, such as the introduction of modalities, and various applications (section 6).[5]

2   THE AGONISTIC ROOTS OF LOGIC

The word 'agonistic' in the title of this section derives from the Greek *agōn* (= game, contest). In what follows I hope to show that originally logic was concerned with a kind of discussion game or verbal contest and that, therefore, contemporary dialogue logic has its roots in the oldest kind of logic.

Mainly because of his syllogistic, Aristotle (384–322 BC) is generally honored as the first logician. However, the first book of his *Prior Analytics*, in which this syllogistic is set out, is only a small part of his logical writings. These are known as the *Organon* (= tool) and comprise: *Categories, On Interpretation, Prior*

---

[4]All the same, both groups of systems will here be covered by the term 'Lorenzen systems'.

[5]Please note as well that the bibliography contains, besides the works referred to in the chapter, a number of other contributions to dialogue logic and its applications.

Analytics (two books), *Posterior Analytics* (two books), *Topics* (eight books), and
On Sophistical Refutations. These works do not present a closely knit whole;
they were written in different periods of Aristotle's career, and underwent various
revisions. Moreover, the traditional order of these work is clearly different from the
order in which they were written and experts differ about the order in which the
works of the *Organon* got their present shape. It is, however, likely that the famous
syllogistic of *Prior Analytics I* was a late discovery and that the last two works of
the *Organon*, the *Topics* and *On Sophistical Refutations*, showing no traces of this
syllogistic, together form the clearest exposition of Aristotle's first logical theory.
The *Categories*, too, may be an early work, and can be said to present a different
aspect of this first logical theory. At the end of *On Sophistical Refutations* — which
is clearly meant to be a conclusion to the *Topics* as well — Aristotle himself, with
uncharacteristic boastfulness, declared that nothing preceded this first theory of
logic.

What was the subject of the *Topics* and of *On Sophistical Refutations*? They
dealt with the art of disputation. So, originally logic was a theory of disputation
or discussion. In the opening lines of the *Topics*, Aristotle himself characterizes
the purpose of his treatise as follows: 'The purpose of the present treatise is to
discover a method by which we shall be able to reason from generally accepted
opinions about any problem set before us and shall ourselves, when sustaining
an argument, avoid saying anything self-contradictory' [Aristotle, 1976, p. 273;
100a18–21]. The largest part of the Topics consists of a discussion of a great
number of *topoi*. The word '*topoi*' is the plural of *topos*, which literally means
'place'. The exact technical meaning of the term *topos* is still a moot question.
According to Slomkowski 'a *topos* is a universal proposition and functions as a
hypothetical premiss [a premise that is a generalized (bi)conditional statement]
in hypothetical syllogisms [deductions based on such premises] which in turn are
constructed with the help of *topoi*' [1997, p. 3]. Examples are principles such as:
'*A* is *B* if and only if *not-B* is *not-A*' and '*A* is *B* if and only if more or less or
equally much *A* is more or less or equally much *B*'. Knowledge of such universal
principles is needed to be successful in certain kinds of structured discussion in
which one party tries to refute the other by means of a deduction (see below).
The work *On Sophistical Refutations* is sometimes listed as the ninth book of the
Topics. It treats of fallacies as stratagems that may be used to trip up one's
opponent as well as of the opportunities for logical self-defense.[6]

But what kind of discussion do these works refer to? This can only partially be
surmised from the *Topics*, even though its eighth and last book gives quite some
information about the course of events in these discussions. Anyhow, the following
features seem characteristic:[7]

(i)   Discussions start with a problem that can be expressed by a question of the

---

[6]*Logical Self-Defense* is the title of a well-known textbook in informal logic [Johnson and
Blair, 1994].

[7]Cf. [Moraux, 1968].

form: Is it the case that ... , or is it not? For instance: Is the world finite, or is it not?

(ii)   There are two roles: the Questioner (Q) and the Answerer (A).

(iii) A chooses as his initial thesis one of the two possible answers to the question that expresses the problem; Q's initial thesis is the other possible answer. For instance, A holds that the world is infinite and Q holds that it is finite.

(iv)   Q's aim is to successfully establish his thesis and thus to refute the Answerer. A's aim is to uphold his thesis and thus to prevent the Questioner from achieving his aim.

(v)   To achieve his aim, Q puts questions to A. These questions take the form: Is it the case that ... ? For instance: Is it the case that everything that has come into being has come into being from something? On the other hand, A does not put questions to Q. So there is an asymmetry in the roles.

(vi)   Possible replies to a question are: (1) to admit the point (A concedes), (2) not to admit the point, (3) to admit the point with certain qualifications (on the one hand... on the other hand ...), (4) to request that the question be clarified, (5) to object to the question (see Rule (ix)).

(vii) In principle, each point conceded by A offers an advantage to Q; nevertheless it is not the case that A can follow a simple strategy of refusing to concede anything. When A is asked to concede something that agrees with what is generally or sufficiently acceptable, he can not refuse the concession, unless granting it would amount to an immediate admission that Q's thesis was right.

(viii) A has a right to clarify or adjust an earlier concession whenever Q twists this concession in an undesirable direction.

(ix)   A can protest against certain of Q's questions or conclusions by formulating objections.

(x)   Q wins whenever he refutes A by deducing Q's thesis from the answers given, and also when A, to avoid this, produces answers that go against generally or sufficiently accepted opinions.

(xi)   A wins whenever Q is unsuccessful. Perhaps there is a time set for discussion so that A wins whenever Q does not win within the preset time. Also, A certainly wins whenever he manages to raise an objection that unsettles his adversary.

One might agree that this set of rules looks quite a bit like a game, with discussion rules functioning as the rules of the game. Unfortunately, no manual survives to provide us with definite answers about the ins and outs of these rules, nor do

we have any direct recordings of the discussions themselves as they occurred in Aristotle's time. But, though Aristotle created the first *theory* about this type of discussion, that does not imply that he designed the type himself. On the contrary, it is plausible that, generally, a particular cultural phenomenon precedes theorizing about that phenomenon. Indeed, in Aristotle's time the cultural phenomenon of a particular type of regimented discussion had existed for at least a century. One may think of Zeno of Elea, the sophists, Socrates, and the Megarian school. Plato's early ('Socratic') dialogues may serve as examples that have at least a lot in common with those discussions to which Aristotle's theorizing pertains. A short fragment will suffice to show that these dialogues sometimes display a gamesome character.

Dionysodorus:	You will admit all this [*among other things that Ctesippus' father is a dog*], if you answer my questions. Tell me, have you got a dog?
Ctesippus:	Yes, and a brute of one too.
D:	And has he got puppies?
C:	Yes indeed, and they are just like him.
D:	And so the dog is their father?
C:	Yes, I saw him mounting the bitch myself.
D:	Well then: isn't the dog yours?
C:	Certainly.
D:	Then since he is a father and is yours, the dog turns out to be your father, and you are the brother of the puppies, aren't you? [*Quickly to keep the other from cutting in*:] Just answer me one more small question: Do you beat this dog of yours?
C (*laughing*):	Heavens yes, since I can't beat you!
D:	Then do you beat your own father?

(Adapted from Plato: *Euthydemus* [1997, 298D–E].)

The crucial move in this fragment: *He is a father and is yours, so he is your father*, is alluded to by Aristotle in his *On Sophistical Refutations*: 'Is the dog your father?' [Aristotle, 1965, p. 121; 179a34–35]. It is there presented as an example of the so-called fallacy of accident.

The *asymmetry* in the roles — see Rule (v) above — is clearly illustrated by the next fragment that occurs a bit earlier in the same dialogue:

[*Socrates just asked a question. Dionysodorus, who anticipates that he is going to be refuted, tries to avoid answering that question. After a brief interlude, there follows a straightforward attempt to have an illegitimate change of roles.*]

Dionysodorus:	[...] so go ahead and answer.
Socrates:	Before you answer me, Dionysodorus?
D:	You refuse to answer then?
S:	Well, is it fair?
D:	Perfectly fair.
S:	On what principle? Or isn't it clearly on this one, that you have come here on the present occasion as a man who is completely skilled in arguments, and you know when an answer should be given and when it should not? So now you decline to give any answer whatsoever because you realize you ought not to?
D:	You are babbling instead of being concerned about answering. But, my good fellow, follow my instructions and answer, since you admit that I am wise.
S:	I must obey then, and it seems I am forced to do so, since you are in command, so ask away.

[Plato, 1997, 287C]

In this fragment the participants interrupt their discussion in order to engage in a *metadiscussion*, i.e. a discussion about the rules of discussion for the original discussion. What they do is analogous to interrupting a game in order to resolve some conflict about the rules of the game. In his ironic way, Socrates depicts his adversary as an expert on matters of discussion and agrees to an — actually most unfair — exchange of roles.

We saw that each participant in a discussion has a particular aim (Rules (iv), (x), and (xi)), just as in most games. These *internal objectives* — one's objectives *within* the discussion — must be distinguished from the *external objective*, the goal of the discussion as a whole. This external goal may consist in the participants' common desire to exercise themselves in finding and expressing arguments, or in their fondness of wrangling, but discussions can also constitute a method of teaching, or a method of inquiry, for instance by generating a survey of the pros and cons in a particular case. Clearly, the external goal of a discussion will (and ought to) influence the form of discussion, i.e. the way the discussion will proceed as determined by the rules.

There is a twofold task for logic here. First, logic should formulate and justify various forms of discussion that are geared to different external goals. Second, it should give strategic advice to those that are going to participate in a discussion of a particular type. As we saw, Aristotle made a start, which was also the start of logic itself.

3   SMALL BEGINNINGS

In this section, we shall analyze two brief papers by Paul Lorenzen, which were to start the development of 20$^{\text{th}}$ century dialogue logic.

3.1   *Logic and Contest*

The revival of the dialectical point of view in logic started in 1958, when Paul Lorenzen read his paper *Logik und Agon*. This paper, published as [1960], remained rather unknown until it was republished in [Lorenzen and Lorenz, 1978, pp. 1–8]. Until then, the best known early paper on dialogue logic had been [Lorenzen, 1961]. In *Logik und Agon*, Lorenzen sharply contrasted the agonistic roots of logic, which I tried to sketch in the preceding section, with the solo-minded and monolectical points of view of our time:

> If one compares this agonistic origin of logic with modern conceptions, according to which logic is the system of rules that, whenever they are applied to some arbitrary true sentences, will lead one to further truths, then it will be but too obvious that the Greek agon has come to be a dull game of solitaire. In the original two-person game only God, secularized: 'Nature', who is in possession of all true sentences, would still qualify as an opponent. Facing Him there is the human individual — or perhaps the individual as a representative of humanity — devoted to the game of patience: starting from sentences that were, so he believes, obtained from God before, or snatched away from Him, and following rules of logic, he is to gain more and more sentences. ([Lorenzen and Lorenz, 1978, p. 1], translated from the German by the present author.)

But bringing back a dialectical point of view is not the same as proposing to do logic in a dialogical way (starting dialogue logic). For the first, it is sufficient to distinguish certain dialectical roles, such as Questioner and Answerer, as well as some types and goals of dialogue. For the second, a dialectical approach is of course indispensable, but moreover one should propose rules of attack and defense that are dependent upon the *logical form* of the sentence attacked or defended. This step was indeed taken in *Logik und Agon*, a paper of no more than eight pages.

   That we need another interpretation of logic than the 'modern conception' sketched in the passage just quoted is evident from the existence of disagreements about what principles are logically valid, such as the disagreement between intuitionists and classical logicians about the Law of Excluded Middle (*tertium non datur*) [Lorenzen and Lorenz, 1978, p. 2]. In order to present such an interpretation, Lorenzen starts with formal calculi consisting of a list of uninterpreted symbols and a number of rules that permit one to derive sequences of these symbols. Such calculi had been studied by Lorenzen in his *Einführung in die operative*

Logik und Mathematik [1955], and the dialogical approach to logic must be seen as an attempt to replace these so-called operative foundations of mathematics by better ones. As an example Lorenzen presents a system with two symbols: +, and 0, and three rules:

(I) $\Rightarrow +$

(II) $x \Rightarrow x0$

(III) $x \Rightarrow +x+$

Here '$\Rightarrow$' serves to express that these are to be read as rules, and 'x' serves as a variable for arbitrary sequences of symbols [Lorenzen and Lorenz, 1978, p. 2]. In itself this calculus has nothing to do with logic. But it can serve as a basis for a logical dialogue game. In this game there are two players: a Proponent (P) and an Opponent (O). A simple case would be that P claims to be able to derive the sequence $++0+$, whereas O challenges him to do so. If P manages to present the derivation, he wins, otherwise he loses. According to Lorenzen, this simple game is all we need in order to understand what it means to claim to be able to derive $++0+$ (written as $\vdash ++0+$), no further semantics is necessary [Lorenzen and Lorenz, 1978, p. 3]. Actually, for what follows, it is inessential that claims (*Behauptungen*) by P refer to derivability in a calculus, as long as it is clear what actions are to be undertaken by P in order to win. These actions may even consist of empirical experiments [Lorenzen and Lorenz, 1978, pp. 3–4].

A somewhat more complex claim would be $\vdash x \rightarrow ++x$. The content of this claim is that, if P is challenged on this account by O, then, for any sequence r, as soon as O claims $\vdash r$, P is committed to claim $\vdash ++r$ [Lorenzen and Lorenz, 1978, p. 4]. By some rules that are not yet spelled out in this paper O can be forced to present her derivation first, so that P can profit from it when it is his turn to deliver. It is clear that this last example brings us close to what, in the introduction, was called a 'logical rule', i.e. a rule that prescribes attacks and defenses on the basis of the logical form of the sentences attacked or defended. The logical operator to which the rule pertains would be '$\rightarrow$'. Only, the antecedent and consequent are not statements or statement-like entities but sequences of symbols and variables. However, if we reinterpret the claim $\vdash x \rightarrow ++x$ as $\forall x (\vdash x \rightarrow \vdash ++x)$, the explanation in the paper can be seen to explain a logical rule for '$\rightarrow$'. If P makes this claim and is challenged by O, O should first select a sequence r to be substituted for 'x', P should then claim $\vdash r \rightarrow \vdash ++r$ and P's further commitments are exactly described by the rule. Thus one may say that this is the place where dialogue logic starts.

The paper gives further examples in which also rules for '$\forall$' and '$\vee$' are used. Negation is introduced as follows. Let the discussants have agreed on the selection of a sequence that is certainly underivable in the calculus. Let us say this is the one-place sequence '0'. Then the claim that a certain sequence, say '0+' is underivable may be expressed as $\vdash 0+ \rightarrow 0$, or, in our reconstruction, as $\vdash 0+ \rightarrow \vdash 0$. Thus, if P makes this claim, the burden of proof is upon O to show that '0+' is derivable,

i.e. O should derive '0+' or otherwise lose the game [Lorenzen and Lorenz, 1978, pp. 4–5].

With some claims, there is a way for P that leads to a win irrespective of the calculus (or other action complex) that functions as a basis for the dialogue game. These are called 'generally admissible' (*allgemein-zulässig*). Lorenzen gives the example $\vdash (x \rightarrow y) \rightarrow ((y \rightarrow z) \rightarrow (x \rightarrow z))$, which in our reconstruction would read as $\forall x \forall y \forall z ((\vdash x \rightarrow \vdash y) \rightarrow ((\vdash y \rightarrow \vdash z) \rightarrow (\vdash x \rightarrow \vdash z)))$. In Lorenzen's explanation of the general admissibility of this claim, which amounts to the Law of Transitivity of Conditionals, many things about the rules of dialogue are taken for granted that were later on to be made explicit, as the field developed. For instance, it is supposed without any comment, that O is committed to (the analogue of) *modus ponens* [Lorenzen and Lorenz, 1978, p. 5].

Some (analogues of) classical laws are not generally admissible. Lorenzen 's example here is the Law of Double Negation: $\vdash ((x \rightarrow 0) \rightarrow 0) \rightarrow x$, or in our reconstruction: $\forall x (((\vdash x \rightarrow \vdash 0) \rightarrow \vdash 0) \rightarrow \vdash x)$, where again it is presumed that '0' is the selected underivable sequence. Unlike the preceding claim, this one is, according to Lorenzen, risky for P because there could be a sequence r such that O knows how to derive r, but P does not [Lorenzen and Lorenz, 1978, p. 6]. Nevertheless, I would say, the Law of Double Negation does admit of a way for P to win, only P may fail to be aware of this fact. The strategy for this claim happens to be dependent on knowledge about what sequences are derivable in the calculus. The strategy for the Law of Transitivity of Conditionals is free of such dependency. Thus, Lorenzen has made a point, which would later be expressed by saying that, given certain rules of the game, there is a formal winning strategy in the second case, but not in the first.

Since the Law of Double Negation is one of the classical principles that were rejected by the intuitionists, and since similar stories as the one above can be told for other such principles, it seems that the interpretation of logical operators as determined by moves in a contest will be inclined to favor Brouwer's intuitionism [1908]. That is precisely what Lorenzen is saying on the last page of his paper, but he also points out that by modification of the game rules (for instance, by granting P the right to withdraw certain moves and to live up to earlier commitments, whereas according to the 'intuitionistic' rules these earlier commitments would have lapsed) one may justify classical logic. This insight was taken up in later developments. On the other hand, Lorenzen's suggestion to identify intuitionistic logic with eristic logic and classical logic with dialectic logic, where the former would be antagonistic and the latter cooperative, and thus to distinguish different types of dialogue did not take on [Lorenzen and Lorenz, 1978, p. 8].[8]

Finally, there are three points I wish to stress before we leave this remarkable paper. One is that the paper makes it crystal clear that originally dialogue logic was not about a game with meaningless formulas, but thoroughly based on a set of meaningful expressions. That is, in the parlance that developed later, dialogue

---

[8]For a further development of the idea of having different types of dialogues, see [Walton and Krabbe, 1995].

logic starts with material, not formal, dialogues. Further it is interesting that the semantics of the elementary statements is 'empractical', that is, describable in terms of actions (such as presenting a derivation in a calculus) without use of abstract semantic jargon [Lorenzen and Lorenz, 1978, p. 3]. Here we see that the paper anticipates the reform program of the Erlangen School, to which we shall briefly return in section 5. Finally the idea that dialogue logic can yield more than one logic (pluralism) was with Lorenzen's project right from the start.

3.2 A Dialogical Criterion for Constructivity

Paul Lorenzen's second paper on dialogue logic, *Ein dialogisches Konstruktivitäts-kriterium* [A Dialogical Criterion for constructivity], received much more attention than his first. It was presented on a symposium on the foundations of mathematics in 1959, published as [1961], and reprinted in [Lorenzen and Lorenz, 1978, pp. 9–16]. The paper purports to clarify the notion of 'constructivity'. To equate 'constructive' with 'recursive', Lorenzen says, yields too narrow a concept, Brouwer's explanations are understood only by few, and his own formulations about operative 'definiteness' are, so he admits, equally unclear [Lorenzen and Lorenz, 1978, p. 9].

A common problem with discussions of constructivity is that it is not always clear what entities are supposed to be either constructive or nonconstructive. Here these entities seem to be statements (*Aussagen*) and statement forms or predicates (*Aussageformen*), rather than methods or proofs; but also logical operators and inductive definitions seem to be entities that may be called constructive, probably in the sense that the new statements or statement forms they yield when applied to constructive statements or statement forms are again constructive.

Statements are called proof-definite (*beweisdefinit*) when it is decidable whether a given alleged proof is indeed a proof (a well-defined class of potential proof must here be understood). Claims to the effect that a certain sequence is derivable in a calculus belong to the proof-definite statements. Proof-definite statements form a basis from which further statements can be built, using logical operators. If these further statements can be discussed in dialogues such that each dialogue ends with a decision whether P has won or lost with respect to his claim, they are called dialogue-definite (*dialogisch-definit*). Lorenzen's proposal is to replace the vague concept of 'constructivity' with the more precise concept of 'dialogical definiteness' [Lorenzen and Lorenz, 1978, p. 10]. Lorenzen now discusses, in order, the dialogical sense of logical operators and that of inductive definitions.

The logical operators Lorenzen treats in this paper are ∧ (conjunction), ∨ (disjunction), → (implication), ¬ (negation), ∀ (universal quantifier), and ∃ (existential quantifier). To show the constructivity, in the sense of dialogue-definiteness, of these operators, it suffices to show that each statement formed by these operators starting from dialogue-definite statements (or, in the case of quantifiers, from predicates whose statemental substitution instances are dialogue-definite) will again be dialogue-definite. For this it suffices to formulate dialogue rules that reduce a di-

alogue about a statement formed by one of these operators to dialogues about its direct constituents (or, in the case of quantifiers, that constituent's statemental substitution instances). If this procedure proves successful, the curious effect will be that all first-order statements will have been shown to be constructive in the sense of being dialogue-definite. This holds for intuitionistically acceptable principles and unacceptable principles (such as the Law of Excluded Middle and the Law of Double Negation) alike. But the difference is that for the intuitionistically acceptable principles there is, supposedly, a winning strategy (*Gewinnstrategie*) for P, whereas for the unacceptable principles there is not [Lorenzen and Lorenz, 1978, pp. 11–13]. Hence 'constructive' in the sense of dialogue-definite must be sharply distinguished from 'constructively valid'.

In this paper, the dialogue rules are not yet spelled out in a satisfactory way. Instead of discussing each of them here we do better to discuss a later and more polished version. Let it suffice to remark that the paper does not contain a description of a dialogue game in game-theoretical terms, or easily translatable into such terms. Some rules underwent considerable changes in later developments. Take for instance, the rule for implication, which here says that if P claims $a \to b$ he is committed to claim b as soon as O has made the claim a and has successfully defended this claim against P [Lorenzen and Lorenz, 1978, p. 11]. But in systems as they were developed later, P need not claim b as long as there are other things for him to do, and he may reach a win before ever reaching a stage where claiming b becomes mandatory.

Since we are left a bit in the dark about the precise rules of the game, it will be a tricky affair to assess Lorenzen's claims that there is or is not a winning strategy in certain cases. The problem is the more urgent since, as we saw, in *Logik und Agon* [1960]. Lorenzen himself pointed out that a change of rules would suffice to justify classical logic. In later publications, these shortcomings were repaired.

Nevertheless, the paper takes a big step forward. There are now indications of logical rules for all the principal operators of first-order logic. There is a clear insight that the universal and the existential quantifier are to be distinguished by the identity of the discussant who is called upon to select a substituent for the bound variable. If P claims $\forall x Q(x)$ it is O who makes the choice, but if P claims $\exists x Q(x)$ it is P himself who selects (upon O's challenge) the term to be substituted for x [Lorenzen and Lorenz, 1978, p. 12]. Also, dialogical tableaux are introduced to describe strategies for P, and the similarity with Beth-tableaux is noted [Lorenzen and Lorenz, 1978, p. 11].[9]

About the second part of the paper, I shall be brief. In it Lorenzen shows how inductive definitions may introduce new dialogue-definite predicates. (A predicate is dialogue-definite if its substitution instances are dialogue-definite.) For instance, let a predicate $Q(x)$ be introduced by an inductive definition based on dialogue-definite predicates and statements, say one with the basic clause $A(x) \Rightarrow Q(x)$

---

[9]Beth [1955; 1959], see also [Beth, 1962]. Lorenzen notified Beth of this similarity in a letter dated 17 August 1959 in which he presents a dialogical tableau in first-order language for the syllogistic argument form *Festino*. This could be the first communication of a dialogical tableau.

and the inductive clause $Q(x) \wedge R(x, y) \Rightarrow Q(y)$. If P claims that $Q(n)$ he may try to defend this by having recourse to an antecedent in either the basic or the inductive clause. Say, P uses the inductive clause stating $Q(m) \wedge R(m, n)$. Assuming that $R(m, n)$ is unproblematic, he must now defend $Q(m)$. This could be done in the same way (or, perhaps, by an appeal to the basic clause). To avoid an endless regress, P is to commit himself in advance to a maximal number of steps (in a generalized version, this becomes a series of choices generating a decreasing sequence of ordinals). If P reaches a statement $A(r)$, he will win the dialogue about $Q(n)$, if he can defend $A(r)$. If he cannot defend $A(r)$ or runs out of steps before even reaching such a statement, he will lose the dialogue about $Q(n)$. By these stipulations the predicate $Q(x)$ is made into a dialogue-definite predicate [Lorenzen and Lorenz, 1978, pp. 13–14]. This approach opens a perspective on having a dialogical arithmetic and, as Lorenzen shows, in its generalized form (with decreasing sequences of ordinals) a dialogue-definite concept of truth [Lorenzen and Lorenz, 1978, p. 15].

4   STRUCTURAL RULES

It may be clear from the preceding section that, after the early stages of dialogue logic, what was most needed was a more precise description of the rules of dialogue. Without a more definite description of what moves would be admissible in a dialogue and when they would be admissible, dialogues would never be able to serve as a basis for a concept of constructivity, nor in any definite sense, yield a logic, whether classical or intuitionist. Thus a more technical approach was called for. In this section we shall see how both Lorenzen and Lorenz worked out proposals to fill the bill. The choice of structural rules, that is rules which determine *when* a move is admissible, provides a main theme for the 1960s.

4.1   *Arithmetic and Logic as Games*

With Kuno Lorenz's dissertation [1961], dialogue logic reached the age of maturity (selections reprinted in [Lorenzen and Lorenz, 1978, pp. 17–95]). Besides giving the first complete formulations of dialogue games, Lorenz's dissertation contains the first detailed reasoning about strategies in these games and establishes some connections with other approaches to logic. This was made possible, because now, for the first time, dialogical games were defined with sufficient precision to make them amenable to metalogical analysis. For this, Lorenz availed himself of mathematical game theory. Concepts such as that of a game rule, a game situation or position, a starting position, a final position, a player, a strategy, a winning strategy, were fruitfully employed. The logical rules (*allgemeine Spielregel*) for attacks and defenses of statements of specific logical forms were listed separately as a common basis for various games. These rules were now clearly distinguished from the structural rules (*spezielle Spielregel*). The former explain *how* one may attack or defend a statement, whereas the latter stipulate *when* one is allowed

to do so [Lorenzen and Lorenz, 1978, pp. 22–23]. The terms 'logical rule' and 'structural rule' for these two sets of dialogue rules were introduced by Wolfgang Stegmüller [1964, pp. 85–87].

Lorenz's logical rules are listed below in Figure 1. This figure is quoted from Lorenz's text [Lorenzen and Lorenz, 1978, p. 38], with only slight changes in notation. The column headed by C lists statements of various logical forms, the one headed by z' lists possible attacks on C, whereas the column under z'' lists the defenses of C against the attacks under z'.

	C	z'	z''
Konjunktion	$A \wedge B$	?l	A
		?r	B
Adjunktion	$A \vee B$	?	A
			B
Subjunktion	$A \rightarrow B$	A	B
Negation	$\neg A$	A	
große Konj.	$\forall x A$	$?n$	$A[n/x]$
große Adj.	$\exists x A$	?	$A[n/x]$

Figure 1. Logical Rules

According to Figure 1, a conjunction can be attacked in two ways, either by challenging the left conjunct or by challenging the right conjunct. The defense consists in each case by a statement of the conjunct challenged. A disjunction can be attacked only by challenging it as a whole. It is then up to the defense to select a disjunct to put forward. In the case of implication, the attack commits one to a statement of its antecedent, whereas the consequent constitutes the defense. A negation can be attacked by stating the statement it negates, there is no specific defense. A universally quantified statement can be attacked by selecting an individual constant, and defended by the corresponding substitution instance. ($A[n/x]$ stands for the statement obtained by substituting n for all free occurrences of x in A.) Finally, an existentially quantified statement can be attacked by a pure challenge, but this time the defense comprises the selection of a constant to be substituted for x and a statement of the resulting substitution instance.

In order to formulate structural rules, Lorenz introduces some terminology. Each dialogue (or initial fragment of a dialogue) can be envisaged as written in two columns. The left column contains the statements and challenges (question marks) produced by O, whereas the right column contains those produced by P. (Lorenz uses 'S' and 'W' instead of 'O' and 'P', but we shall here stick to the more common designations.) Each row in the diagram is called a *round*. A round can not contain more than one expression (statement or question mark) in each column. Rounds may either be *open* or *closed*. They are open if they contain just one attack in one column and nothing in the other column. They are closed when

the attack in the one column is parried by a defense in the other column. On top
of the diagram, there is an (improper) open round containing, in P's column, the
dialogue's initial statement, later to be called 'the thesis'. It will never be closed.
Let us assume that the thesis is logically complex, so that there is a way to attack
it. The first (proper) round is opened by O as she attacks the thesis. Then P either
closes this round by putting in a defense, or opens a new round by a new attack.
The first option does not obtain if the thesis was a negation. The last option only
exists if O's first move presented a statement and not a question mark, i.e. if the
initial thesis was either an implication or a negation. Moreover O's first statement
must be logically complex, or otherwise open to attack. The rest of the dialogue
is constructed in the same way: the discussants move alternately, opening a new
round which is then put below the others, or returning to close rounds that are
still open [Lorenzen and Lorenz, 1978, pp. 40–41].

Structural rules can now be formulated as restrictions on the opening and clos-
ing of rounds. For instance the pure rule (*reine Spielregel*) stipulates that for each
move by one of the discussants, the other discussant may introduce only one coun-
termove, either an attack or a defense. Consequently, each move must consist of a
reaction to the preceding move: attacks must take place in the round immediately
following the one which contains the statement attacked, and a defense move can
only close the bottommost round [Lorenzen and Lorenz, 1978, p. 41].

An alternative is the strict rule (*strenge Spielregel*) according to which attacks
may be postponed. It is now possible to react to a move once by an attack and
once by a defense (but not more). Further, according to the strict rule, a defense
move must close the bottommost open round, which need not be the bottommost
round [Lorenzen and Lorenz, 1978, p. 56].

The effective rule (*effektive Spielregel*) introduces an asymmetry in the roles of
Opponent and Proponent. In this type of game, a move by P can only be attacked
once, as with the strict rule, whereas there is no limit to the number of attacks
on a move by O. The rule for defenses is strict for both roles: each defense must
close the bottommost open round [Lorenzen and Lorenz, 1978, p. 63].

The antieffective rule (*antieffektive Spielregel*) mirrors the effective rule. The
rule for attacks is strict for both roles: each move may be attacked only once,
but the attack can be postponed. Defenses by O must close the bottommost open
round, but defenses by P may either do that or open a new kind of open round (a
so-called defense round). Defense rounds are never closed [Lorenzen and Lorenz,
1978, p. 67].

Finally, the maximally P-friendly rule is the classical rule (*klassische Spielregel*).
It combines the strict rule for attacks and defenses by O with the amplified possibil-
ities for attacks and defenses by P as granted by the effective and the antieffective
rule respectively [Lorenzen and Lorenz, 1978, pp. 69–70].

In all these games the rule for winning and losing stipulates that whoever's turn
it is to move will have lost when he or she cannot make a legal move, whereas the
other discussant will in that case have won. If no final position is ever reached,
which means that the dialogue is infinitely long, O will count as having won, and

P as having lost [Lorenzen and Lorenz, 1978, p. 28]. Lorenz does not tell how
this is to be carried out, but it may obviously be done by having P choose some
maximal number of steps beforehand. Thus the games are fixed, except for the
way in which to handle elementary (logically simple) statements. This leads us to
another innovation by Lorenz: the introduction of *formal* dialogue games.

In one sense of 'formal' all dialogue games are formal because they are based on
logical rules that pertain to the logical forms of statements. They are also formal
in the sense of displaying rigorous procedures. But, as we saw, Lorenzen's first
examples of dialogue games were based on meaningful statements and therefore,
in another sense, material. In his dissertation, Lorenz wanted to make a precise
distinction between logical and factual truth, and for that purpose he introduced
formal dialogue games (*formale Dialogspiele*) to complement the material games
(*faktische Dialogspiele*) [Lorenzen and Lorenz, 1978, pp. 48–50]. Whereas mate-
rial games operate with statements, formal games operate more abstractly with
statemental schemata (formulas). We saw that with Lorenzen's material dialogues,
it was not immediately obvious which strategies would establish that some state-
ment expressed a logical truth instead of a merely factual one. Lorenz tackles this
problem from the other side, by using the formal games to define which statemen-
tal schemata are logically valid (*allgemeingültig*). A dialogue-definite statemental
schema is valid iff there is, in the formal game, for P a winning strategy pertaining
to that schema [Lorenzen and Lorenz, 1978, p. 53]. If one wishes, one may then
say that statements that are substitution instances of a logically valid schema are
logically true.

Since we want a winning strategy in a formal dialogue game to show how P,
whatever concrete substitutions are made for the elementary statement-schemata
(atomic formulas), will be able to win in the corresponding material games, we
must take care that a statement of an elementary schema cannot become inde-
fensible after substitution. This makes Lorenz introduce an asymmetry between
the roles of O and P by stipulating, for formal dialogues, an additional clause to
the logical rules (*Zusatz zur allgemeine Spielregel*), which was called 'basic rule' in
[Stegmüller, 1964, pp. 85–86]. It is a rule to the effect that P may state an elemen-
tary schema only if this schema was stated before by O.[10] Let a formal dialogue
proceed according to this basic rule, and suppose that P states a propositional
variable q, then q must have been stated before by O. If then some concrete sub-
stitution is made for q, say the statement Q is substituted for q, P will, generally,
be able to defend Q. Given that Q will also have been substituted for the occur-
rence of q that was stated by O, it is obvious that P may just copy O's moves as
she starts to attack his statement of Q. Following that strategy, P will always have
the last word. (Here I presume that the structural rules will not prevent P from
following this strategy.) In general: if, in a formal dialogue, concrete substitutions
are made for elementary statement-schemata, P may copy O's moves with respect
to the substituted statements. Thus a winning strategy for a position in a formal

---

[10]Besides, one may, as a part of the basic rule, also introduce a clause that stipulates how
often P may copy an elementary schema [Lorenzen and Lorenz, 1978, p. 52].

game will (generally) guarantee winning strategies for its substitution instances in material games.

It will be clear, however, that this approach will not work if the pure rule is chosen as a structural rule. For, in games under the pure rule, O's former statements are no longer accessible, so that there can be no question of copying O's moves. Therefore the pure rule does not yield an interesting logic in this way: '*Die "reine" Logik ist leer*' (' "Pure" logic is empty') [Lorenzen and Lorenz, 1978, p. 25]. But the other types of structural rule are used by Lorenz to define formal dialogue games, each yielding a different formal logic.

It is time to present an example of a formal dialogue. The dialogue in Figure 2 is not compatible with the pure rule, but it is compatible with the strict rule and, therefore, with all the other choices of structural rules presented here. It is (with slight modifications in the notation) quoted from [Lorenzen and Lorenz, 1978, p. 57]. Here 'p' and 'q' are propositional variables.

O		P	
		0.	$((p \rightarrow p) \rightarrow q) \rightarrow q$
1.	$(p \rightarrow p) \rightarrow q$ [0]	6.	q
5.	q	2.	$p \rightarrow p$ [1]
3.	p [2]	4.	p

Figure 2. A Strict Dialogue

In Figure 2, the moves are numbered in the order in which they were brought forward in the dialogue. Moves 5 and 6 show how one may return to earlier rounds to close them. A number in brackets is put behind each attack to indicate which formula was attacked. Thus the formulas with a number in brackets behind them are attacks, they open rounds, whereas formulas without a number in brackets represent defenses and close their round. Notice that, according to the basic rule, the defense move number 6 could not have occurred immediately after move 1. It is only permissible after O stated q in move 5.

A remarkable feature of Lorenz's formal dialogue games is that they all allow the Opponent to attack other statement-schemata than the one introduced by P in his last move and to defend against another attack than the one executed by P in his last move. For instance, in Figure 2, move 5 fails to react to P's preceding move (move 4). The following structural rule, here quoted from [Krabbe, 1985a, p. 304], is adopted in none of Lorenz's formal dialogue games:

D6   After the first move (O's attack on the initial thesis), each further move by O consists of a reaction on the immediately preceding move by P.

According to a terminology introduced by Felscher [1985; 1986] Lorenz's dialogue games with the effective rule, which also lack D6, are systems of type D. Below we

shall see that Lorenzen formulated systems which incorporate D6, and which in Felscher's terminology are said to be of type E. It has proven to be rather difficult to show that these two types of systems are equivalent to one another.

4.2 Metamathematics

In his monograph *Metamathematik* [1962], Paul Lorenzen used the dialogical approach to give an interpretation of the logical operators and to define a concept of effective logical truth (validity). It seems unlikely that, when writing this monograph, Lorenzen was already fully taking into account the work by his pupil Lorenz. Anyhow, they both agree in making a clear distinction between material and formal dialogues. Though Lorenzen in [1962] did not yet distinguish them by special names, it is obvious that he starts with dialogues about statements (material dialogues) [1962, pp. 20–24] and then goes on to construct a game with dialogues about statemental schemata (formal dialogues) [1962, pp. 24–28]. The latter system is used to define effective validity [Lorenzen, 1962, pp. 28–29].

In *Metamathematik* the material dialogues are merely sketched, as they were in Lorenzen's earlier papers. There is no clear account of the structural rules for these dialogues. The rule for implication remains less than satisfactory. It now says that when one discussant, say P, claims $a \rightarrow b$, then O is either to concede this and to lose the dialogue, or to claim that a. If O claims that a, P is committed to claim b in case O can defend a [1962, p. 23]. We are not told how it is to be determined that O *can* defend a, unless it is meant that O must have completed a successful defense of a, which stipulation would make this rule for implication return to the equally unsatisfactory formulation in [1961]. Thus the precise formulation of logical rules for material dialogues is still on the way. The same holds for the structural rules: there is now an awareness that something must be said about these and an attempt to stipulate such a rule, but the result is not very convincing [1962, pp. 23–24].

With formal dialogues, however, the progress made looks more promising. In *Metamathematik* one finds a precise description of the formal language to be used in such dialogues. This is a first-order language with the same logical operators as in [Lorenzen, 1961]: $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (implication), $\neg$ (negation), $\forall$ (universal quantifier), and $\exists$ (existential quantifier); moreover there are two logical constants which I shall write as $\top$ (*verum*) and $\bot$ (*falsum*). The logical rules for these constants stipulate that $\top$ may not be challenged and that whoever claims that $\bot$ has lost the dialogue [Lorenzen, 1962, p. 24]. The second rule must have been a slip of the pen: it may hold for O, but P must be allowed to claim that $\bot$ and yet win the dialogue whenever O contradicts herself. Otherwise, the equivalence with Heyting's logic [1930], which Lorenzen claims to hold [1962, p. 31], would break down. The logical rules for the operators are not listed in a survey, as in Lorenz's dissertation. Yet one may find out what they are from Lorenzen's presentation of a formal calculus for the construction of winning strategies, for which he uses the notational devices of Beth-tableaux [1955; 1959;

1962]. From the rules of this calculus [Lorenzen, 1962, pp. 25–28] one may infer what the logical rules are supposed to be like and that they exactly agree with the logical rules formulated by Lorenz. Further, the structural rules are constrained by the requirement that this calculus for the construction of winning strategies has to agree with them. The rule for elementary formulas in this formal game also agrees with Lorenz's basic rule: There are no challenges of elementary claims and P can only make an elementary claim if O made the same claim before [Lorenzen, 1962, pp. 25–26]. The rules for winning and losing must be the logical rule for $\perp$, as amended above, and a rule to the effect that if it is someone's turn to move and he cannot make a legal move, this discussant loses the dialogue, whereas his adversary wins. We saw that Lorenz, too, adheres to the latter principle.

Thus we see that by writing *Metamathematik*, Lorenzen much clarified his position. He gave a better idea about what the dialogues looked liked, clearly distinguished between material and formal dialogues, and provided tableau techniques to establish winning strategies. Also, the connections with Gentzen's sequent calculi [1934] and Heyting's logic (intuitionistic logic) [1930] were clarified. On most issues Lorenzen and Lorenz were in agreement, yet the development of structural rules in Lorenzen's systems would head into a direction different from that taken by Lorenz.

4.3 Dialogue Logic in Kindergarten

In 1967 Wilhelm Kamlah and Paul Lorenzen first published their *Logische Propädeutik oder Vorschule des vernünftigen redens* (Logical Propaedeutic or Pre-School of Reasonable Discourse [1967; 1973], English translation 1984). Together with its sequel ([Lorenzen and Schwemmer, 1973] 2nd edition 1975) which was in the preface to its second edition characterized as an elementary school of technical and practical reason, the Logical Propaedeutic forms a systematic introduction to the constructive approach to language, science, and ethics of the Erlangen School. That words like 'pre-school' and 'elementary school' are not to be taken literally, may be obvious. Nevertheless, these terms are important in as far as they indicate the principle that all of the language of science and of ethics should be reconstructible and teachable on the basis of language uses grounded in everyday practice.

Somewhere in the process of reconstructing language the student is to relearn the use of logical operators as well as their connections with truth and with logical truth. It is here that dialogue logic comes in. The dialogue games provide a context in which the meaning of logical operators and the semantic and logical notions that are based on these meanings can be clarified. In [Kamlah and Lorenzen, 1967] this is done in chapter 6, which was written by Lorenzen. In the second edition, this chapter — renumbered as chapter 7 — underwent various modifications.

When writing this chapter, Lorenzen realizes that if the notion of truth of a thesis is to be given in terms of winning strategies for the Proponent in a material dialogue, a more precise determination of the structural rules is called for [Kamlah and Lorenzen, 1967, p. 200]. In going through several attempts to formulate an adequate structural rule, the text then shows something of the ongoing struggle about this issue that was so characteristic for the sixties and seventies. Since the same steps toward a structural rule were taken in Lorenzen's *John-Locke-Lectures* (1967–1968), published as Normative Logic and Ethics [1969], I shall insert references to the latter publication as well.

A *first version* makes the following stipulations: P starts by putting forward his thesis. After that, the players move alternately. Each player may attack only one of the statements put forward by his adversary or defend himself against one attack by his adversary [Kamlah and Lorenzen, 1967, p. 201]; [Lorenzen, 1969, p. 27].

After discussion of a particular example from the point of view of what constitutes reasonable argument, Lorenzen deems it plausible that defenses should not be given in any order, but rather obey the last-in-first-out principle. This leads to a *second version*: Each player may attack only one of the statements put forward by his adversary or defend himself against the last attack by the adversary against which he has not yet defended himself [Kamlah and Lorenzen, 1967, p. 202]; [Lorenzen, 1969, p. 28].

It is then observed that this does not prevent O from protracting the dialogue by endlessly repeating her attacks. The same holds for P, but it is in P's interest to complete the defense of his thesis and therefore to try and avoid repetitious behavior. On the other hand, O has not made a statement at the initial stage of the dialogue and therefore is only interested in preventing P from completing his task. So, some limit on repetitions by O has to be imposed. Therefore, as a *third version* it is suggested to stipulate that after the thesis has been put forward, O should select a natural number m that is to stand for the maximal number op repetitions of attacks by O on a statement by P. A thesis T is to be called 'true' if P can win the dialogue about T, whatever the choice of m [Kamlah and Lorenzen, 1967, pp. 202–203]; [Lorenzen, 1969, p. 28].

But, says Lorenzen, there is a simpler structural rule (*allgemeine Spielregel*) that leads to the same set of true theses (true in the sense that P can always win). Here he refers to a technical paper by Kuno Lorenz (of which Lorenzen must have seen the manuscript and have known the date of publication) in which various aspects of dialogue theory were closely investigated [1968]. This simpler structural rule is no other than the rule D6 mentioned above. Therefore, the *fourth version* of the structural rule runs as follows:

    1. The Proponent may attack only one of the statements put forward by the Opponent or defend himself against the last move of attack by the Opponent.

> 2. The Opponent may attack only the statement that was put forward in the preceding move by the Proponent or defend himself against the attack in the preceding move by the Proponent

([Kamlah and Lorenzen, 1967, pp. 203–204], translated from the German by teh present author, cf. [Lorenzen, 1969, p. 29].)

In *Normative Logic and Ethics* he adds:

> That this simplification of the general rule [= structural rule] does not affect the defensibility of any thesis is a logically composite metadialogical assertion. Since the meta-dialogue may be played with the unmodified general rule, there is no circularity here; but this metadialogue is too complicated to be dealt with in these lectures. The modification of the general rule serves only to simplify the dialogue; therefore I will continue to use it, although it could be dispensed with [1969, p. 29].

Thus, in the end Lorenzen proposes a system of type E (in Felscher's terminology). Though Lorenzen refers to [Lorenz, 1968] for the equivalence of E and D systems, not everyone was convinced that a complete proof could be found there. This equivalence problem was the subject of [Kindt, 1970]. There is a sketch of a proof in [Haas, 1980, §1.4], a very detailed proof in [Felscher, 1985], and a relatively quick proof, in [Krabbe, 1985a], which was based on ideas in [Lorenz, 1968].

To complete his description of the rules for material dialogues, Lorenzen formulates a simple rule for winning and losing: The Proponent has won [if and] only if the Opponent can no longer make a move [Kamlah and Lorenzen, 1967, p. 204]; [Lorenzen, 1969, p. 28]. He explains that the last move by which P wins the dialogue must either be a defense of an elementary statement, or an attack on an elementary statement that the Opponent can not defend.

This discussion of structural rules may appear rather intricate, especially if we consider that we are still in the kindergarten of reasonable discourse. Yet, we are not at the end of it. In the second edition of the *Logische Propädeutik*, the whole passage we just discussed was deleted from the text. Instead, Lorenzen now proposes three alternative rules [Kamlah and Lorenzen, 1973, pp. 213–215]. The first is called strictly constructive (*streng-konstruktiv*);[11] it agrees with Lorenz's pure rule (*reine Spielregel*). The second and the third liberalize the conditions for P, while leaving those for O as in the strictly constructive rule. All these rules, therefore, imply rule D6 and yield E-dialogues. The second rule is called the constructive structural rule (*konstruktive allgemeine Dialogregel*), it agrees with the rules quoted above from the first edition. Also, this rule is analogous to Lorenz's effective rule for D-dialogues. The third, called the classical rule of dialogue (*klassische Dialogregel*), is analogous to Lorenz's classical rule for D-dialogues. It liberalizes the conditions for P under the constructive rule as follows:

---

[11]Later Lorenzen preferred to call this rule simply *streng*, whereas he called the constructive structural rule *effektiv* [Lorenzen, 1987].

The Proponent attacks *one* of the statements put forward by the other,
or defends himself against *one* attack by the other.
([Kamlah and Lorenzen, 1973, p. 215], translated from the German by
the present author.)

These three kinds of dialogue yield three different concepts of truth, all of them
based on the idea that a statement is true if and only if the Proponent will always
be able to win the dialogue with that statement as its initial thesis: strictly con-
structive, constructive, and classical truth [Kamlah and Lorenzen, 1973, p. 217].
There follows a brief discussion about the justifications for the proposed liberal-
izations. One must, according to Lorenzen, take care that with each liberalized
rule strictly constructive truth is preserved. But this is trivial for rules that only
enlarge P's possibilities to move. The other main requirement concerns the consis-
tency of the resulting dialogue system: For each statement A, it should not be the
case that both A and $\neg A$ are true in the dialogical sense, that is, P should not have
a winning strategy for both. If we assume that $\bot$ is a statement for which P has
no winning strategy and that $\neg A$ is dialogically equivalent to $A \to \bot$, consistency
is a special case of a metatheorem to the effect that, for all statements A and B,
if P has winning strategies for both $A \to B$ and for A, he has one for B. This
theorem, which is similar to Gentzen's main theorem (Gentzen's *Hauptsatz*), is a
little above the kindergarten level. Lorenzen refers the reader to the elementary
school of [Lorenzen an dSchwemmer, 1973; 1975]. Consistency follows from the
theorem by simply letting B be $\bot$ [Kamlah and Lorenzen, 1973, pp. 217–218].

Whatever one may think of these justifications, they work equally well for con-
structive and classical dialogues, as Lorenzen recognizes. The only advantage of
the constructive rule is that it allows for a maximum of distinctions between for-
mulas (a minimum of equivalence). But some may think of this as a drawback
rather than an advantage.

Next ,let us consider formal dialogues, dialogues that use formulas (statement-
schemata) instead of statements. How do they fare in kindergarten? Lorenzen's
motivation for formal dialogues is the same as Lorenz's: some strategies in the
material games depend on truth values of elementary formulas, but others do not.
In these latter cases, the Proponent may arrange the dialogue so that he will
finally have to defend some elementary statement that has been put forward by
the Opponent as well. If P attacks this statement and O cannot defend it, P will
have won. If O can defend it, P may win the dialogue by copying O's defense.
But, unlike Lorenz,[12] Lorenzen does not propose a rule to the effect that P may
state an elementary formula only if this formula was stated by O before. At some
places, the text may suggest this, but if taken literally it does not say that this
is one of the rules. For instance the formal rule for winning and losing runs as
follows:

---

[12]And unlike Lorenzen himself in [Lorenzen, 1962].

> The Proponent has won when he has to defend an elementary formula
> after the Opponent's bringing forward of an identical elementary for-
> mula.
> ([Kamlah and Lorenzen, 1967, p. 207; 1973, p. 221], translated from
> the German by the present author.)

Though this may suggest that, for this rule to apply, O must first state the ele-
mentary formula, before P states it and O attacks it, this need not be the only
possibility. It could also be that P first states the formula, O then attacks it, and
P forces O later to state the formula herself. That this is the right interpretation is
borne out by one of the examples later on in the same text [Kamlah and Lorenzen,
1967, p. 210; 1973, p. 223].

What Lorenzen does stipulate for formal dialogues, is that in these dialogues the
Proponent is not allowed to attack elementary statements, whereas the Opponent
is free to do so (using a simple sign of challenge '?'). There is, however, no way for
the Proponent to defend himself to such an attack [Kamlah and Lorenzen, 1967, p.
206; 1973, p. 220]. Further, the structural rules are the constructive ones [Kamlah
and Lorenzen, 1967, p. 207; 1973, p. 221] (or the classical ones [1973, p. 222]) as
they were stipulated for material E-dialogues. Constructive and classical logical
truths are defined as formulas that can be defended by the Proponent against
any possible opposition in the constructive or the classical formal dialogue game,
respectively. Thus the formal games yield a foundation for logic.

As an illustration, a constructive formal E-dialogue is presented in Figure 3. The
thesis is the same as in Figure 2. Again, the moves are numbered consecutively,
but the dialogue is not analyzed as a sequence of rounds. Attacks are indicated
by the presence of a question mark

O		P	
		0.	$((p \rightarrow p) \rightarrow q) \rightarrow q$
1.	$(p \rightarrow p) \rightarrow q$?	2.	q
3.	?	4.	$p \rightarrow p$?
5.	q		
		P wins	

Figure 3. A Constructive Formal E-Dialogue

In the dialogue of Figure 3, P wins because, in the situation after move 5, O has
stated q herself, while P still has to defend his statement of q in move 2 against
O's attack in move 3 (this being the last of O's attacks).

Interestingly, formal dialogues are exclusively a kindergarten subject. They
do not return in the elementary school [Lorenzen and Schwemmer, 1973; 1975;
Lorenzen, 1987]. The reason must be that, for the foundations of logic, it is
sufficient to have the notion of a formal strategy (a strategy that does not depend
on the meaning of elementary statements) in material dialogues, and that actually,

for this purpose, the set-up of separate formal dialogue games, though meaningful, is not indispensable. Anyhow, Lorenzen gave priority to the material dialogues. In *Normative Logic and Ethics* he wrote:

> However, in order to understand the formal game, that is, in order to answer the question, why it is reasonable to spend our time with this game, we will have to remember that the formal game is a formalization of the material game. The material game has to be understood first, then it has to be formalized. The result is the formal game. With the formal game we are simulating material dialogues. [Lorenzen, 1969, p.35]

After the second edition of *Logische Propädeutik*, Lorenzen kept to the three kinds of structural rule for material dialogues he had then defined and henceforth explained logical concepts in terms of the formal strategies for material dialogues determined by these rules.

5   COMPARISONS

In this section we shall briefly discuss the general character of Jaakko Hintikka's game-theoretical semantics and of Hamblin's formal dialectic. Both were developed independently of Lorenzen's work (and of each other), and it remains to be seen to what extent these enterprises are related.

5.1   *Language-Games for Quantifiers*

In 1968 Hintikka published his paper *Language-Games for Quantifiers*, in which he introduced games of seeking and finding to explain the meaning of quantifiers. The rules of the game are very much like those we find summarized in Lorenz's table (Figure 1, above), but this similarity is not immediately obvious, because Hintikka does not analyze the game in terms of attacks and defenses. To give a good idea of what the game is like, I shall quote the rules as they are expounded in *Quantifiers, Language-Games, and Transcendental Arguments*, which is chapter 5 of [1973]. The language used is a first-order language (with $\wedge$, $\vee$, $\neg$, $\forall$, and $\exists$), which is interpreted within a domain of individuals D. Hence, the dialogues are material. We are informed that there are two players: myself and Nature. I slightly adapt the notations, to make for ease of comparison.

> At each stage of the game, a substitution-instance G of a (proper or improper) subformula of F is being considered. The game begins with F, and proceeds by the following rules:
>
> > (G.$\exists$) If G is of the form $\exists x G_0$, I choose a member of D, give it a name, say 'n' (if it did not have one before). The game is continued with respect to $G_0[n/x]$.

Here $G_0[n/x]$ is of course the result of substituting 'n' for 'x' in G_0.

(G.$\forall$) If G is of the form $\forall x G_0$, Nature likewise chooses a member of D.

(G.$\vee$) If G is of the form $(G_1 \vee G_2)$, I choose G_1 or G_2, and the game is continued with respect to it.

(G.$\wedge$) If G is of the form $(G_1 \wedge G_2)$, Nature likewise chooses G_1 or G_2.

(G.$\neg$) If G is of the form $\neg G_0$, the game is continued with respect to G_0 with the roles of the two players interchanged.

[Hintikka, 1973, pp. 100–101]

In a finite number of moves the game will reach an elementary statement. This statement is either true or false in the underlying interpretation. If it is true, I (that is the player who is then performing the role of myself) win and Nature (that is the player who is than performing the role of Nature) loses. If it is false, it will be the other way around.

It is not hard to see that this game, which will henceforth be called the Hintikka game, is equivalent to a material dialogue game with the pure (or, strictly constructive) rule. Call the player that takes on to play the role of myself at the start 'Proponent', and call his adversary 'Opponent' (remember that the roles of myself and Nature may be interchanged). Now a Lorenzen dialogue corresponding to the Hintikka game about F will start with P's putting forward of F as the thesis. P is myself and O is Nature. Until we get to a negation, each move by a rule of the Hintikka game can be analyzed as an attack by O and a defense by P in the Lorenzen game. All formulas will appear in P's column. The reversal of roles, when a formula $\neg G_0$ appears can be analyzed as an attack according to the rule for negation , so that G_0 now appears in O's column. This means that from now on O is myself and P is Nature. Each step by a Hintikka rule (until another negation operator appears) is now analyzed as an attack by P and a defense by O. This analysis can be continued until the end of the game. Also, winning and losing will concur. So each tournament according to the Hintikka game corresponds to an equivalent tournament in a material Lorenzen game. It may be seen that the converse holds as well.

So the similarities between the Hintikka games and the dialogue games in the preceding sections are obvious. What are the differences? Hintikka characterizes the Lorenzen games as 'indoor games' and his own games as 'outdoor':

In fact it seems to me that a sharp distinction has to be made between such 'outdoor' games of exploring the world in order to verify or falsify certain (interpreted) statements by producing suitable individuals and such 'indoor' games as, e.g., proving that certain uninterpreted formulae are logical truths by manipulating sequences of symbols in a suitable way. [Hintikka, 1973, p.81]

But this way to see the distinction between Hintikka games and Lorenzen Games is untenable. The basis of Lorenzen's approach to logical operators is to be found in material dialogues, which are every bit as 'outdoor' as the Hintikka game. Hintikka, however takes the formal games to be typical for the Lorenzen approach. We saw already that formal games are merely a (dispensable) tool for the study of formal strategies. They do not constitute the locus where the logical operators, foremost among them the quantifiers, are introduced.

Then what is the difference? Technically, I think, with respect to some basic games there is none. So the approaches overlap. But also, either approach pursues interests that are not mirrored by the other. On the one hand, we saw that in their treatment of material dialogues in a context where all elementary statements have a well-determined truth value, both approaches are in complete agreement. On the other hand, game-theoretic semantics does not consider different options for structural rules, and Lorenzen dialogue logic never entered the field of branching quantifiers.

However, the two approaches also display a difference in their principal orientation towards logic. Around 1970, there were three major orientations towards (or pictures of) logic: the derivational orientation, the semantic orientation, and the dialogical orientation.[13] According to the derivational orientation logical operators are implicitly defined by the system of axioms and inference rules to which they belong. Logical validity (logical consequence) is defined as derivability in such a system. According to the semantic orientation logical operators are defined by means of semantic rules that serve to calculate semantic values for linguistic objects (relative to a model). Logical validity is reconstructed as immunity from counterexample. According to the dialogical orientation logical operators are implicitly defined by the dialogue game to which they belong, whereas logical validity is defined as the availability of a formal winning strategy for the Proponent.[14]

When this rough classification is applied, Hintikka and Lorenzen systems will be subsumed in different groups. The Lorenzen approach is a truly dialogical approach, but the Hintikka approach is, notwithstanding appearances to the contrary, primarily semantic. A small piece of evidence for this is the absence of a rule for implication, which is dialogically the most interesting operator. Generally there is in game-theoretical semantics a lack of discussion about typical dialogical issues such as the different sets of structural rules and their consequences for the possibilities of winning and losing. Rather, game-theoretical semantics competes successfully with other approaches in semantics, notably model-theoretic semantics [Saarinen, 1979]. Seen in this way, game-theoretic semantics provides an ingenious way to describe semantic values, but does not, primarily, constitute an approach to various kinds of logical validity defined in terms of winning strategies.

---

[13]Orientations were called 'garbs' in [Barth and Krabbe, 1982, ch. 1].

[14]There are more orientations possible. Recently an information-theoretic orientation has been but forward [Veltman, 1996].

5.2   Formal Dialectic

Charles Hamblin's book *Fallacies* [1970] is as renowned for its historical as for its systematic chapters. Among its systematic chapters the one entitled 'Formal Dialectic' (chapter 8) invites a comparison with the work of Paul Lorenzen and Kuno Lorenz. Hamblin starts with the concept of a *dialectical system.* 'This is no more or less than a regulated dialogue or family of dialogues' [1970, p. 255]. For instance, a regulated interchange of statements about the weather, would count as a dialectic system [1970, p. 256]. Dialectic is the study of dialectic systems, and it can proceed in two ways:

> The study of dialectical systems can be pursued *descriptively*, or *formally.* In the first case, we should look at the rules and conventions that operate in actual discussions [. . .] . A formal approach, on the other hand, consists in the setting up of simple systems of precise but not necessarily realistic rules, and the plotting of the properties of the dialogues that might be played out in accordance with them. Neither approach is of any importance on its own; for description of actual cases must aim to bring out formalizable features, and formal systems must aim to throw light on actual, describable phenomena. [. . .]Dialectic, whether descriptive or formal, is a more general study than Logic; in the sense that Logic can be conceived as a set of dialectical conventions. It is an ideal of certain kinds of discussion that the rules of Logic should be observed by all participants, and that certain logical goals should be part of the general goal. [Hamblin, 1970, p. 256]

A word is in order on what is here meant by 'formal'. Since this term is used by Hamblin as an opposite of 'descriptive', its meaning must be different from any of the three meanings we met before (rigorous, related to logical form, operating with formulas rather than statements). Rather, 'formal' here means nondescriptive or normative. We may observe that the Lorenzen dialogues (whether material or not) are formal in this sense as well, and that, therefore, their study (dialogue logic) must be belong to formal dialectic. Moreover, as Hamblin says, formal dialectic is a more general study than logic, and since dialogue logic is a part of logic, formal dialectic must be a more general study than dialogue logic.

This will even hold for that part of formal dialectic that studies dialectical systems in which rules of logic are observed. For, in formal dialectic these rules of logic need not take the shape of rules of dialogue logic as in Figure 1, above. They could appear as consistency requirements based on a logic formulated in a non-dialogical way (e.g. in terms of derivations). Indeed, none of the formal dialectic systems that Hamblin introduces in his chapter on formal dialectic comprises formal dialogue rules giving an analysis of the meaning of logical operators in terms of possible attacks and defenses. Also, there is no definition of logical validity in terms of winning strategies. We conclude that Hamblin's formal dialectic is not

an analogue of dialogue logic, but a more comprehensive study which shares with
dialogue logic an interest in dialectic rules.

The same observation holds for numerous other studies that are concerned with
dialectic, dialogue, or language games. I'm thinking of, among other things, Jaakko
Hintikka's papers on information-seeking dialogues (e.g. [1981]), Lauri Carlson's
dialogue theory [1983], Nicholas Rescher's dialectics [1977], and the follow-up on
Charles Hamblin's work on dialectical systems as undertaken by himself [1971],
by Jim Mackenzie, in many papers, e.g. [1979; 1985; 1990], and by John Woods
and Douglas Walton, also in many papers, see [1989], as well as monographs, (e.g.
[Walton, 1984; 1998]). These studies can not be discussed in this chapter, which
is exclusively concerned with dialogue logic in the more restricted sense. That is
not to deny that there are notable relationships and influences between dialogue
logic and these other types of dialectic. In fact, Lorenzen dialogues and Hamblin
dialogues were at a certain point integrated into one dialectical system [Walton
and Krabbe, 1995].

6   FURTHER DEVELOPMENTS AND APPLICATIONS

Returning to dialogue logic proper, I shall now briefly mention some of the de-
velopments that followed upon the kindergarten era. These developments concern
the introduction to modalities, a reframing of initial situations of dialogues and of
dialectical roles, new motivations for structural rules, and metalogical proofs. At
the end I shall review the various uses that have been made of dialogue logic.

6.1   Modalities in Dialogue

For an early formulation of dialogical rules for modal operators, we must actually
go back to the sixties. Lorenzen's *Normative Logic and Ethics* [1969] already
contains a section on modal logic. In it Lorenzen introduces a symbol for necessity
(here I shall use '$\Box$') that can be indexed by a reference to a system of (modality-
free) sentences Σ. The symbol '$\Box_\Sigma$' can then be read as 'necessary relative to Σ',
and is defined as follows:

$$\Box_\Sigma A \textbf{ iff } \Sigma \text{ logically implies } A.$$

Here 'A' stands for a modality-free sentence [Lorenzen, 1969, p. 62]. The notion
of logical implication can of course be clarified by dialogue logic:

> Σ logically implies A **iff** A is defensible as a thesis when the formulas Σ
> are given as hypotheses (initial statements made by O). (Cf. [Lorenzen,
> 1969, pp. 30, 34–35].)

Lorenzen observes that there are theorems for which the choice of Σ does not
matter, so that we may suppress the index. These theorems can be called 'modal-
logically true', for instance:

$$\Box A \wedge \Box B \rightarrow \Box (A \wedge B) [1969, \text{ p. } 62]$$

Lorenzen then characterizes the new, unindexed, operator by dialogical rules. These rules are to be formulated without any reference to Σ, which remains suppressed. The attack-defense-rule for $\Box$-formulas, that is their logical rule, is displayed in Figure 4.

	Statement form	Attack	Defense
Necessity	$\Box A$	?	A

Figure 4. The Logical Rule for Necessity[14]

This rule would make '$\Box$' into a void operator (like 'it is the case that ...'), if it weren't for a special structural rule:

> $\Box$-*defense*-rule: If the proponent defends a $\Box$-formula he may attack only the $\Box$-formulae (the beginning $\Box$ deleted) put by the opponent beforehand [Lorenzen, 1969, p. 62].

We must take care not to interpret this rule in a way that would prevent P from defending an attacked thesis $\Box A$ by a counterattack against an hypothesis $B \wedge \Box A$, say using ?r in order to obtain $\Box A$ as a new hypothesis. This P is allowed to do, even though $B \wedge \Box A$ is not a $\Box$-formula. It is only *after* P has executed his defense by asserting A, as prescribed by the logical rule for necessity, that the non-$\Box$-formulas are as it were removed from the set of hypotheses [Krabbe, 1982a, pp. 207–208].

   Lorenzen thus formulates a dialogical system for necessity in a nutshell. He then goes on to discuss modal syllogistic and deontic modalities, but he seldom returns to the dialogical basis. Apparently, Lorenzen did not set great store by these rules of modal dialectics, for they do not reappear in his later treatments of modality. In these later treatments, e.g. [Kamlah and Lorenzen, 1973; Lorenzen and Schwemmer, 1973; 1975; Lorenzen, 1987], modalities are still critically reconstructed in terms of dialogical procedures, but the dialogues in question are not modal dialogues but material dialogues, in a metalanguage, about logical implications. The notion of modal implication is explained in terms of generally applicable winning strategies on this metalevel (i.e. strategies that do not depend on Σ). To characterize the class of correct modal implications one may indeed introduce a system of modal dialogues, but it can also be done in other ways. This explains why the modal dialogues of *Normative Logic and Ethics* [Lorenzen, 1969] could gradually disappear from the scene. In *Normative Logic and Ethics* there are still two dialogue rules (quoted above). In [Kamlah and Lorenzen, 1973, p. 227] these are replaced by a $\Box$-rule (Δ-*Regel*) without an indication of how

---

[15][Lorenzen, 1969, p. 64]

the players are to apply this rule in dialogue (i.e. of who should, or may, perform what act). In [Lorenzen and Schwemmer, 1975, p. 116] the □-rule (now called Δ-*Schritt*) is no longer presented as a dialogue rule of an independently formulated modal dialogue system, but as a reduction rule in a system for the construction of modal tableaux (similar to the deductive Beth-tableaux of [1959; 1962]), the tableau system being designed to characterize exactly the correct modal implications [Krabbe, 1982a, pp. 208–209].

All the same, the considerations in [Lorenzen, 1969] served as a starting point for the development of a truly dialogical modal logic. In [Krabbe, 1982a; 1985; 1986] this line of research has led to multiply modal systems (i.e. systems with many different necessity operators) on the basis of an intuitionistic propositional logic and displaying an S4-like character. Jaap Hoepelman and Toine van Hoof applied dialogue logic to generic statements and conditionals and introduced interesting role reversals in logical dialogues [Hoepelman and Van Hoof, 1993; Van Hoof, 1995]. Also modal dialogue logic (and many other special dialogue logics) were studied by Shahid Rahman and Helge Rückert in a number of papers, e.g. [1999; 2001b].

6.2   The Initial Situation and Dialogical Roles

In most systems of dialogue rules the Opponent is to make the first move and this move must be an attack directed at the initial thesis. Sometimes P's assertion of the initial thesis is counted as an improper move (move zero) that precedes the actual dialogue. The initial situation most often is supposed to comprise just one statement, the thesis, and there are no initial statements made by O. But we just saw that in *Normative Logic and Ethics* [Lorenzen, 1969] the possible presence of initial hypotheses was taken into account. So by a more general definition the initial situation comprises a thesis (put forward by P) and set of statements put forward by O. The latter set may be empty, bringing us back to the simpler type of initial situation.

In an article published in [1980], Gerrit Haas staunchly defended the importance of having a definition that admits initial hypotheses. But he also proposed another and more radical change: the first proper move should be made by P instead; nevertheless the case of an empty set of initial hypotheses should not be excluded. What then would P's first move put forward when there is, for P, no statement by O to attack? The solution was that P should (in an improper zero move) *announce* the thesis but not yet *assert* it. So, when the dialogue starts, P may as his first proper move assert the announced thesis. This guarantees that P can always make a first move, whether the set of hypotheses is empty or not. If there are statements by O that P can attack, P may opt to do that first and to assert the thesis later. The status of an announced but not yet asserted thesis equals that of a potential defense that P has not yet realized. For instance, let O's last attack have been on a statement $A \to B$, then P has the option to assert B in a defense move, as long as he did not yet do so. Hence, as long as the defense has not been realized one

can say of such a potential defense B as well that it was announced, but has not yet been asserted. Haas's approach leads to elegant systems and works out very nicely in the metatheory.

Closely related to the way one constructs the initial situation is the way one views the two dialectical roles. These became gradually to be seen as more radically divergent ('asymmetric'). The assertions on O's side were seen not to constitute theses, but only hypotheses, or better concessions, that P may use to defend his thesis. So the parlance of 'attack and defense' came to be felt as inappropriate. This was also caused by its unnecessary martial sound, but mainly by the difference between the two dialectical roles. In [Walton and Krabbe, 1995] this set of terms is replaced by different sets for O and P: O challenges and P defends, P questions and O answers.

Mostly, P is pictured as a serious discussant that tries to convince O of a thesis he himself beliefs to be true. But beliefs play no part in formal dialogues and an insincere Proponent is formally indistinguishable from a sincere one, as long as he fulfills all his dialectical obligations; nevertheless it is important to have a picture of the normal situation. In [Krabbe, 1982] another picture of the normal situation has been proposed. According to this picture it is O who has an opinion, and P who puts this to the test. O's hypotheses, or concessions, together express a theory O wants to maintain. P as a critic puts forward as a thesis, some statement O would be inclined to reject. This is called a 'provocative thesis'. Then P claims that given O's position she must also accept the provocative thesis. This kind of criticism of O's position is called 'immanent criticism'. Clearly, there is no need for P to believe his thesis to be true, even if P is sincere. The thesis could even be $\perp$, meaning that P claims that O's theory is inconsistent. From this picture, some of the customary rules of dialogue can be made more plausible.

6.3 Structural Rules Again

In the meantime, the debate on structural rules did not stop. There were proposals of other motivations to support particular choices of rules, as well as investigations about the effect of certain choices. These matters were a concern of, for instance, [Barth and Krabbe, 1978; 1982; Haas, 1980; Felscher, 1986; Valerius, 1990].

A problem for some motivations was that it was felt to be highly desirable to end up with a system that yielded a respectable logic, most often intuitionistic logic, but that, at least by some, it was not seen as permissible to let this desire be part of the motivation. Yet, wittingly or unwittingly, dialecticians could be influenced by the desire to lay the foundations for some interesting logic (and prove an equivalence theorem, see below). Even the extensive motivations given by Barth and Krabbe, though officially resting on an analysis of norms for processes of conflict resolution, may have been biased by the prospective to generate foundations for intuitionistic logic.

The analysis in [Barth and Krabbe, 1978; 1982] starts from the concept of a conflict of avowed opinions. The idea is that such a conflict can be resolved by

a discussion that satisfies certain norms. There is a proposal for a number of fundamental norms, which are stated in rather general terms, and an attempt to implement these norms by more specific norms. Acting in accordance with the more specific norms is supposed to help achieve the ends of the more general norms. Going down to ever more specific norms one finally reaches the level of rules of dialogue which define a dialogue game (or a number of alternative games).[16] For instance, take *the fundamental norm of dynamic dialectics*:

> FD D1   The system of FD-rules applied in a discussion shall be designed *to promote the revision and flux of opinions* in any company in which these rules are adopted. [Barth and Krabbe, 1982, p. 29]

If one accepts this norm, one may implement it by also accepting:

> FD D2   The rules shall be such that *unavoidable* decisions as to the outcome of discussions will be reached *as soon as possible*. [Barth and Krabbe, 1982, p. 79]

This can be implemented by a number of rules that avoid repetitions and verbosity. Clearly, this road to structural rules is in principle feasible in any kind of formal dialectic, not just dialogue logic.

6.4   Metalogic

Metalogic is the (mathematical) study of properties of logic systems and the relations between these systems. In the case of dialogue logic, the logic yielded by a system of dialogue rules may be identified with the set of sentences for which there is a P-winning strategy (i.e. a wining strategy for P), given that P starts with a statement of the sentence in his improper move (move zero), and that O starts with the empty set of hypotheses. In metalogic, the logics yielded by dialogue systems are compared with one another and with logics yielded by other types of system.

A P-winning strategy can be pictured as a tree of which the branches represent all possible dialogues in which P uses the strategy. The nodes are labeled by moves (or, alternatively, by positions in the dialogue game). Whenever it is O's turn to move, all options for O are represented (this generally leads to a branching of the tree), but whenever it is P's turn to move, just one particular option is represented. This is a P-strategy. A P-winning strategy is a P-strategy such that the final nodes all represent moves that determine a win for P (or, alternatively, a position in which P has won), and such that there are no infinite branches.[17] Starting with [Lorenzen, 1961], P-winning strategies are mostly represented by

---

[16]For a brief exposition, see [Barth, 1982].

[17]In game-theoretical terms, P-strategies and P-winning strategies are parts of the game in extensive form.

dialogical tableaux. In Figure 5 an example is given of a dialogical tableau representing a P-winning strategy for the same thesis as in Figure 2 and Figure 3. The dialogue system is that of constructive and formal E-dialogues. To read a dialogical tableaux, one should remember that, when a tableau splits its columns, this represents a branching of the tree. In that case the leftmost O-column is to be associated with the leftmost P-column to get one branch of the tree, and the rightmost O-column is to be associated with the rightmost P-column to get the other branch.[18]

O		P	
		0. $((p \to p) \to q) \to q$	
1. $(p \to p) \to q$?		2. q	
3. ?		4. $p \to p$?	
5. p?	5. q	6. p	
7. ?			P wins
		P wins	

Figure 5. A Dialogical Tableau

The P-winning strategies of a system of dialogue rules must be fully analyzed in order to prove correctness (soundness) and completeness theorems for the logic yielded by the system. Here correctness and completeness are relative concepts. A logic system $\mathbf{L}_1$ is correct relative to a logic system $\mathbf{L}_2$ if whatever is valid in $\mathbf{L}_1$ is also valid in $\mathbf{L}_2$. Most often $\mathbf{L}_1$ is a derivational or dialogical system, and $\mathbf{L}_2$ a semantic system, but here we shall not require this to be the case. A logic system $\mathbf{L}_1$ is complete relative to a logic system $\mathbf{L}_2$ if whatever is valid in $\mathbf{L}_2$ is also valid in $\mathbf{L}_1$. A logic system $\mathbf{L}_1$ is equivalent to a logic system $\mathbf{L}_2$ if $\mathbf{L}_1$ is both correct and complete relative to $\mathbf{L}_2$. Much metatheoretic research in dialogue logic has been focusing on such relations, both to relate different dialogical logics and to make comparisons with logic systems defined by derivational or semantic methods.

The problem with metalogical proofs about a dialogue system is that one first has to specify the system in mathematical terms; that is, one has to define the set of dialogue situations and the game rule, using for instance the apparatus of set theory. This can often be done in various ways that are not always obviously equivalent. Therefore the proofs are either very informal, and therefore not convincing for everyone, or very rigorous, but limited to a particular specification in mathematical terms. Nevertheless it is clear by now that for the main systems of dialogue logic such proofs are available.

The equivalence between logics generated by (particular kinds of) D-dialogues and (particular kinds of) E-dialogues proved a notoriously hard nut to crack. To show this equivalence, one needs rather complex transformations of strategies. Above, when discussing Lorenzen's simplification of the structural rule (see *4.3 Dialogue Logic in Kindergarten*), we saw that this matter was taken up by a

---

[18]These conventions are, of course, known from Beth-tableaux [1955; 1959; 1962].

number of authors.[19] As far as I can see, Felscher's proof is correct, but very complicated, whereas mine is also correct, but not as formal and much simpler.

As to the equivalence between dialogue logics and other (semantic, derivational) logics, this has been studied by tableau methods since [Lorenz, 1961]. A number of proofs and sketches of proofs have been published since, widely diverging in methods and pertaining to different types of system: [Kindt, 1972, §10; Thiel, 1978; Haas, 1980; Mayer, 1981; Stegmüller and Varga von Kibéd, 1984; Felscher, 1985; Barth and Krabbe, 1982; Krabbe, 1982b; 1985a; 1988]. Kindt's work is very mathematical, and it is uncertain whether his specification corresponds to the dialogue systems in the literature, Haas treats systems with his own kind of initial situation, Mayer treats intuitionistic and classical logic, Stegmüller and Varga von Kibéd treat classical logic only, Felscher presents very detailed proofs for intuitionistic logic, Barth and Krabbe treat besides intuitionistic and classical logic also minimal logic, but restrict themselves to propositional logic, [Krabbe, 1982b] gives the additions for predicate logic, [Krabbe, 1985a; 1988] present ever quicker ways to prove equivalence.[20]

6.5   A Summary of the Uses of Dialogue Logic

What is the use of logical dialogue games? The answer has to be that with various authors, or even with the same author at different times, one finds rather distinct purposes. In his first publications on dialogue logic, Paul Lorenzen wanted to amend the foundations of mathematics. The goal was *to establish a constructive mathematics and to justify some constructive logic* (e.g. the intuitionistic logic formulated by Arend Heyting [1930]), but to do so without absorbing the solipsistic (one mind) philosophy of Heyting's teacher, L. E. J. Brouwer. Soon it became evident that, by varying the rules of dialogue, a dialogical foundation could also be given to other logics than Heyting's, including classical logic to which all constructivists were in some way opposed. This is not to say that the case for constructive (intuitionistic) logic was now lost, because one could still try to show that those rules of dialogue that yield a constructive logic are for some reason to be preferred.

In the years that followed, dialogue logic developed into an independent dialogical orientation towards logic that can compete with the semantic and with the derivational orientations as a means *to give a characterization of logical operators (logical constants)* — i.e. to determine the meaning of logical operators. In dialogue logic this is done by rules that stipulate in what way statements of various logical forms can be attacked and defended.

This dialogical orientation towards logic can also compete with the semantic and the derivational orientations as a means *to provide precise meanings for concepts of theoretical logic*, such as 'validity' or 'consistency'. In dialogue logic, definitions for these concepts are framed in terms of strategies. As we saw, the logical validity of a thesis can be defined as the existence of a winning strategy for the Proponent.

---

[19][Lorenz, 1968; Kindt, 1970; Haas, 1980; Felscher, 1985; Krabbe, 1985].
[20]See [Felscher, 1986, pp. 349–353] for a more extensive evaluation.

Consistency of a set of statements can be defined as the existence of a winning strategy for the Opponent who grants these statements vis-à-vis a Proponent who asserts a conventionally indefensible sentence, codified as '⊥'.

Lorenzen realized that the dialogical approach could also be deployed *to contribute to the foundation of other than mathematical uses of language.*[21] Together with a number of German philosophers (known as the 'constructivists' or as the 'Erlangen School'), he worked for a considerable period towards a critical reconstruction of the *Bildungssprache* (the language of culture, i.e. the language of science and philosophy, and so on). Their aim was to provide the intellectual means to end the present lack of discipline as people are writing nineteen to the dozen and talking at cross-purposes; to end the chaos in communication, for short [Kamlah and Lorenzen, 1967; 1973, p. 11]. For this we must reconstruct our language step by step, making sure that each part is thoroughly understood by its users. Point of departure in this enterprise are those speech acts for which it is sufficiently clear how they should be executed and for what purpose (for instance simple orders). This so-called 'empractical' (*empraktisch*, also: *empragmatisch*) use of language is safely kept in check by nonlinguistic action [Lorenzen and Schwemmer, 1975, p.22]; [Lorenzen, 1987, p. 20]. All further steps in the construction of language must be teachable and verifiable as to their purpose [Lorenzen and Schwemmer, 1975, pp. 10–11]; [Lorenzen, 1987, p. 10]. Going through this process, we may finally reach terms such as 'synthetical *a priori* truth', 'coincidental', 'social structure', or 'or', and admit them in the reconstructed language. The language thus reconstructed is called 'ortholanguage' (*Orthosprache*). Now, what is the role of dialogue games in this program? It is that at a certain stage in this process of reconstruction one has to reintroduce the logical operators, and this is done by explaining the rules of logical dialogue games. In the process, the dialogue games are preceded by a 'rational grammar'(*rationale Grammatik*) [Lorenzen and Schwemmer, 1975, p. 55]; [Lorenzen, 1987, p. 52] for parts of speech and elementary sentences, including a survey of 216 well-founded locative prepositions. They are followed by further reconstructions, pertaining to arithmetic, geometry, ethics, politics, etc.

Another objective of dialogue logic is *to provide models for argumentation theory*. This aim is not inconsistent with the ideal of an ortholanguage, but neither need it be restricted to that context. Let me briefly point out why argumentation theory, more specifically *dialectical* argumentation theory, needs models of dialogue. In the dialectical approach to argumentation it is assumed that in arguments there are always two roles in play, even when just one person is putting forward an argument so that the role of the Opponent remains implicit (monologues). The starting point for all arguments is found in differences of opinion. The goal of an argumentative process is to resolve a difference of opinion so as to reach a solid and well-founded agreement. It is not sufficient just to settle the difference by negotiation or to put an end to it in some ad hoc way. Therefore, the argumentative process must consist of a serious and critical discussion of the

---

[21]In these last paragraphs I paraphrase some of the remarks I made in [Hodges and Krabbe, 2001, pp. 36–38].

issues. The ideal format of this process is to be given by a model of discussion. Real life argumentative discussion may not be in accordance with this ideal format, but the theorist needs a model to analyze and evaluate what actually goes on.

Models of argumentative discussion can be more or less formal. In the approach to argumentation called 'pragma-dialectics' informal models based on speech act theory are used [Van Eemeren and Grootendorst, 1982; 1984; 1992; 2004], in a formal dialectic approach Lorenzen-type models are nowadays combined with Hamblin-type models [Walton and Krabbe, 1995].[22] Formal dialogue games, and in general formal dialectic systems, of many different types constitute a kind of laboratory for the argumentation theorist, where small scale conceptual experiments are possible regarding concepts such as: making a claim, granting a concession, useless versus useful repetition, burden of proof, blunder, fallacy, relevance, being in the right versus being put in the right, etc. Further experiments concern different options for rules of dialogue as well as the interaction of rules of dialogue that seem separately plausible. Thus the study of dialogue games, together with other kinds of formal dialectic, may stepwise contribute to a comprehensive theory of argumentation.[23]

Finally, it can be seen that dialogue logic, together with other dialectically oriented studies in logic, continues *to relate with work in artificial intelligence and linguistics.*[24] A recent issue of *Synthese*, edited by Shahid Rahman and Helge Rückert testifies of the influence of dialogue logic in several directions and of the continuing interest in dialogue logic itself [2001a].

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

[Aristotle, 1965] Aristotle. On sophistical refutations. In Aristotle, *On Sophistical Refutations,*

---

[22]See [Hegselmann, 1985] for another kind of formal model, combining dialogues and internal reasoning.

[23]For a survey of contemporary argumentation theory see [Van Eemeren, *et al.*, 1996]. Dialogue logic and formal dialectic are discussed in chapter 9.

[24]I mention a few works that appeared in the nineties: Gerard Vreeswijk on defeasible reasoning [1993], Toine van Hoof on conditionals [1995], Joachim Machate on stress rules [1996], Richard Starmans on argument based reasoning [1996], Peter Fach on man-machine interaction between man and computer [1997], Maria Strobel on man-machine interaction by telephone [1997]. Nowadays an arguing machine is no longer an outrageous concept: see the book edited by Chris Reed and Timothy J. Norman, *Argumentation Machines* [2003].

On Coming-To-Be and Passing-Away On the Cosmos. Volume [400/442] of *The Loeb Classical Library.* Harvard University Press–William Heinemann, Cambridge, MA–London, 1965. Translated by E.S. Forster. First edition published in 1955.

[Aristotle, 1976] Aristotle. Topica. In Aristotle, *Posterior Analytics, Topica.* Volume 391 of *The Loeb Classical Library.* Harvard University Press–William Heinemann, Cambridge, MA–London, 1976. Translated by E.S. Forster. First edition published in 1960.

[Astroh, *et al.*, 1997] Michael Astroh, Dietfried Gerhardus, and Gerhard Heinzmann, editors. *Dialogisches Handeln: Eine Festschrift für Kuno Lorenz.* Spektrum Akademischer Verlag, Heidelberg, Berlin & Oxford, 1997.

[Barth, 1982] Else M. Barth. A normative-pragmatical foundation of the rules of some systems of formal$_3$ dialectics. In [Barth and Martens, 1982, pp. 159–170].

[Barth and Krabbe, 1978] Else M. Barth and Erik C.W. Krabbe. Formele$_3$ dialectiek: instrumenten ter beslechting van conflicten over geuite meningen. [Formal$_3$ Dialectics: Instruments for the Resolution of Conflicts of Avowed Opinion], *Spektator: tijdschrift voor neerlandistiek*, 7:307–341, 1978.

[Barth and Krabbe, 1982] Else M. Barth and Erik C.W. Krabbe. *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation.* Walter de Gruyter, Berlin & New York, 1982.

[Barth and Martens, 1977] Else M. Barth and Jo L. Martens. Argumentum ad Hominem: From chaos to formal dialectic: The method of dialogue-tableaus as a tool in the theory of fallacy. *Logique et Analyse*, n.s. 20e Année:76–96, 1977.

[Barth and Martens, 1982] Else M. Barth and Jo L. Martens, editors. *Argumentation: Approaches to Theory Formation: Containing the Contributions to the Groningen Conference on the Theory of Argumentation, October 1978.* John Benjamin, Amsterdam, 1982.

[Beth, 1955] Evert W. Beth. *Semantic Entailment and Formal Derivability.* North-Holland, Amsterdam, 1955. *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde, n.r. 18.* Reprinted in Jaakko Hintikka, editor, *The Philosophy of Mathematics*, pages 9–41. Oxford University Press, London, 1969.

[Beth, 1959] Evert W. Beth. Considérations heuristiques sur les méthodes de déduction par séquences. *Logique et Analyse*, n.s. 2e Année:153–159, 1959.

[Beth, 1962] Evert W. Beth. *Formal Methods: An Introduction to Symbolic Logic and to the Study of Effective Operations in Arithmetic and Logic.* Reidel, Dordrecht, 1962.

[Brouwer, 1908] L.E.J. (Bertus) Brouwer. De onbetrouwbaarheid der logische principes [The unreliability of the logical principles]. *Tijdschrift voor wijsbegeerte* 2:152–158, 1908. Translation in L.E.J. Brouwer, *Collected Works I: Philosophy and Foundations of Mathematics*, edited by Arend Heyting, pages 107–111. North-Holland–American Elsevier, Amsterdam–New York, 1975.

[Carlson, 1983] Lauri Carlson. *Dialogue Games: An Approach to Discourse Analysis.* Reidel, Dordrecht, 1983.

[Drieschner, 1966] Rudolf Drieschner. *Untersuchungen zur dialogischen Deutung der Logik.* PhD Disseration, Hamburg, 1966.

[Fach, 1997] Peter W. Fach. *Dialogspiele in der Mensch-Computer Interaktion: Eine Fallstudie am Beispiel von Online-Hilfen.* PhD Dissertation, Stuttgart University, Stuttgart, 1997. Waxmann, Münster, 1997.

[Felscher, 1985] Walter Felscher. Dialogues, Strategies, and Intuitionistic Provability. *Annals of Pure and Applied Logic*, 28:217–254, 1985.

[Felscher, 1986] Walter Felscher. Dialogues as a Foundation for Intuitionistic Logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic III: Alternatives to Classical Logic*, pages 341–72. Reidel, Dordrecht, 1986.

[Fritz, 1994] Gerd Fritz. Formale Dialogspieltheorien. In Gerd Fritz and Franz Hundsnurscher, editors, *Handbuch der Dialoganalyse*, pages 131–152. Max Niemeyer, Tübingen, 1994.

[Gentzen, 1934] Gerhard Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift* 39:176–210 and 405–431, 1934. Translated as 'Investigations into Logical Deduction'. In M.E. Szabo, editor, *The Collected Papers of Gerhard Gentzen*, pages 68–131. North-Holland, Amsterdam and London, 1969.

[Gethmann, 1980] Carl Friedrich Gethmann, editor. *Theorie des wissenschaftlichen Argumentierens.* Suhrkamp, Frankfurt-on-Main, 1980.

[Gethmann, 1982] Carl Friedrich Gethmann, editor. *Logik und Pragmatik: Zum Rechtfertigungsproblem logischer Sprachregeln.* Suhrkamp, Frankfurt am Main, 1982.

[Haas, 1980] Gerrit Haas. Hypothesendialoge, konstruktiver Sequenzenkalkül und die Rechtfertigung von Dialograhmenregeln. In [Gethmann, 1980, pp. 136–161].

[Haas, 1984] Gerrit Haas. *Konstruktive Einführung in die formale Logik*. Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1984.

[Hamblin, 1970] Charles L. Hamblin. *Fallacies*. Methuen, London, 1970. Reprinted with a preface by John Plecnik and John Hoaglund. Vale Press, Newport News, VA, 1986.

[Hamblin, 1971] Charles L. Hamblin. Mathematical Models of Dialogue. *Theoria: A Swedish Journal of Philosophy* 37:130–155, 1971.

[Hegselmann, 1985] Rainer Hegselmann. *Formale Dialektik: Ein Beitrag zu einer Theorie des rationalen Argumentierens*. Felix Meiner, Hamburg, 1985.

[Heyting, 1930] Arend Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preuß ischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, pages 42–56, 1930.

[Hintikka, 1968] Jaakko Hintikka. Language-Games for Quantifiers. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 46–72. Blackwell, Oxford, 1968. Reprinted in [Hintikka, 1973, pp. 53–82].

[Hintikka, 1973] Jaakko Hintikka. *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*. Clarendon & Oxford University Press, Oxford, 1973.

[Hintikka, 1981] Jaakko Hintikka. The Logic of Information-Seeking Dialogues: A Model. In Werner Becker and Wilhelm K. Essler, editors, *Konzepte der Dialektik*, pages 212–231. Vittorio Klostermann, Frankfurt am Main, 1981.

[Hodges and Krabbe, 2001] Wilfrid Hodges and Erik C.W. Krabbe. Dialogue Foundations: Part I: A Sceptical Look & Part II: Dialogue Logic Restituted [title misprinted as 'Dialogue Logic Revisited']. *The Aristotelian Society (The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of York July 2001)*, Supplementary Volume 75:17–32 & 33–49, 2001. Part I written by Wilfrid Hodges and Part II written by Erik C.W. Krabbe.

[Hoepelman and Van Hoof, 1993] Jaap Hoepelman and Toine van Hoof. Default and Dogma. In [Krabbe, et al., 1993, pp.141–158].

[Johnson and Blair, 1994] Ralph H. Johnson and J. Anthony Blair. *Logical Self-Defense*. Third edition. McGraw-Hill Ryerson, Toronto, 1994. First edition published in 1977.

[Kambartel, 1979] Friedrich Kambartel. *Überlegungen zum pragmatischen und zum argumentativen Fundament der Logik*. In [Lorenz, 1979, pp. 216–228].

[Kambartel and Mittelstraß, 1973] Friedrich Kambartel and Jürgen Mittelstraß, editors. *Zum normativen Fundament der Wissenschaft*. Athenäum, Frankfurt-on-Main, 1973.

[Kamlah and Lorenzen, 1967] Wilhelm Kamlah and Paul Lorenzen. *Logische Propädeutik oder Vorschule des vernünftigen Redens*. Revised edition. Volume 227, *Hochschultaschenbücher*. Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1967.

[Kamlah and Lorenzen, 1973] Wilhelm Kamlah and Paul Lorenzen. *Logische Propädeutik: Vorschule des vernünftigen Redens*. Second improved and enlarged editon. Volume 227, *Hochschultaschenbücher*. Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1973. First revised edition published in 1967.

[Kamlah and Lorenzen, 1984] Wilhelm Kamlah and Paul Lorenzen. *Logical Propaedeutic: Pre-School of Reasonable Discourse*. University Press of America, Lanham, MD & London, c1984. Translation of *Logische Propädeutik* by Hoke Robinson.

[Kindt, 1970] Walther Kindt. *Dialogspiele*. Diplomarbeit, Albert-Ludwigs-Universität, Freiburg im Breisgau, 1970.

[Kindt, 1972] Walther Kindt. *Eine abstrakte Theorie von Dialogspielen*. PhD Dissertation, Albert-Ludwigs-Universität, Freiburg im Breisgau, 1972.

[Kindt, 1980] Walther Kindt. Theorie der Dialogspiele, die Einführung des Wahrheitsprädikats und die Logik von Sprachen mit unfundierten Sätzen. In Thomas Ballmer and Walther Kindt, editors, *Zum Thema Sprache und Logik: Ergebnisse einer interdisziplinären Diskussion*, pages 95–145. Helmut Buske Verlag, Hamburg, 1980.

[Krabbe, 1978] Erik C.W. Krabbe. The Adequacy of Material Dialogue-Games. *Notre Dame Journal of Formal Logic*, 19:321–330, 1978.

[Krabbe, 1982] Erik C.W. Krabbe. Formal Dialectics as Immanent Criticism of Philosophical Systems. In [Barth and Martens, 1982, pp. 233–243]. Also in [Krabbe, 1982a, pp. 70–79].

[Krabbe, 1982a] Erik C.W. Krabbe. *Studies in Dialogical Logic*. PhD Dissertation, Groningen University, Groningen, 1982.

[Krabbe, 1982b] Erik C.W. Krabbe. Essentials of the Dialogical Treatment of Quantifiers. In [Krabbe, 1982a, pp. 249–257].

[Krabbe, 1985] Erik C.W. Krabbe. Noncumulative Dialectical Models and Formal Dialectics. *Journal of Philosophical Logic*, 14:129–168, 1985.

[Krabbe, 1985a] Erik C.W. Krabbe. Formal Systems of Dialogue Rules. *Synthese*, 63:295–328, 1985.

[Krabbe, 1986] Erik C.W. Krabbe. A Theory of Modal Dialectics. *Journal of Philosophical Logic*, 15:191–217, 1986.

[Krabbe, 1988] Erik C.W. Krabbe. Dialogue Sequents and Quick Proofs of Completeness. In Jakob Ph. Hoepelman, editor, *Representation and Reasoning: Proceedings of the Stuttgart Conference Workshop on Discourse Representation, Dialogue Tableaux Theory and Logic Programming*, pages 135–140 & 174. Max Niemeyer, Tübingen, 1988.

[Krabbe, 1988a] Erik C.W. Krabbe. Creative Reasoning in Formal Discussion. *Argumentation* 2:483–498, 1988.

[Krabbe, 1991] Erik C.W. Krabbe. Logica en spel [Logic and Games]. *Wijsgerig perspectief op maatschappij en wetenschap*, 31:102–111, 1990–1991.

[Krabbe, 1997] Erik C.W. Krabbe. Lorenzen Dialogue Games: An Approach to Interpersonal Reasoning: Extended Abstract. In Paul Dekker, Martin Stokhof, and Yde Venema, editors, *Proceedings of the Eleventh Amsterdam Colloquium: December 17–20, 1997*, pages 19–24. Institute for Logic, Language, and Computation and the Department of Philosophy, University of Amsterdam, Amsterdam, 1997.

[Krabbe, 1997a] Erik C.W. Krabbe. Arguments, Proofs, and Dialogues. In [Astroh, *et al.*, 1997, pp. 63–75].

[Krabbe, 1998] Erik C.W. Krabbe. Dialogical Logic. In Edward Graig, editor, *Routledge Encyclopedia of Philosophy*, vol. 3, pages 59–62. Routledge, London & New York, 1998.

[Krabbe, *et al.*, 1993] Erik C.W. Krabbe, Renée José Dalitz, and Pier A. Smit, editors. *Empirical Logic and Public Debate: Essays in Honour of Else M. Barth*. Rodopi, Amsterdam & Atlanta, GA, 1993.

[Lorenz, 1961] Kuno Lorenz. *Arithmetik und Logik als Spiele*. PhD Dissertation, Christian-Albrechts-Universität, Kiel, 1961. Selections reprinted in [Lorenzen and Lorenz, 1978, pp. 17–95].

[Lorenz, 1968] Kuno Lorenz. Dialogspiele als semantische Grundlage von Logikkalkülen. *Archiv für mathematische Logik und Grundlagenforschung*, 11:32–55 & 73–100, 1968. Reprinted in [Lorenzen and Lorenz, 1978, pp. 96–162].

[Lorenz, 1973] Kuno Lorenz. Die dialogische Rechtfertigung der effektiven Logik. In [Kambartel and Mittelstraß, 1973, pp. 250–280]. Reprinted in [Lorenzen and Lorenz, 1978, pp. 179–209].

[Lorenz, 1973a] Kuno Lorenz. Rules Versus Theorems: A New Approach for Mediation Between Intuitionistic and Two-Valued Logic. *Journal of Philosophical Logic*, 2:352–369, 1973.

[Lorenz, 1979] Kuno Lorenz, editor. *Konstruktionen versus Positionen: Beiträge zur Diskussion um die konstruktive Wissenschaftstheorie, I: Spezielle Wissenschaftstheorie*. Walter de Gruyter, Berlin and New York, 1979.

[Lorenz, 1984] Kuno Lorenz. Logik, dialogische. In Jürgen Mittelstraß, editor, *Enzyklopädie Philosophie und Wissenschaftstheorie, II*, pages 643–646. Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1984.

[Lorenzen, 1955] Paul Lorenzen. *Einführung in die operative Logik und Mathematik*. Springer, Berlin, 1955.

[Lorenzen, 1960] Paul Lorenzen. Logik und Agon. In *Atti del XII Congresso Internazionale di Filosofia (Venezia, 12–18 Settembre 1958)*, Volume IV *Logica, linguaggio e comunicazione*, pages 187–194. . Sansoni, Firenze, 1960. Reprinted in [Lorenzen and Lorenz, 1978, pp. 1–8].

[Lorenzen, 1961] Paul Lorenzen. Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics, Warsaw, 2–9 September 1959*, pages 193–200. Pergamon Press–Panstwowe wydawnictwo naukowe, Oxford–Warsaw, 1961. Reprinted in [Lorenzen and Lorenz, 1978, pp. 9–16].

[Lorenzen, 1962] Paul Lorenzen. *Metamathematik*. Bibliographisches Institut, Mannheim, 1962.

[Lorenzen, 1969] Paul Lorenzen. *Normative Logic and Ethics*. Volume 236, *Hochschultaschenbücher*. Bibliographisches Institut, Mannheim, 1969. Second annotated edition published by Bibliographisches Institut, Wissenschaftsverlag, 1984.

[Lorenzen, 1972] Paul Lorenzen. Dialogkalküle. *Archiv für mathematische Logik und Grundlagenforschung*, 15:99–102, 1972.

[Lorenzen, 1972a,] Paul Lorenzen. Zur konstruktiven Deutung der semantischen Vollständigkeit klassischer Quantoren- und Modalkalküle. *Archiv für mathematische Logik und Grundlagenforschung*, 15:103–117, 1972.

[Lorenzen, 1978] Paul Lorenzen. *Theorie der technischen und politischen Vernunft*. Philipp Reclam jun., Stuttgart, 1978.

[Lorenzen, 1980] Paul Lorenzen. Die dialogische Begründung von Logikkalkülen. In [Gethmann, 1980, pp. 43–69]. A nearly identical version was published in [Barth and Martens, 1982, pp. 23–54].

[Lorenzen, 1987] Paul Lorenzen. *Lehrbuch der konstruktiven Wissenschaftstheorie*. Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1987.

[Lorenzen and Lorenz, 1978] Paul Lorenzen and Kuno Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.

[Lorenzen and Schwemmer, 1973] Paul Lorenzen and Oswald Schwemmer. *Konstruktive Logik, Ethik und Wissenschaftstheorie*. First edition. Volume 700, *Hochschultaschenbücher*. Bibliographisches Institut, Mannheim, 1973.

[Lorenzen and Schwemmer, 1975] Paul Lorenzen and Oswald Schwemmer. *Konstruktive Logik Ethik und Wissenschaftstheorie*. Second improved edition. Volume 700, *Hochschultaschenbücher*. Bibliographisches Institut, Mannheim, 1975. First edition published in 1973.

[Machate, 1996] Joachim Machate. *Fokussierung im Dialog: Aspekte der Fokusintonation im Deutschen*. PhD Dissertation, Stuttgart University, Stuttgart, 1995. Volume 230, *Forschung und Praxis*. Springer, Berlin, 1996.

[Mackenzie, 1979] J.D. (Jim) Mackenzie. How to Stop Talking to Tortoises. *Notre Dame Journal of Formal Logic*, 20:705–717, 1979.

[Mackenzie, 1985] J.D. (Jim) Mackenzie. No Logic Before Friday. *Synthese*, 63:329–341, 1985.

[Mackenzie, 1990] J.D. (Jim) Mackenzie. Four Dialogue Systems. *Studia Logica*, 49:567–583, 1990.

[Mayer, 1981] Gregor Mayer. *Die Logik im Deutschen Konstruktivismus: Die Rolle formaler Systeme im Wissenschaftsaufbau der Erlanger und Konstanzer Schule*. PhD Dissertation, Ludwig-Maximilians-Universität, München, 1981.

[Moraux, 1968] Paul Moraux. La joute dialectique d'après le huitième livre des Topiques. In Gwilym E.L. Owen, editor, *Aristotle on Dialectic: The Topics: Proceedings of the Third Symposium Aristotelicum*, pages 277–311. Clarendon, Oxford, 1968.

[Plato, 1997] Plato. Euthydemus. (Translated by Rosamond Kent Sprague.) In Plato, *Complete Works*, edited by John M. Cooper and D.S. Hutchinson, pages 708–745. Hackett Publishing Company, Indianapolis, IN, and Cambridge, MA, 1997.

[Rahman and Rückert, 1999] Shahid Rahman and Helge Rückert. Dialogische Modallogik (für T, B, S4 und S5). *Logique et Analyse*, 42(167–168):243–282, 1999.

[Rahman and Rückert, 2001a] Shahid Rahman and Helge Rückert, editors. *New Perspectives in Dialogical Logic. Synthese*, 127, 2001.

[Rahman and Rückert, 2001b] Shahid Rahman and Helge Rückert. Dialogical Connexive Logic. In [Rahman and Rückert, 2001a, pp. 105–139].

[Reed and Norman, 2003] Chris Reed and Timothy J. Norman, editors. *Argumentation Machines: New Frontiers in Argument and Computation*. Kluwer, Dordrecht, 2003.

[Rescher, 1977] Nicholas Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, Albany, NY, 1977.

[Saarinen, 1979] Esa Saarinen, editor. *Game-Theoretical Semantics: Essays on Semantics by Hintikka, Carlson, Peacocke, Rantala, and Saarinen*. Reidel, Dordrecht, 1979.

[Slomkowski, 1997] Paul Slomkowski. *Aristotle's Topics*. Brill, Leiden, 1997.

[Starmans, 1996] Richard J.C.M. Starmans. *Logic, Argument, and Common Sense*. PhD Dissertation, Catholic University of Brabant, Tilburg, 1996.

[Stegmüller, 1964] Wolfgang Stegmüller. Remarks on the Completeness of Logical Systems Relative to the Validity-Concepts of P. Lorenzen and K. Lorenz. *Notre Dame Journal of Formal Logic*, 5:81–112, 1964.

[Stegmüller and Varga von Kibéd, 1984] Wofgang Stegmüller and Matthias Varga von Kibéd. *Strukturtypen der Logik*. Berlin, Springer, 1984. *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie III*.

[Strobel, 1997] Maria Strobel. *Flexibilisierung und Beschränkung von Benutzeräusserungen innerhalb von telefonischen Mensch-Maschine Dialogen*. IBM Deutschland, Scientific Centre, Institute for Logic and Linguistics, Heidelberg, 1997.

[Thiel, 1978] Christian Thiel. Duality Lost? Transforming Gentzen Derivations into Winning-Strategies for Dialogue Games. *Theorema* 8:57–66, 1978. University of Valencia.

[Valerius, 1990] Rolf D.H. Valerius. *Die Logik von Rahmen- und Stopregeln in Lorenzen-Spiele*. PhD Dissertation, Stuttgart University, Stuttgart, 1990.

[Van Dun, 1972] Frank van Dun. On the Modes of Opposition in the Formal Dialogues of P. Lorenzen. *Logique et Analyse*, n.s. 15e Année:103–136, 1972.

[Van Eemeren and Grootendorst, 1982] Frans H. van Eemeren and Rob Grootendorst. *Regels voor redelijke discussies: Een bijdrage tot de theoretische analyse van argumentatie ter oplossing van geschillen* [Rules for Reasonable Discussions: A Contribution to the Theoretical Analysis of Argumentation Aiming at Conflict Resolution]. PhD Dissertation, University of Amsterdam, Amsterdam, 1982.

[Van Eemeren and Grootendorst, 1984] Frans H. van Eemeren and Rob Grootendorst. *Speech Acts in Argumentative Discussions: A Theoretical Model for the Analysis of Discussions Directed Towards Solving Conflicts of Opinion*. Dordrecht and Cinnaminson, NJ: Foris, 1984.

[Van Eemeren and Grootendorst, 1992] Frans H. van Eemeren and Rob Grootendorst. *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective*. Hillsdale NJ, etc.: Lawrence Erlbaum, 1992.

[Van Eemeren and Grootendorst, 2004] Frans H. van Eemeren and Rob Grootendorst. *A Systematic Theory of Argumentation*. Cambridge: Cambridge University Press, 2004.

[Van Eemeren, *et al.*, 1996] Frans H. van Eemeren, Rob Grootendorst, Francisca Snoeck Henkemans *et al. Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Mahwah, NJ: Lawrence Erlbaum, 1996.

[Van Hoof, 1995] Antonius J.M. (Toine) van Hoof. *Der logisch-pragmatischen Gebrauch von Konditionalsätzen: Eine dialog-logische Analyse*. PhD Dissertation, Stuttgart University, Stuttgart, 1994. Volume 216 *Forschung und Praxis*, Springer, Berlin, 1995.

[Veltman, 1996] Frank Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[Vreeswijk, 1993] Gerard A.W. Vreeswijk. *Studies in Defeasible Argumentation*. PhD Dissertation, Vrije Universiteit, Amsterdam, 1993.

[Walton, 1984] Douglas N. Walton. *Logical Dialogue-Games and Fallacies*. Lanham, New York and London, 1984.

[Walton, 1998] Douglas N. Walton. *The New Dialectic: Conversational Contexts of Argument*. University of Toronto Press, Toronto, 1998.

[Walton and Krabbe, 1995] Douglas N. Walton and Erik C.W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany NY, 1995.

[Wittgenstein, 1953] Ludwig J.J. Wittgenstein. *Philosophische Untersuchungen / Philosophical Investigations*. Routledge and Kegan Paul, London, 1953. Edited by G.E.M. Anscombe and R. Rhees.

[Woods and Walton, 1989] John Woods and Douglas N. Walton. *Fallacies: Selected Papers 1972–1982*. Dordrecht & Providence, RI: Foris, 1989.

INDEX

This Page is Intentionally Left Blank