Ondrej Majer
Ahti-Veikko Pietarinen
Tero Tulenheimo
Editors

LOGIC, EPISTEMOLOGY, A

Games: Uni
Logic, Lang

LOGIC, EPISTEMOLOGY, AND THE UNITY OF SCIENCE

VOLUME 15

Logic, Epistemology, and the Unity of Science aims to reconsider the question of the unity of science in light of recent developments in logic. At present, no single logical, semantical or methodological framework dominates the philosophy of science. However, the editors of this series believe that formal techniques like, for example, independence friendly logic, dialogical logics, multimodal logics, game theoretic semantics and linear logics, have the potential to cast new light no basic issues in the discussion of the unity of science.

This series provides a venue where philosophers and logicians can apply specific technical insights to fundamental philosophical problems. While the series is open to a wide variety of perspectives, including the study and analysis of argumentation and the critical discussion of the relationship between logic and the philosophy of science, the aim is to provide an integrated picture of the scientific enterprise in all its diversity.

For other titles published in this series, go to
www.springer.com/series/6936

Games: Unifying Logic, Language, and Philosophy

Edited by

Ondrej Majer
Czech Academy of Sciences, Czech Republic

Ahti-Veikko Pietarinen
University of Helsinki, Finland

Tero Tulenheimo
University of Helsinki, Finland

Springer

Editors
Dr. Ondrej Majer
Academy of Sciences of the
Czech Republic
Institute of Philosophy
Jilska 1
110 00 Prague 1
Czech Republic
majer@site.cas.cz

Dr. Tero Tulenheimo
University of Helsinki
Department of Philosophy
Siltavuorenpenger 20 A
FI-00014 Helsinki
P.O. Box 9
Finland
tero.tulenheimo@helsinki.fi

Dr. Ahti-Veikko Pietarinen
University of Helsinki
Department of Philosophy
Siltavuorenpenger 20 A
FI-00014 Helsinki
P.O. Box 9
Finland
ahti-veikko.pietarinen@helsinki.fi

Cover image: Adaptation of a Persian astrolabe (brass, 1712–13), from the collection of the Museum of the History of Science, Oxford. Reproduced by permission.

ISBN 978-1-4020-9373-9 e-ISBN 978-1-4020-9374-6

Library of Congress Control Number: 2008938971

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Contributing Authors

Boudewijn de Bruin
Faculty of Philosophy
University of Groningen
Oude Boteringestraat 52
9712 GL Groningen
The Netherlands
b.p.de.Bruin@philos.rug.nl

Petr Cintula
Institute of Computer Science
Academy of Sciences
of the Czech Republic
Pod Vodarenskou vezi 2
182 07 Prague 8
Czech Republic
cintula@cs.cas.cz

Robin Clark
Department of Linguistics
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19104-6305
USA
rclark@babel.ling.upenn.edu

Christian G. Fermüller
Technische Universität Wien
Favoritenstr. 9-11/E1852
A-1040 Wien
Austria
chrisf@logic.at

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

Hykel Hosni
Scuola Normale Superiore
Piazza dei Cavalieri
56100 Pisa
Italy
hykel.hosni@sns.it

Giorgi Japaridze
Department of Computing Sciences
Villanova University
800 Lancaster Avenue
Villanova, Pennsylvania 19085
USA
giorgi.japaridze@villanova.edu

Mathieu Marion
Département de philosophie
Université du Québec à Montréal
Case postale 8888, succursale
Centre-ville
Montréal, Québec
Canada H3C 3P8
marion.mathieu@uqam.ca

Merlijn Sevenster
Philips Research
Prof. Holstlaan 4
5656 AA Eindhoven
The Netherlands
merlijn.sevenster@philips.com

Shahid Rahman
U.F.R. de Philosophie
Domaine Universitaire "Pont de Bois"
Université Lille III
59653 Villeneuve d'Ascq
France
shahid.rahman@univ-lille3.fr

Manuel Rebuschi
Laboratoire d'Histoire des Sciences et de Philosophie
Archives Henri Poincaré
UMR 7117 CNRS - Nancy-Université
Université Nancy 2
23, Bd. Albert Ier-BP 3397. F-54015 NANCY Cedex
France
Manuel.Rebuschi@univ-nancy2.fr

Wim Veldman
Institute for Mathematics, Astrophysics and Particle Physics
Radboud University Nijmegen
P.O. Box 9010
6500 GL Nijmegen
The Netherlands
W.Veldman@math.ru.nl

John Woods
Department of Philosophy
University of British Columbia
1866 Main Mall E370
Vancouver, BC
Canada V6T 1Z1
jhwoods@interchange.ubc.ca
woods@dcs.kcl.ac.uk

Ondrej Majer
Institute of Philosophy
Academy of Sciences of the Czech Republic
Jilská 1, 110 00 Prague
Czech Republic
majer@site.cas.cz

Ahti-Veikko Pietarinen
Department of Philosophy
P.O. Box 9
00014 University of Helsinki
Finland
ahti-veikko.pietarinen@helsinki.fi

Tero Tulenheimo
Department of Philosophy
Academy of Finland
P.O. Box 9
00014 University of Helsinki
Finland
tero.tulenheimo@helsinki.fi

Introduction

Ondrej Majer, Ahti-Veikko Pietarinen, and Tero Tulenheimo

1 Games and logic in philosophy

Recent years have witnessed a growing interest in the unifying methodologies over what have been perceived as pretty disparate logical 'systems', or else merely an assortment of formal and mathematical 'approaches' to philosophical inquiry. This development has largely been fueled by an increasing dissatisfaction to what has earlier been taken to be a straightforward outcome of 'logical pluralism' or 'methodological diversity'. These phrases appear to reflect the everyday chaos of our academic pursuits rather than any genuine attempt to clarify the general principles underlying the miscellaneous ways in which logic appears to us.

But the situation is changing. Unity among plurality is emerging in contemporary studies in logical philosophy and neighbouring disciplines. This is a necessary follow-up to the intensive research into the intricacies of logical systems and methodologies performed over the recent years.

The present book suggests one such peculiar but very unrestrained methodological perspective over the field of logic and its applications in mathematics, language or computation: games. An allegory for opposition, cooperation and coordination, games are also concrete objects of formal study.

As a metaphor for argumentation Aristotle's *Topics* and its reincarnations such as the scholastic *Ars Obligatoria* are set up as dialogical duels (Pietarinen, 2003a). Logics exploiting this idea resurface in the twentieth century attempts to clarify the concepts of argument and proof. The game metaphor has retained its strength in contemporary theories of computation (Pietarinen, 2003b, Japaridze, this volume), in which computation is recast in terms of the symbiosis between the Computing System ('Myself') and its Environment ('Nature'). In mathematics, the benefits of doing so were noted decades ago by Stanislaw Ulam (1960), who wrote how amusing it is "to consider how one can 'gamize'

various mathematical situations (or perhaps the verb should be 'paizise' from the Greek word $\pi\alpha\iota\varsigma\iota\eta$, to play)."

Games as explications of the core philosophical questions concerning the scientific methodologies were on the brink of being born in the writings of the early unificators, including Rudolf Carnap, Otto Neurath, Charles Morris and Carl Gustav Hempel. But they never operationalised the key notions. The term 'operationalisation' is apt, since what was attempted was to give meaning to 'operationalisation'. According to operationalism, a concept is synonymous with the set of operations correlated with it. Influenced by Percy Bridgman's and Alfred Einstein's thoughts, the early workers on what was later to become the Unity of Science Movement inherited the better parts of the Viennese verificationism in the methodology of science which, in turn, was allied to, though also significantly different from, the pragmaticism of Charles Peirce. Moreover, Pietarinen and Snellman (2006) show that the kernel of pragmaticism is, in turn, essentially game-theoretical in nature.

Accordingly, a sustained attempt has existed in the history and philosophy of science to articulate the interactive, the strategic and the pragmatic in logic. The chief reason for the failure of the early philosophers working on uniting the foundations of scientific methodology was their stout belief in the explanatory capacities of singular behaviour. In game theory, in contrast, the success lies in the possibility of there being general, or strategic, habits of acting in a certain way whenever certain kinds of situations are confronted.

How coincidental it must have been that many of the logicians working on the operative definitions of logical concepts, including Hugo Dingler and Paul Lorenzen, were not only champions of the Husserlian notion of *Spielbedeutungen* (Pietarinen, 2008), but also immersed in the continental branch of operationalism, which in various forms had already been in vogue around the exiting new projects emerging in the philosophy of science since the 1920s. Meanwhile, game theory proper was in the making, first in the urban atmospheres of the continental triangle of Berlin, Vienna and Göttingen, and later on in the singular intellectual concentrate of the ludic post-war Princeton Campus.

But these historical events constitute just the beginnings of the story, the impact of which is only beginning to unravel. The present book itself constitutes only a modest fragment of that narrative. The book consists of 12 chapters divided into four parts: *Philosophical Issues* (Part I), *Game-Theoretic Semantics* (Part II), *Dialogues* (Part III), and *Computation and Mathematics* (Part IV). The individual topics covered include, in Part I, the philosophy of logical games (Chapter 1, Mathieu Marion), the epistemic characterisation results in game theory, scientific explanation and the philosophy of the social sciences (Chapter 2, Boudewijn de Bruin), rationality, strategic interaction, focal points, radical interpretation and the selection of multiple Nash-equilibria (Chapter 3, Hykel Hosni) and the notion of cognitive agency, cognitive economy

and fallacies (Chapter 4, John Woods and Dov M. Gabbay). In Part II, the central methodology is that of game-theoretic semantics, where the germane topics are independence-friendly (IF) logic, imperfect-information games and weak dominance (Chapter 5, Merlijn Sevenster), fuzzy logic (Chapter 6, Petr Cintula and Ondrej Majer) and generalised quantifiers and natural-language semantics (Chapter 7, Robin Clark). Part III is devoted to the method of dialogues, and it deals with the relationships between the game-theoretic and dialogic notions of truth and validity (Chapter 8, Shahid Rahman and Tero Tulenheimo), fuzzy logic, vagueness, supervaluation and betting (Chapter 9, Christian G. Fermüller) and epistemic and intuitionistic logic (Chapter 10, Manuel Rebuschi). Part IV is on the application and use of games in computation and mathematics. Topics covered have to do with computability logic, game semantics and affine linear logic (Chapter 11, Giorgi Japaridze) and determinacy, infinite games and intuitionism in mathematics (Chapter 12, Wim Veldman).

As is evident from this impressive list of topics, the method of games is so widespread across studies in logic and the neighbouring disciplines—including applications to linguistic semantics and pragmatics, the social sciences, philosophy of science, epistemology, economics, mathematics and computation—that it prompts us to take seriously the possibility that there is some "greater conceptual rationale of what it is to be a *bona fide* science" (Margolis, 1987, p. xv). Games, as applied to logic, philosophy, epistemology, linguistics, cognition, computation or mathematics, provide at the same time a notably modern, rigorous and creative formal toolkit that lays bare the structures of logical and cognitive processes—be they proofs, dialogues, inferences, models, arguments, negotiations, bargaining, or computations—while being the product of an age-old enquiring mind and human rational action.

To what extent such methods and tools are able ultimately to reconcile the human and natural sciences (Margolis, 1987) remains to be seen. After all, the first steps in any expansion over multiple disciplines must begin from the beginning; in logic, it would begin from charting what the foundational perspectives are that logic provides to those fields of intellectual pursuit amenable to fruitful formalisations. But we believe that the existence of methods inescapably linked with the ways in which human rational thought processes and actions function supports the wider scenario.

Whether the unity holds in those nooks and corners of scientific and intellectual pursuits covered in the present essays we leave for the readers to judge—it is a question of not only method of logic but also ontology, history of ideas, scientific practices, and, ultimately, of the fruits that the applications of games to the multiplicity of intellectual tasks are capable of bearing.

In the remainder of this introduction, we outline the essentials of two major approaches to how games have been used to explicate logical notions: game-theoretical semantics and dialogical logic.

2 Game-Theoretical Semantics

Hintikka (1968) introduced Game-Theoretical Semantics (GTS) for first-order logic. From the very beginning, the idea was driven by philosophical considerations. Hintikka's goal was not merely to provide an alternative characterisation of truth for first-order logic, but to lay down a theory of meaning making use of—and sharpening—Wittgenstein's idea of 'language game', relating these considerations to Kantian thought and to the idea that logic has to do with synthetic activity (Hintikka, 1973).

Hintikka extended the game-theoretic interpretation that Henkin (1961) had in effect provided to quantified sentences in prenex normal form; this interpretation will be discussed further below. He explained how a semantic game is played with an arbitrary first-order sentence as input.[1] He observed that conjunctions and disjunctions can be treated on a par with universal and existential quantifier, respectively. After all, $(\phi \wedge \chi)$ holds if and only if all of the sentences ϕ, χ hold, and $(\phi \vee \chi)$ holds if and only if at least one of the sentences ϕ, χ holds. Accordingly, a game for $(\phi \wedge \chi)$ proceeds by the "universal" player picking out one of the conjuncts $\theta \in \{\phi, \chi\}$, after which the play is continued with respect to the sentence θ. Similarly, in connection with a game for $(\phi \vee \chi)$, it is the "existential player" who makes a choice of a disjunct $\theta \in \{\phi, \chi\}$. (The objects chosen are syntactic items in connection with conjunction and disjunction, whereas the moves for quantifiers involve choosing objects out there in the domain.)

What about negation, then? Hintikka observed that negation has the effect of changing the roles of the players. After any sequence of moves that the players have made while playing a game, one of the players has the role of 'Verifier' and the other that of 'Falsifier'. Now a game corresponding to $\neg\phi$ continues with respect to ϕ, with the players' roles reversed: the player having the role of 'Verifier' relative to $\neg\phi$ assumes relative to ϕ the role of 'Falsifier', and vice versa.

GTS provides a game-theoretic counterpart to the model-theoretic notion of truth. In this way, the notions of truth for a great variety of logics can be provided. Cases in point are propositional logic, first-order logic, modal and temporal logics, independence-friendly logics (Hintikka, 1995, 1996; Sandu, 1993; Hintikka and Sandu, 1989, 1997), logics with Henkin quantifiers (Henkin, 1961; Krynicki and Mostowski, 1995), infinitely deep languages (Hintikka and Rantala, 1976; Karttunen, 1984; Hyttinen, 1990) and the logic of Vaught sentences (Vaught, 1973; Makkai, 1977).

Semantic games are two-player games; we may call the two players Eloise or the 'initial Verifier' and Abelard or the 'initial Falsifier'. The truth of a sentence φ in a model $\mathcal{M}$ corresponds to the existence of a winning strategy for

[1] The game interpretation goes back to Charles Peirce's investigation in the algebra of logic and graphical logic (Hilpinen, 1982; Pietarinen, 2006b).

Eloise in the semantic game $G(\varphi, \mathcal{M})$ correlated with φ and played on $\mathcal{M}$. The falsity of φ corresponds to the existence of a winning strategy for Abelard. Intuitively, Eloise can be thought of as defending the claim "φ is true in $\mathcal{M}$" against any attempts of Abelard to refute this claim. Similarly, Abelard defends the claim "φ is false in $\mathcal{M}$" against any attempted refutations of this claim by Eloise. The games $G(\varphi, \mathcal{M})$ are so defined that φ is indeed true (false) in $\mathcal{M}$ iff there exists a method for Eloise (Abelard) to win against all sequences of moves by Abelard (Eloise).

The mathematical reality behind semantic games may be less picturesque than the above description in terms of defences against refutations suggests. Given a semantic game $G(\varphi, \mathcal{M})$, the existence or non-existence of a winning strategy for either player is an objective fact about the model $\mathcal{M}$. Whether the players' actions bear relevance to the truth or falsity of the sentence is thus arguable.[2]

The roots of semantic games go back to the Tarskian definition of truth. According to Tarski, to test whether a sentence such as $\forall x \exists y P(x, y)$ is true in a model $\mathcal{M}$, reference to objects a and b of the domain M of $\mathcal{M}$ is needed. The sentence is true iff it is the case that for any a there is an object b such that $P(a, b)$ holds. Thus understood, the truth of the sentence $\forall x \exists y P(x, y)$ does not require the existence of a function $f\colon M \to M$ such that $b = f(a)$ for any $a \in M$. It only requires the existence of a relation $R \subseteq M \times M$ such that for every a there is at least one b with $R(a, b)$ such that $P(a, b)$ holds in $\mathcal{M}$. To get from the statement involving relations to the statement concerning functions, the Axiom of Choice is, in general, needed (Hodges, 1997a). On the other hand, assuming the Axiom of Choice, the truth-condition of $\forall x \exists y P(x, y)$ can indeed be stated as the requirement that there be a function f such that for any value a interpreting $\forall x$, the function produces a witness $b = f(a)$ for $\exists y$. Such functions, introduced by Skolem (1920), are known as Skolem functions.

Henkin (1961) considered logical systems in which infinitely long formulas with infinitely many quantifier alternations are allowed; one of the examples he mentions is the formula

$$\exists x_1 \forall x_2 \exists x_3 \forall x_4 \ldots P(x_1, x_2, \ldots). \tag{1}$$

In connection with such formulas, Henkin suggested that the procedure of picking up objects corresponds to moves in a game between two players, which we might for simplicity call the universal player (Abelard) and the existential

[2]Hodges (2006a, b; Hodges and Krabbe, 2001) has levelled critique on the idea that logical games shed new light on the semantics of quantifiers, or that logical games could actually have conceptually important roles to play in justifying certain logical procedures or in defining meanings. But see the rejoinders in Pietarinen (2006b, Chapter 9) and Hodges and Krabbe (2001) and Marion, this volume, as well as earlier discussion in Hand (1989).

player (Eloise). The former is responsible for choosing objects corresponding to universally quantified variables while the latter similarly interprets existentially quantified variables.

Admittedly, Henkin used the notion of game quite metaphorically. But he pointed out that logical games are related to Skolem functions and observed that winning strategies for the existential player are sequences of Skolem functions. For instance, when evaluating the above formula (1) relative to a model $\mathcal{M}$, any sequence $\langle f_1, f_3, f_5, \ldots \rangle$ of Skolem functions, one for each existential quantifier $\exists x_{2n+1}$ in (1), gives a winning strategy for the existential player in the game correlated with the formula (1) in the model $\mathcal{M}$. In other words, the formula (1) is true in $\mathcal{M}$ if and only if the following second-order formula is true in $\mathcal{M}$[3]:

$$\exists f_1 \exists f_3 \exists f_5 \ldots \forall x_2 \forall x_4 \forall x_6 \ldots P(f_1, x_2, f_3(x_2), x_4, f_5(x_2, x_4), x_6, \ldots). \quad (2)$$

Let us give a precise definition of semantic games for first-order logic. First we agree on some terminology. If τ is a vocabulary, ψ is a first-order τ-formula and c is an individual constant (not necessarily from the vocabulary τ), then $\psi[x/c]$ will stand for the $(\tau \cup \{c\})$-formula that results from substituting c for all free occurrences of the variable x in ψ. Whenever $\mathcal{M}$ is a τ-structure (model), by convention M will stand for the domain of $\mathcal{M}$. If $\mathcal{M}$ is a τ-structure, $\mathcal{M}'$ is a τ'-structure, and $\tau \subset \tau'$, then $\mathcal{M}'$ is an expansion of $\mathcal{M}$, provided that $M = M'$ and $\mathcal{M}'$ agrees with $\mathcal{M}$ on the interpretations of the symbols from τ.

With every vocabulary τ, τ-structure $\mathcal{M}$ and first-order τ-sentence φ, a two-player, zero-sum game $G(\varphi, \mathcal{M})$ of perfect information is associated. The games are played with the following rules.

- If $\varphi = R(a_1, \ldots, a_n)$, the play has come to an end. If $(a_1^{\mathcal{M}}, \ldots, a_n^{\mathcal{M}}) \in R^{\mathcal{M}}$, the player whose role is 'Verifier' wins, and the one whose role is 'Falsifier' loses. On the other hand, if $(a_1^{\mathcal{M}}, \ldots, a_n^{\mathcal{M}}) \notin R^{\mathcal{M}}$, then 'Falsifier' wins and 'Verifier' loses.

- If $\varphi = (\psi \vee \chi)$, then 'Verifier' chooses a disjunct $\theta \in \{\psi, \chi\}$, and the play continues as $G(\theta, \mathcal{M})$.

- $\varphi = (\psi \wedge \chi)$, then 'Falsifier' chooses a conjunct $\theta \in \{\psi, \chi\}$, and the play continues as $G(\theta, \mathcal{M})$.

[3] In order for the second-order sentence (2) to be equivalent to the sentence (1), the standard interpretation of second-order logic in the sense of Henkin (1950) must be applied (the other requisite assumption being the Axiom of Choice). In particular, n-ary function variables are taken to range over arbitrary n-ary functions on the domain. Note that in (2) a Skolem function f_{2n+1} for the quantifier $\exists x_{2n+1}$ is a function of type $M^n \to M$. Hence a Skolem function for $\exists x_1$ is a zero-place function, that is, a constant.

- If $\varphi = \exists x\psi$, then 'Verifier' chooses an element $b \in M$, gives it a name, say n_b, and the play goes on as $G(\psi[x/n_b], \mathcal{N})$, where $\mathcal{N}$ is the $(\tau \cup \{n_b\})$-structure expanding $\mathcal{M}$ and satisfying $n_b^{\mathcal{N}} = b$.
- If $\varphi = \forall x\psi$, then 'Falsifier' chooses an element $b \in M$, gives it a name, say n_b, and the play goes on as $G(\psi[x/n_b], \mathcal{N})$, where $\mathcal{N}$ is the $(\tau \cup \{n_b\})$-structure expanding $\mathcal{M}$ and satisfying $n_b^{\mathcal{N}} = b$.
- If $\varphi = \neg\psi$, then the play continues as $G(\psi, \mathcal{M})$, with the players' roles switched: the 'Verifier' of game $G(\neg\psi, \mathcal{M})$ is the 'Falsifier' of game $G(\psi, \mathcal{M})$, and vice versa.

In applying the above game rules, any play of $G(\varphi, \mathcal{M})$ reaches an atomic sentence and hence comes to an end after finitely many moves. These rules follow Hintikka's original definition (Hintikka, 1968); in particular, whenever $G(\varphi, \mathcal{M})$ is a game, φ is a sentence—formula with no free occurrences of variables. However, no conceptual difficulties are involved in generalising the definition so as to apply to first-order formulas with any number of free variables. This is accomplished by providing variable assignments γ as an extra input when specifying games. Accordingly, for every τ-formula φ, τ-structure $\mathcal{M}$, and assignment γ mapping free variables of φ to the domain M, a game $G(\varphi, \mathcal{M}, \gamma)$ can be introduced. The game rules for quantifiers become simpler when phrased in terms of variable assignments. If for instance $\varphi = \exists x\psi$, then game $G(\varphi, \mathcal{M}, \gamma)$ proceeds by 'Verifier' choosing an element $b \in M$, whereafter the play continues as $G(\psi, \mathcal{M}, \gamma')$, where γ' is otherwise like γ but maps x to b. Unlike in the games defined for sentences, now the vocabulary considered is not extended by a name for the element b, and the model $\mathcal{M}$ is not expanded.

To make proper use of games for semantic purposes, having laid down a set of game rules is not enough. We also need the notion of strategy. To this end, some auxiliary notions must be defined. A history (or, partial play) of game $G(\varphi, \mathcal{M})$ is any sequence of moves, made in accordance with the game rules. A terminal history (or, play) is a history at which it is neither player's turn to move. The set of non-terminal histories can be partitioned into two classes $\mathcal{P}_\exists$ and $\mathcal{P}_\forall$: those at which it is Eloise's turn to move and those at which it is Abelard's turn to move.

Write $O_\exists$ for the set of those tokens of logical operators in φ for which it is Eloise's turn to move in $G(\varphi, \mathcal{M})$, namely for all existential quantifiers and disjunction signs with positive polarity, and for all universal quantifiers and conjunction signs with negative polarity.[4] Likewise, write $O_\forall$ for the set of the tokens of operators for which it is Abelard's turn to move. Then the histories in

[4] A logical operator has a positive polarity in a formula φ, if it appears in φ subordinate to n negation signs with $n \in \{2m : m \in \mathbb{N}\}$; otherwise it has a negative polarity.

the set $\mathcal{P}_\exists$ can be further partitioned according to the logical operator to which they correspond: for each $O \in \mathcal{O}_\exists$ there is a subset $\mathcal{P}_\exists^O$ of $\mathcal{P}_\exists$ of those histories at which Eloise must make a move to interpret O. The set $\mathcal{P}_\forall$ is similarly partitioned by $\mathcal{P}_\forall^O$ with $O \in \mathcal{O}_\forall$.

For each $O \in \mathcal{O}_\exists$, Eloise's strategy function is a function that provides a move for her at each history belonging to $\mathcal{P}_\exists^O$. It is commonplace to stipulate that at a history $h \in \mathcal{P}_\exists^O$, Eloise's strategy function for O takes as its arguments Abelard's moves made in h.[5] A strategy for Eloise is a set of her strategy functions, one function for each operator in $\mathcal{O}_\exists$. A strategy for Eloise is winning, if it leads to a play won by Eloise against any sequence of moves by Abelard. The notions of strategy function, strategy, and winning strategy are similarly defined for Abelard.

Assuming the Axiom of Choice, it can then be shown that a first-order sentence φ is true (false) in a model $\mathcal{M}$ in the usual Tarskian sense if and only if there exists a winning strategy for Eloise (Abelard) in game $G(\varphi, \mathcal{M})$, (see Hodges, 1983; Hintikka and Kulas, 1985).[6]

The fact that any formula φ is either true or false in any given model $\mathcal{M}$ manifests on the level of games in that all semantic games for first-order logic are determined: in any game $G(\varphi, \mathcal{M})$, either Eloise or Abelard has a winning strategy. Semantic games are zero-sum, two-player games of perfect information with finite horizon. The fact that they are determined follows from the Gale-Stewart theorem (Gale and Stewart, 1953).

The framework of semantic games makes it possible to pursue research at the interface of game theory and logic. Once a parallel between logical and game-theoretic notions has been successfully drawn—as it has, for instance, in connection with the notion of truth-in-a-model for first-order logic and the game-theoretic notion of the existence of a winning strategy for Eloise in a semantic game—one can meaningfully bring in further game-theoretic notions and go on studying the resulting logical systems.

One such avenue is opened up by subjecting games to imperfect information. The goal is then to study the 'information flow' in logical formulas, or the various relations of dependence and independence between logical constants. This type of research has led to the investigation of a family of independence-friendly logics (IF logics), studied in various publications by Jaakko Hintikka, Gabriel Sandu and many others (Hintikka, 1995, 1996; Hintikka and Sandu, 1989, 1997; Hodges 1997a, b; Pietarinen, 2001b, 2006a; Sandu, 1993; Väänänen, 2007).

[5] Normally, allowing Eloise's own moves as arguments of her strategy functions would not make it any easier for Eloise to have a winning strategy.

[6] The Axiom of Choice could be avoided when formulating the relation of the game-theoretic truth-definition to the Tarskian truth-definition, if strategies in the above sense, namely deterministic strategies, were replaced by nondeterministic strategies (Hodges, 2006b; Väänänen, 2006).

The framework of semantic games with imperfect information has been applied to a host of variants of IF logic, including IF propositional logic (Pietarinen, 2001a; Sandu and Pietarinen, 2001, 2003; Sevenster, 2006a), IF modal logic (Bradfield, 2006; Bradfield and Fröschle, 2002; Hyttinen and Tulenheimo, 2005; Pietarinen, 2001c, 2003c, 2004; Tulenheimo, 2003; Tulenheimo and Sevenster, 2006; Sevenster, 2006b), IF fixpoint logic (Bradfield, 2004) and IF fuzzy logics (Cintula and Majer, this volume).

Another example of game-theoretic conceptualisations in connection with logic is furnished by systematically investigating how far the common ground between logic and game theory can be pushed (van Benthem, 2001). The paper of Sevenster (this volume) belongs to that tradition.

3 Dialogical logic

Dialogical logic (a.k.a. dialogic) offers a game-theoretic approach to the logical notions of validity and satisfiability. In so doing, it contributes to two of the four objectives mentioned by Erik C. W. Krabbe in his apology of the dialogical standpoint, "Dialogue Logic Restituted" (Hodges and Krabbe, 2001): the foundations of mathematics and the addition of a third approach to logic next to model theory and proof theory. The two further objectives are related to argumentation theory and systematic reconstruction of the language of science and politics. Let us concentrate here on dialogical logic seen from the logic-internal viewpoint.

Given a formula φ of, say, propositional logic, it is associated with a game $\mathcal{D}(\varphi)$ referred to as dialogue about φ. Such games are between two players, called the Proponent and the Opponent. Games are so defined that a formula φ of classical propositional logic is valid under the usual criteria (that is, true under all valuations) iff there is a winning strategy for the Proponent in the dialogue about φ. The framework is flexible—a game-theoretic characterisation is obtained similarly, for instance, for validity in first-order logic and in various modal logics. It has also been applied to paraconsistent, connexive and free logics (Rahman et al., 1997; Rahman and Rückert, 2001; Rahman and Keiff, 2005). What is more, the contrast between classical and intuitionistic logic has a clear-cut characterisation in terms of dialogues. Indeed, Paul Lorenzen's characterisation of validity in intuitionistic propositional logic in his 1959 talk "Ein dialogisches Konstruktivitätskriterium" (Lorenzen, 1961) in terms of dialogues was of crucial importance to the very birth of dialogical logic. With hindsight, we may observe that, given rules that define dialogues corresponding to intuitionistic propositional logic, there is a systematic liberalisation that can be effected with respect to these rules so as to yield classical propositional logic (Lorenz, 1968).

The rules of dialogues are divided into two groups—particle rules and structural rules. The former rules specify, for each logical operator (or 'logical particle'), how a formula having this operator as its outmost form can be criticised, and how such a critique can be answered. Structural rules, by contrast, lay down the ways in which the dialogues can be carried out—they specify, for instance, how the dialogue is begun, what types of attacks and defenses are allowed, and what counts, for a given player, as a win of a play of a dialogue. As it happens, dialogues for intuitionistic logic are obtained from those of classical logic by changing a single structural rule, while keeping the particle rules intact. (In classical dialogues, a player may defend himself or herself against *any* previously effected challenge, including those that the player has already defended at least once; while in intuitionistic dialogues, the player may only defend himself or herself against the most recent of those challenges that have not yet been defended.)

Dialogical logicians tend to see dialogues as a sui generis approach to logic, a third realm in addition to proof theory and model theory. Be that as it may, there is a clear sense in which dialogical logic is naturally coupled with proof theory, whereas game-theoretical semantics, in contrast, is coupled with the study of model-theoretic properties. Think of a logic $\mathcal{L}$ that admits, as a matter of fact, a sound and complete proof system, say classical propositional logic or classical first-order logic. Dialogues provide such a proof system for $\mathcal{L}$. A winning strategy of the Proponent in a dialogue about φ counts as a proof of φ. Crucially, dialogues for the logic $\mathcal{L}$ serve to recursively enumerate the set of valid formulas of $\mathcal{L}$. (Given a valid formula of $\mathcal{L}$, the Opponent's choices can only give rise to finitely many moves before a play is reached which is won by the Proponent and which cannot be further extended.) It is natural to consider systems of semantic tableaux (Hintikka, 1955; Beth, 1959; Smullyan, 1968; Fitting, 1969) as mediating the connection between proof theory and dialogues; there is an important, yet straightforward connection between tableaux on the one hand, and the totality of plays of dialogues on the other (Rahman and Keiff, 2005). In particular, for a given refutable formula φ of, say, propositional logic, there is a one-one correspondence between open maximal branches of a tableau for the signed formula $F\varphi$ and winning strategies of the Opponent in the dialogue about φ. And for a given valid propositional formula φ, there is a way of mechanically transforming the totality of closed branches of a tableau for $F\varphi$ to a winning strategy of the Proponent, and vice versa.

The moves in dialogues are formal, they do not involve objects out there (elements of the domains of models). All that is involved is manipulation of linguistic items, such as individual constants substituted for variables. Hintikka (1973) has called his semantic games 'games of seeking and finding', or 'games of exploring the world'. Semantic games are 'outdoor' games, they are related to the activities of verifying or falsifying (interpreted) formulas,

while dialogues are 'indoor' games, related to proving—by suitably manipulating sequences of symbols—that certain (uninterpreted) formulas are valid (Hintikka, 1973, pp. 80–81). From Hintikka's vantage point, only 'outdoor' games can build a bridge between logical concepts and the meaningful use of language.

Naturally, the realism-antirealism dispute looms large here.[7] As is typical in connection with logics driven by proof theory, philosophically dialogical logic tends to be associated with antirealism or justificationism, namely the idea that semantic properties such as truth or validity can only be ascribed to sentences which can be recognised as having this property. In the transition from premises to conclusion, inference rules preserve assertibility rather than truth in abstracto. Therefore, a dialogician would typically not accept Hintikka's arguments for the 'semantic irrelevance' of dialogues. Rather, he or she would argue in favour of a justificationist theory of meaning, whereby an informal notion of proof would become a central semantic notion. A dialogician might further hold that dialogues capture such a notion of informal proof. It would be possible, but not necessary, to combine this view with the conception that dialogues actually introduce a third realm for logical theorising, adding to what proof theory and model theory have on offer.

Without entering philosophical discussions on the fundamental nature of dialogues, it can be observed that the notion of proof or inference to which dialogues give rise is distinct from the fully formal notion of proof operative in sound and complete proof systems. One may, at least so it seems, formulate reasonable dialogues—and reasonable tableau systems—even for pathologically incomplete logics, namely logics which simply do not admit of any sound and complete proof system. If so, the type of inference with which dialogues are concerned is semantic inference—with no a priori claim to always yield a recursive enumeration of the (uninterpreted) formulas of the language considered. If dialogues were all about formal proofs, it would be a contradiction in terms to speak of formal dialogues for incomplete logics.[8]

Acknowledgments

Supported by The Academy of Finland (Grant No. 207188), the University of Helsinki (Grant No. 2104027), the Institute of Philosophy, Academy of Sciences of the Czech Republic and the Grant Agency of the Czech Republic (Grant No. GA401/04/0117). The editors would like to express their thanks to those whose comments helped to improve the quality of this volume: Johan van Benthem (Universities of Amsterdam and Stanford), Jaroslav Peregrin (Acad-

[7] On antirealism (see, e.g., Dummett, 1978, 2004, 2006) and Marion, this volume.

[8] For discussion, see the contribution of Rahman and Tulenheimo in this volume (Subsection 7.2).

emy of Sciences of the Czech Republic), Shahid Rahman (University of Lille), Gabriel Sandu (University of Helsinki), and Wim Veldman (Radboud University Nijmegen). Special thanks will go to our typesetters Marie Benediktová (Prague) and Jukka Nikulainen (Helsinki) in producing the final version of the manuscript.

References

Beth, E. W. (1959). *The Foundations of Mathematics*. Amsterdam: North-Holland.

Bradfield, J. (2004). On independence-friendly fixpoint logics. *Philosophia Scientiae*, 8:125–144.

Bradfield, J. (2006). Independence: logics and concurrency. In Aho, T. and Pietarinen, A.-V. (eds.), *Truth and Games: Essays in Honour of Gabriel Sandu*, Acta Philosophica Fennica 79, Helsinki: Societas Philosophica Fennica, 47–70.

Bradfield, J. and Fröschle, S. (2002). Independence-friendly modal logic and true concurrency. *Nordic Journal of Computing*, 9:102–117.

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: MacMillan.

Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.

Dummett, M. (2004). *Truth and the Past*. New York: Columbia University Press.

Dummett, M. (2006). *Thought and Reality*. Oxford: Oxford University Press.

Fitting, M. (1969). *Intuitionistic Logic—Model Theory and Forcing*. Amsterdam/London: North-Holland.

Gale, D. and Stewart, F. M. (1953). Infinite games with perfect information. In Kuhn, H. W. and Tucker, A. W. (eds.), *Contributions to the Theory of Games II*, Annals of Mathematics Studies 28, pages 245–266. Princeton, NJ: Princeton University Press.

Hand, M. (1989). Who plays semantical games? *Philosophical Studies*, 56:251–271.

Henkin, L. (1950). Completeness in the theory of types. *Journal of Symbolic Logic*, 15(2):81–91.

Henkin, L. (1961). Some remarks on infinitely long formulas. In *Infinitistic Methods*, pages 167–183. Oxford: Pergamon.

Hilpinen, R. (1982). On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics. *The Monist*, 65:182–188.

Hintikka, J. (1955). Form and content in quantification theory. *Acta Philosophica Fennica*, 8:7–55.

Hintikka, J. (1968). Language-games for quantifiers. In *American Philosophical Quarterly Monograph Series 2: Studies in Logical Theory*, pages 46–72. Oxford: Blackwell.

Hintikka, J. (1973). *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*. Oxford: Clarendon.

Hintikka, J. (1995). What is elementary logic? Independence-friendly logic as the true core area of logic. In Gavroglu, K., Stachel, J., and Wartofsky, M. W. (eds.), *Physics, Philosophy and the Scientific Community*, pages 301–326. New York: Kluwer.

Hintikka, J. (1996). *The Principles of Mathematics Revisited*. New York: Cambridge University Press.

Hintikka, J. and Kulas, J. (1985). *Anaphora and Definite Descriptions*. Dordrecht: Reidel.

Hintikka, J. and Sandu, G. (1989). Informational independence as a semantical phenomenon. In Fenstad, J. E., et al. (eds.), *Logic, Methodology and Philosophy of Science vol. 8*, pages 571–589. Amsterdam: Elsevier.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A. (eds.), *Handbook of Logic and Language*, pages 361–410. Amsterdam: Elsevier.

Hintikka, J. and Rantala, V. (1976). A new approach to infinitary languages. *Annals of Mathematical Logic*, 10:95–115.

Hodges, W. (1983). Elementary predicate logic. In Gabbay, D. and Guenthner, F. (eds.) *Handbook of Philosophical Logic, vol. I*. pages 1–131. Dordrecht: Reidel.

Hodges, W. (1997a). Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL*, 5:539–563.

Hodges, W. (1997b) Some strange quantifiers. In Mycielski, J., Rozenberg, G., and Salomaa, A. (eds.) *Structures in Logic and Computer Science*, Lecture Notes in Computer Science, Vol. 1261, pages 51–65. London: Springer.

Hodges, W. (2006a). Logic and games. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. (Summer 2006 Edition). http://plato.stanford.edu/entries/logic-games.

Hodges, W. (2006b). The logic of quantifiers. In R. E. Auxier and L. E. Hahn (eds.), *The Philosophy of Jaakko Hintikka*, Library of Living Philosophers, vol. 30, pages 521–534. Chicago, IL: Open Court.

Hoges, W. and Krabbe, E. C. W. (2001). Dialogue foundations. Part I (Nilfrid Hodges): "A sceptical look" (17–32); Part II (Krabbe, E. C. W.): "Dialogue logic restituted" (33–49). In *Proceedings of the Aristotelian Society, Supplementary Volume 75*, pages 17–49.

Hyttinen, T. (1990). Model theory for infinite quantifier logics. *Fundamenta Mathematicae*, 134:125–142.

Hyttinen, T. and Tulenheimo, T. (2005). Decidability of IF modal logic of perfect recall. In Schmidt, R., Pratt-Hartmann, I., Reynolds, M., and Wansing, H. (eds.), *Advances in Modal Logic vol. 5*, pages 111–131. London: King's College London Publications.

Karttunen, M. (1984). *Model Theory for Infinitely Deep Languages*. Annales Academiae Scientiarum Fennicae, Series A, Mathematica, Dissertationes, vol. 50. University of Helsinki.

Krynicki, M. and Mostowski, M. (1995). Henkin quantifiers. In Krynicki, M., Mostowski, M., and Szczerba, L. (eds.), *Quantifiers: Logics, Models and Computation, vol. 1*, pages 193–262. Dordrecht: Kluwer.

Lorenz, K. (1968). Dialogspiele als semantische Grundlagen von Logikkalkülen. *Archiv für mathematische Logik und Grundlagenforschung*, 11:32–55 & 73–100.

Lorenzen, P. (1961). Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods*, pages 193–200. Oxford: Pergamon.

Margolis, J. Z. (1987). *Science without Unity: Reconciling the Human and Natural Sciences*. Oxford: Blackwell.

Makkai, M. (1977). Admissible sets and infinitary logic. In Barwise, J. (ed.), *Handbook of Mathematical Logic*, pages 233–281. Amsterdam: North-Holland.

Pietarinen, A.-V. (2001a). Propositional logic of imperfect information: foundations and applications. *Notre Dame Journal of Formal Logic*, 42:193–210.

Pietarinen, A.-V. (2001b). Intentional identity revisited, *Nordic Journal of Philosophical Logic*, 6:144–188.

Pietarinen, A.-V. (2003a). Games as formal tools versus games as explanations in logic and science. *Foundations of Science*, 8:317–364.

Pietarinen, A.-V. (2003b). Logic, language games and ludics. *Acta Analytica*, 18:89–123.

Pietarinen, A.-V. (2003c). What do epistemic logic and cognitive science have to do with each other? *Cognitive Systems Research*, 4:169–190.

Pietarinen, A.-V. (2004). Peirce's diagrammatic logic in IF perspective. In Blackwell, A., Marriott, K., and Shimojima, A. (eds.), *Diagrammatic Representation and Inference, Lecture Notes in Artificial Intelligence* 2980, pages 97–111. Berlin: Springer.

Pietarinen, A.-V. (2006a). Independence-friendly logic and incomplete information. In van Benthem, J., Heinzmann, G., Rebuschi, M., and Visser, H. (eds.), *The Age of Alternative Logics: Assessing Philosophy of Logic and Mathematics Today*, pages 243–259. Dordrecht: Springer.

Pietarinen, A.-V. (2006b). *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication* (Synthese Library 329). Dordrecht: Springer.

Pietarinen, A.-V. (2008). Who plays games in philosophy? In Hale, B. (ed.), *Philosophy Looks at Chess*, pages 119–136. Chicago IL: Open Court.

Pietarinen, A-V. and Snellman, L. (2006). On Peirce's late proof of pragmaticism. In Aho, T. and Pietarinen, A.-V. (eds.), *Truth and Games: Essays in Honour of Gabriel Sandu*, Acta Philosophica Fennica 79, Helsinki: Societas Philosophica Fennica, 275–288.

Rahman, S. and Keiff, L. (2005). On how to be a dialogician. In Vanderveken, D. (eds.), *Logic, Thought and Action*, Logic, Epistemology and the Unity of Science, vol. 2, pages 359–408. Dordrecht: Springer.

Rahman, S., and Rückert, H. (2001). Dialogical connexive logic. *Synthese*, 127(1–2):105–139.

Rahman, S., Rückert, H., and Fischmann, M. (1997). On dialogues and ontology. The dialogical approach to free logic. *Logique et Analyse*, 160:357–374.

Sandu, G. (1993). On the logic of informational independence and its applications. *Journal of Philosophical Logic*, 22:29–60.

Sandu, G. and Pietarinen, A.-V. (2001). Partiality and games: propositional logic. *Logic Journal of the IGPL*, 9:101–121.

Sandu, G. and Pietarinen, A.-V. (2003). Informationally independent connectives. In Mints, G. and Muskens, R. (eds.), *Logic, Language and Computation vol. 9*, pages 23–41. Stanford: CSLI.

Sevenster, M. (2006a). On the computational consequences of independence in propositional logic. *Synthese*, 149:257–283.

Sevenster, M. (2006b). *Branches of imperfect information: games, logic, and computation*. PhD thesis, ILLC, Universiteit van Amsterdam.

Skolem, Th. (1920). Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. In *Skrifter utgit av Videnskabsselskapet i Kristiania, I*, pages 1–36. Matematisk-naturvidenskabelig klasse no. 4.

Smullyan, R. (1968). *First-Order Logic*. New York: Dover Publications, 1995 (Originally appeared 1968).

Tulenheimo, T. (2003). On IF modal logic and its expressive power. In Balbiani, Ph., Suzuki, N.-Y., Wolter, F., and Zakharyaschev, M. (eds.), *Advances in Modal Logic vol. 4*, pages 475–498. King's College London Publications. Milton Keynes, UK.

Tulenheimo, T. and Sevenster, M. (2006). On modal logic, IF logic and IF modal logic. In Governatori, G., Hodkinson, I., and Venema, Y. (eds.), *Advances in Modal Logic vol. 6*, pages 481–501. King's College London Publications. Milton Keynes, UK.

Ulam, Stanislaw M. (1960). *A Collection of Mathematical Problems*. Groningen: Interscience.

Väänänen, J. (2006). A remark on nondeterminacy in IF logic. In Aho, T., and Pietarinen, A.-V. (eds.), *Truth and Games: Essays in Honour of Gabriel Sandu*, Acta Philosophica Fennica 79, Helsinki: Societas Philosophica Fennica, 71–78.

Väänänen, J. (2007). *Dependence Logic: A New Approach to Independence Friendly Logic*. New York: Cambridge University Press.

van Benthem, J. F. A. K. (2001). Logic and games (lecture notes). Amsterdam: ILLC and Stanford: CSLI (printed version 2001), unpublished.

Vaught, R. L. (1973). Descriptive set theory in $L_{\omega_1\omega}$. In Mathias, A. and Rogers, H. (eds.), *Cambridge Summer School in Mathematical Logic*, Lecture Notes in Mathematics vol. 337, pages 574–598. Berlin: Springer.

Part I

PHILOSOPHICAL ISSUES

Chapter 1

WHY PLAY LOGICAL GAMES?

Mathieu Marion
Département de philosophie, Université du Québec à Montréal
marion.mathieu@uqam.ca

Abstract Game semantics has almost achieved the status of a paradigm in computer science but philosophers are slow to take notice. One reason for this might be the lack of a convincing philosophical account of logical games, what it means to play them, for the proponent to win, etc., pointedly raised by Wilfrid Hodges as the 'Dawkins question'. In this paper, I critically examine two available answers: after a brief discussion of an argument by Tennant against Hintikka games, I focus on Lorenzen's attempt at providing a direct foundation for his game rules in the life-world, showing some of the difficulties inherent to that project. I then propose an alternative based on the theory of assertions developed by Dummett and Brandom.

We owe to Paul Lorenzen an extraordinarily rich intuitive idea, first presented in his papers 'Logik und Agon' (Lorenzen, 1960) and 'Ein dialogisches Konstruktivitätskriterium' (Lorenzen, 1961), the fruitfulness of which we are barely encompassing today. In the language of game theory, it is the idea of defining logical particles in terms of rules for non-collaborative, zero-sum games with perfect information between two-persons, a proponent and an opponent, and to define truth in terms of the existence of a winning strategy for the proponent.[1] In a nutshell,[2] a 'dialogical' or 'Lorenzen game', as I shall call them is always played in alternate moves between an opponent *O* and a proponent *P*, who begins the game by asserting a given statement; games are thus always in the form of a finite sequence of alternate moves. Lorenzen distinguished between 'particle' and 'structural' rules for these games. Particle rules provide the meaning of the logical connectives. One should distinguish here between players and their role, as attacker or defender: in the initial round *P*

[1] The idea that one should use game theory here was not, however, Lorenzen's, we owe it to his student, Kuno Lorenz.

[2] Alas, there are no textbooks for dialogical logic, but one may find useful introductions in Felscher (1986), Lorenz (1981) or Rückert (2001); I follow for the most part the latter presentation in this paper. A number of seminal papers are collected in Lorenzen and Lorenz (1978). As the point of this paper is purely philosophical, I shall remain at an informal level throughout.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

has undertaken to be the defender, with O the attacker, hence the choice of their names, but these roles might be inverted during the course of a play, as will be pointed out shortly. Therefore, informally stated, for P as the defender and O as the attacker, the rules are as follows: when P asserts $\varphi \,\&\, \psi$, O chooses one of the conjuncts and P must defend it, so the game continues for that conjunct. Since O chooses, P has to have a defence of both conjuncts up her sleeve if she hopes to win. When P asserts $\varphi \vee \psi$, then O asks that P chooses and defends one of the disjuncts; she can thus choose the disjunct for which she has a defence. For an implication $\varphi \rightarrow \psi$, O will assert φ thus forcing P to defend ψ (P also has the possibility to attack φ). With $\neg\varphi$, roles are exchanged, as O has now to defend φ against attacks from P. For quantifiers, when P asserts $\forall x A(x)$, O chooses a value for x and P must then show that it has the property A, and when P asserts $\exists x A(x)$, then O asks that P exhibits an x that has the property A.

Structural rules concern the structure of the games, e.g., the already-mentioned rule stating that players move alternatively or a rule forbidding delaying tactics. One important rule concerns atomic formulas: P can only assert an atomic formula if it had been previously asserted by O, so that the winning strategy for P would be independent of any information about atomic facts. This is known as the 'formal rule', which makes room for 'formal' as opposed to 'material' games, and for a definition of logical validity. In material games, this rule is replaced by a rule stating that *only* true atomic propositions may be asserted, leading to a definition of validity as general truth. Lorenzen games always terminate in a finite number of steps and the winning rule states that the one who has no possible moves left has lost. Thus, if P has at least a move for any move chosen by O, then P will win and a formula is valid if and only if P has a (formal) winning strategy for that formula (Rückert, 2001, 173–174).

Another proposal for 'logical games', as they may generally be called, was put forth by Jaakko Hintikka some years later (Hintikka, 1968, 1972),[3] the heart of which being a game-semantic reading of the quantifiers first suggested by Leon Henkin (Henkin, 1961).[4] One should recall that in a formula of the form:

$$\forall x \exists y A(x, y) \tag{1}$$

the choice of (a value for) y depends on the prior choice of (a value for) x and that one may replace it with a second-order formula, involving what is known as a 'Skolem function':

[3]For an introduction to game-theoretic semantics, see Hintikka and Sandu (1997).

[4]It was found out later on that C. S. Peirce, one of the inventors of quantification theory, was actually the first to propose a game-semantic interpretation of the quantifiers. See Hilpinen (1982) and Pietarinen (2006). For an nice example of exegesis using game semantics, see Pietarinen and Snellman (2006). Incidentally, Henkin's paper appeared in the same volume as one of Lorenzen's first papers (Henkin, 1961, Lorenzen, 1961). This coincidence is not so fortuitous, as they were both at the Institute for Advanced Studies in the late 1950s, possibly being influenced by von Neumann and Morgenstern, both pioneers of game theory, as well as from Tarski.

$$\exists F \forall x A(x, F(x)) \tag{2}$$

Hintikka's suggestion was to follow Henkin in reading (1) in terms of a game between an opponent, variously named 'Nature', 'initial falsifier', or '∀belard', and a proponent, 'Myself', 'initial verifier' or '∃loise'. In (1), the opponent makes the first move and chooses an x, then the proponent must find for that x a y such that $A(x, y)$. If ∃loise finds a y for every x that ∀belard throws at her, so to speak, then she wins the game, otherwise, she loses. Of course, if a function such as F in (2) is available to the proponent, she only has to apply it to find a y for any x chosen by the opponent and thus win the game; this is why the existence of a Skolem function is equivalent to the existence of a winning strategy for the proponent. Furthermore, the existence of a winning strategy is necessary and sufficient for the sentence to be true. This reading of the quantifiers has been extended by Hintikka to 'branching' or 'Henkin quantifiers', first introduced by Henkin in the very same paper in which he introduced his game-semantic interpretation (Henkin, 1961). I shall come back to this briefly.

Hintikka's next move was to extend this reading of the quantifiers to conjunction and disjunction, by reading $\varphi \,\&\, \psi$ as 'All the sentences φ and $\psi \ldots$', where the falsifier chooses one of φ and ψ and the game proceeds accordingly, and $\varphi \vee \psi$ as 'There is one of φ or $\psi \ldots$', where the verifier chooses instead. To obtain the complete set of rules, one must add a rule for negation as the exchange of roles.[5] There is, however, no specific rule for material implication $\varphi \rightarrow \psi$, which is simply defined classically as $\neg\varphi \vee \psi$. One can thus see that Hintikka's reading of the quantifiers is at the heart of his game semantics. Another important semantic idea is the idea of negation as a 'responsibility shift'.

Hintikka has made numerous controversial claims on behalf of his game-theoretical semantics, especially in connection with Independence-Friendly Logic which covers the fragment of second-order logic involving branching quantifiers, including an ambitious plea for re-thinking of the very nature of logic in *The Principles of Mathematics Revisited* (Hintikka, 1996). I shall not discuss these here. One should note, however, that Lorenzen also made radical claims concerning his games. He was a logical monist and his original intention was to provide philosophical foundations for intuitionistic logic, i.e., the 'Heyting calculus' in the following quotation from his John Locke Lectures.[6]

> Philosophically there is no reason to start with the historical fact that Heyting published a certain calculus or to look for an interpretation of that calculus. It

[5]Hintikka's approach is model-theoretic: games will terminate in a finite number of steps with atoms, then one looks at the model, if the atom is valued as true, then the verifier wins, otherwise, she loses.

[6]More precisely, Lorenzen was expressly hoping to recover Beth's tableaux rules for intuitionistic logic. See Barth and Krabbe (1982, 12–13).

> is however, reasonable to start with material dialogues, to formalize this game, to look for admissible rules for winning-positions; this procedure leads us directly to an interpretation of the Gentzen calculus and then indirectly to an interpretation of the Heyting calculus. I would claim, therefore, that the dialogical approach justifies the logical intuitions of Brouwer and Heyting. (Lorenzen, 1969, 39)

His claim was thus that his dialogical approach justifies intuitionistic logic and not, conversely, that intuitionistic logic justifies his choice of rules. This is indeed a rather ambitious claim (to which I shall come back). The equivalence theorem necessary between proofs in Gentzen's natural deduction system for intuitionistic logic and strategies for winning dialogues was obtained only in 1985 by Walter Felscher (Felscher, 1985), at the end of a long search for the right set of restrictions on structural rules for dialogues needed to obtain intuitionistic provability. Kuno Lorenz had in the meantime realized that a slight variation in one of the structural rules would give classical logic (Lorenz, 1968)—yet another point to which I shall come back.

After a period of neglect, dialogical logic has enjoyed a revival recently, after Andreas Blass first proposed (Blass, 1992) to use Lorenzen's ideas to provide a semantics for the then newly invented linear logic of Jean-Yves Girard (Girard, 1987). Blass's paper sparked numerous developments, with new competing semantics: Hyland-Ong games (Hyland, 1997; Hyland and Ong, 2000), Abramsky games (Abramsky and Jagadeesan, 1994; Abramsky, 1997; Abramsky, 2006), Japaridze games (Japaridze, 1997), and even further logical developments, with Japardize's computability logic (Japaridze, 2003), and Girard's 'ludics' (Girard, 2001). Game semantics allows one to provide semantics to a variety of logical systems and programming languages, and has thus emerged as a new paradigm within computer science. However, while computer scientists might have perfectly good reasons for turning to game semantics, the idea is only slowly picking up within philosophical circles. The obvious reason for this is that better-known paradigms, for example, truth-conditional semantics, have more firmly established pedigrees. Philosophers won't budge until they are shown that, in some sense, game semantics is a better alternative and they will only shrug their shoulders when pointed out that, e.g., it allows for the construction of syntax-independent, 'fully abstract' models for programming languages. Some prejudices definitely need to be overcome before game semantics is to displace its rivals in their minds. Some objections are devoid of any merit, such as the claim, often voiced, according to which dialogical games are needlessly complicated: strategies for Lorenzen games amount merely to reading proofs in Gentzen's natural deduction systems upside down (Lorenzen, 1987, 81, 96), and one can learn to do so probably as easily as one learns how to drive on the left side of the road, once one has learned to drive on the right side. From a philosophical point of view, however, more reticence needs to be overcome, so the main task is to provide a coherent,

believable story for seeing logic in terms of dynamic interaction between two players. In other words, one must give not a contrived but a natural answer to the question: Why play logical games? The point of playing a game is to win, but what is a defender doing when trying to win a logical game? What is the motivation for the attacker? These questions have been raised recently by Wilfrid Hodges:

> In most applications of logical games, the central notion is that of a winning strategy for the [proponent]. Often these strategies (or their existence) turn out to be equivalent to something of logical importance that could have been defined without using games—for example a proof. But games are felt to give a better definition because they quite literally supply some motivation: [the proponent] is trying to win. This raises a question that is not of much interest mathematically, but it should concern philosophers who use logical games. If we want [the proponent's] motivation in a game G to have any explanatory value, then we need to understand what is achieved if [the proponent] does win. In particular we should be able to tell a realistic story of a situation in which some agent called [the proponent] is trying to do something intelligible, and doing it is the same thing as winning in the game. (Hodges, 2004, § 2)

Hodges' question is thus a request for a description of a realistic situation in which the proponent is trying to do something which is the same as winning in a logical game. An answer to it is rather important, as it is from this story that the particle and structural rules should emerge, so to speak. It is also not just the obvious prerequisite to any attempt at convincing sceptics about the value of game semantics, it strikes right at the heart of claims made on the behalf of logical games, such as Lorenzen's claim that his games justify intuitionistic logic, and not vice-versa, or Hintikka's proposals to reform logic: if no good answer is forthcoming, these claims will simply fail to convince, as they have done so far. In the remainder of this paper, my task will be to assess available answers and to provide a new one.

At the moment, there are only two answers to Hodges' question in the literature, for Lorenzen games and for Hintikka games, and Hodges has provided criticisms of both (Hodges, 2001, 2004, 2006). He took a rather stern view in both cases, concluding some harsh comments on Hintikka with the remark "it is a little disappointing that nobody took the trouble to look for a better story" (Hodges, 2004, § 3). And Lorenzen does not fare better: "it turns out to be embarrassingly easy to make mincemeat out the fine details of Lorenzen's claims" (Hodges, 2001, 22). Hodges is at any rate not looking forward to be convinced: "each claim of this kind needs its own deconstruction" (Hodges, 2001, 25). As it is not possible fully to discuss both programmes here, the emphasis in what follows will be on Lorenzen games, with remarks on Hintikka games, immediately below, kept to a minimum. At all events, I have already given reasons to reject Hintikka's answer—especially concerning his use of Wittgenstein's

notion of 'language-games' in that context—, and I shall not repeat them here (Marion, 2006).

In the case of Hintikka, the problem of finding a convincing story is reduced to that of providing a story for the quantifiers, since, as I have shown, his central semantic idea concerns the quantifiers (forgetting here the other crucial thesis about negation as 'responsibility-shift'). This is why Hintikka introduced his 'language-games' of 'seeking and finding' (Hintikka, 1973, Chapter 3). It seems to me that the main objection to Hintikka's answer was already put forth by Neil Tennant in the 1970s (Tennant, 1979); I shall briefly rehearse it because I wish to add a proviso. As I pointed out, Hintikka extended his game-semantic reading to 'branching' or 'Henkin quantifiers'. When we wish to say that for all x there is a y and that for all z there is a w, such that $A(x, y, z, w)$, if we want the choice of y to depend on x and the choice of w to depend on z but *not* on x, then the usual notation is inappropriate, since, according to the usual conventions about scope, the expression

$$\forall x \exists y \forall z \exists w A(x, y, z, w) \tag{3}$$

makes the choice of w depend not only on z but also on x. To express this, one needs 'branching' quantifiers, for which Hintikka devised a 'slash' notation:

$$(\forall x)(\forall z)(\exists y/\forall z)(\exists w/\forall x)A(x, y, z, w) \tag{4}$$

The slash in '$\exists w/\forall x$' means that the choice of w is made independently of that of x. Here too, there is a corresponding second-order formula:

$$\exists F \exists G \forall x \forall z A(x, F(x), z, G(z)) \tag{5}$$

where functions F and G will provide the winning strategy for the proponent. Again, I must skip here discussing further claims by Hintikka, e.g., about non-compositionality, so that Tarski's well-known truth definition for first-order logic could not be extended to provide a semantics for Independence-Friendly Logic, etc.[7] One should merely note that the initial verifier's winning strategy is provided here by a set of Skolem functions for which one can merely claim 'existence' in the classical, non-constructive sense of the term. This means, therefore, that the initial verifier has no available knowledge of the set of functions that would provide her with a winning strategy, she could only know that such a set exists. Although Hintikka remains undeterred, e.g., at Hintikka (1998, 171, n. 34), this means that, as Tennant pointed out, "no person could apply these functions in a way that exhibits strategic intent" (Tennant, 1979, 305). This does not mean that the initial verifier could not win a given game, as when one

[7]Hodges has given a compositional semantics in Hodges (1997), but see also Abramsky (2006). For a discussion of this issue and further references, see Hodges (2006).

plays chess, for example. Nevertheless, it thus seems hardly to make sense, in light of Hodges question, to speak of asserting a sentence that is true in reference to a game for which a winning strategy exists but cannot be known to the initial verifier. This point must be qualified, however, as the analogy with chess already suggests: if one distinguishes properly between game level, where anyone who has mastered the particle rules can go on playing—and can thus be said to know their meaning—, and the strategy level, which should involve the handling of some constructive procedure, the point at stake being that there is none here to handle.[8] So, Tennant's argument only applies at the strategy level, and this considerably weakens it.

This critique does not nullify Hintikka games in the least. If anything, it shows that Hodges was right in finding Hintikka's original motivation unsatisfactory.[9] It leaves the door open to other suggestions. For example, to a rather promising approach, which looks at these games as modelling cases where players have to cope with imperfect information, e.g., the proponent has to make a choice without knowledge of the opponents previous move.[10] This is the direction in which Hintikka games have already evolved, in particular in the work of Gabriel Sandu and Ahti-Veikko Pietarinen,[11] and Johan van Benthem has shown how it overlaps further with game theory and with recent developments in dynamic epistemic logic (van Benthem, 2003, 2006). There is thus a different answer to Hodges' question in gestation here, and this approach might provide a better understanding of these games and their applications.

*

The answer to Hodges' question that can be garnered from Lorenzen's writings also has its own difficulties.[12] Like many logicians, Lorenzen felt dissatisfied with the usual Tarski-style semantical definitions, e.g, '*A* & *B* is true if and only if *A* is true *and B* is true' and '*A* & *B* is false if and only if *A* is false *or B* is false', since these presuppose the availability of a metalinguistic 'and' and 'or' (Lorenzen, 1987, 60, 88). As Jean-Yves Girard once put it: to understand Tarski, you need 'Mr. Metatarski', and so on (Girard, 1999, Section 23).

[8] I owe this point to Helge Rückert.

[9] Not, however, for his stated reasons, e.g., his inability to make sense of the role of Nature as the initial falsifier in Hodges (2004).

[10] There is a very brief suggestion to that effect in Marion (2006, 268).

[11] E.g., Sandu and Pietarinen (2003).

[12] The following remarks are not based on a exhaustive review of the literature, and I owe an apology to German readers for my excessive reliance on a handful of English translations. (For example, Lorenzen (1987) is a translation of Lorenzen (1968) and Lorenzen (1974), along with some papers, including Lorenzen (1982). For a short overview and useful bibliography of the Erlangen school, see Gethman and Siegwart (1994), one should also consult the papers collected in Butts and Brown (1989), but beware of the unsympathetic, biased overview in Bubner (1981, 142–153).

Lorenzen was thus looking around for a strict foundation[13] and found what seems to be his key idea (and that of the Erlangen School, which spawned around his work), in the philosophy of Hugo Dingler.[14] Dingler's ideas can be captured in terms of Hans Albert's 'Münchhausen Trilemma', according to which any attempt at a foundation is bound either to lead to an infinite regress, to be circular (as one presupposes what one wishes to ground), or to end arbitrarily, "in the middle" (Albert, 1985, 18). Dingler chose the last option (Dingler, 1931, 21; Dingler, 1955, 97; Albert, 1985, 19n. and 41f.), and Lorenzen followed him, using Neurath's metaphor of the boat that has to be rebuilt at the sea (Lorenzen, 1987, 16). Dingler asked indeed that our scientific discourse be methodically reconstructed 'from the ground up', step by step—to avoid circles, every step must be constructed only on the basis of steps already carried out (Dingler, 1964, 26)—so that it could be open to rational discussion. But he rejected the sort of reductionist programmes typical of his days, such as Carnap's *Aufbau*,[15] i.e., he rejected the idea that the vocabulary of physics can be reduced to an empirical base vocabulary, be it a phenomenal or physicalist language.[16] And to begin 'in the middle' did not mean, as implied by Albert, to begin at an arbitrary point: it meant for Dingler that we have to start the reconstruction in the middle of our 'civil life', i.e., in an hypothetical state of scientific innocence,[17] where all we have is our concrete *actions*. Thus, according to Dingler, the buck stops at them: "all sciences must have their ultimate basis in the theory of action" (Dingler, 1931, 32). To take the example of geometry, what we know about space is said to depend on operations performed within this life-world.[18] The same goes for measuring time or chronometry, and these operations will serve as the basis upon which

[13]That he ultimately succeeds or not in avoiding any recourse to the metalanguage in the rules for his games is not an issue that can be discussed here, although it is implied below that he did not.

[14]For Dingler's bibliography, see Schroeder-Heister (1981). Dingler's collected works are now available in electronic form as Dingler (2004).

[15]On the relation between Carnap and Dingler, see Wolters (1985).

[16]One should therefore note here the connection between Dingler and what Robert Brandom has recently called the pragmatic challenge to the classical project of analysis (Brandom, 2008, 3). Brandom has in mind primarily the later Wittgenstein and Wilfrid Sellars, but it is clear that Lorenzen and the Erlangen School should be understood as belonging to that camp. See, for example, Gethmann (1979).

[17]One should note *en passant* that, although Dingler rejected attempts at a transcendental foundation, his programme is here related to Husserl's project in Appendice III to the *Crisis of European Sciences* on the origins of geometry (Husserl, 1970), i.e., to Husserl's claims about the possibility and the necessity to re-activate the 'evidences' on which the first geometers build geometry—one would say: its 'proto-foundation'—, and whose validity is meant to trickle down the chain of inferences. On the relation between Husserl and Dingler, see Wolters (1991). It is for reasons of this sort that the "constructive philosophy" of Lorenzen and of the Erlangen School has been characterized as "phenomenology after the linguistic turn" (Gethman and Siegwart, 1994, 228).

[18]Dingler speaks of a *Lebensstandpunkt* (Dingler, 1964, 42). However, one should not read too much in my use here of my expression 'life-world'. The connections with Husserl's *Lebenswelt* (see preceding footnote) has been noted many times, e.g., (Gethmann, 1979, 39) with the usual proviso.

physics can be reconstructed. Dingler believed that he could thus show that the axioms of Euclidean geometry are the only *operationally* true ones, hence his life-long opposition to Einstein's relativity theory, which requires the validity of a non-Euclidean geometry.[19]

Influenced here by Oskar Becker as well as Hugo Dingler, Lorenzen originally proposed in *Einführung in die operative Logik und Mathematik* a reconstruction of mathematics on a partly formalist 'operative' basis, i.e., on mechanical operations on strings of symbols in accordance to given rules (Lorenzen, 1955). For example, the basic arithmetical operation is counting, numerals being constructed by the operation schematized as follows:[20]

$$\begin{array}{c} | \\ n \rightarrow n\,| \end{array}$$

The domain of these rules of transition, to which one needs merely to add the above rule for numerals in order to build mathematics (minus geometry) (Lorenzen, 1987, 69–70), is called 'protologic' (Lorenzen, 1955, 1987, 61 and 67). Protologic is thus a theory of formal systems within which one studies principles for the admissibility of inference rules. One of Lorenzen's lasting achievements in that book, to use Peter Schroeder-Heister's words (Schroeder-Heister, 2008, 229), is to have been the first to formulate an inversion principle (Lorenzen, 1955, 30f.), and to apply it to infer elimination rules from introduction rules. Indeed, Lorenzen was the first to introduce in proof-theory the concept of admissibility: a rule is admissible if adding it to the set of rules of a given system does not enlarge its set of derivable sentences (Lorenzen, 1955, § 2), and he further introduced a notion of elimination procedure in order to give an operative meaning to the notion of admissibility: a rule is admissible if every application of it from every derivation in the system to which it is added can be eliminated (Lorenzen, 1955, § 3). This constitutes, as Schroeder-Heister aptly notes, the true intuitionistic core of Lorenzen's conceptions (Schroeder-Heister, 2008, 217–218).

The inversion principle was taken up and generalized by Dag Prawitz in his classical study, *Natural Deduction; A Proof-Theoretical Study* (Prawitz, 1965), and used it further as the basis of a well-known argument which purports to show the more natural character of intuitionistic logic, because the elimination

[19]Dingler mounted a failed challenge at the 86th *Naturforschersammlung* held at Bad Neuheim in 1920, which was, as it turns out, a turning point in German physics, as relativity theory was then finally adopted by the German physicists and Dingler became isolated; this is the beginning of numerous professional problems, that led eventually to his siding with Lenard's *Deutsche Physik*, with all the obvious consequences. A similar but slightly different operational reconstruction of geometry and chronometry was carried out in the Erlangen school, see, e.g., Janich's 'proto-physics' of time in Janich (1985), or, for geometry the essays in Lorenzen (1987, Part vi).

[20]Since there are only operations on signs, and no impredicative definitions, the 'operative' mathematics developed in Lorenzen (1955) and, further, in Lorenzen (1971) stands closest to Weyl's predicativism.

rule for double negation in classical logic does not respect this inversion principle.[21] However, by the time these developments took place, Lorenzen had more or less lost interest in the topic. Nevertheless, it should be noted that Lorenzen had in the 1960s a parallel argument from the point of view of his protologic:

> In fact, the usual logic can be operatively—that is on the basis of schematic operations—interpreted this way. With the exception of negation, everything is exactly as it is in the classical theory. For negation we have, in contrast, at first only intuitionistic logic with which, however, we know that we can justify two-valued logic as at least a fiction. (Lorenzen, 1987, 68)

One also should note on historical matters that Lorenzen did not have yet his insight into the dialogical nature of logic at the time he introduced his 'operative logic' in *Einführung in die operative Logik und Mathematik*. After publishing that book, he went to Princeton to meet Weyl (alas, Weyl died before Lorenzen's arrival), and the idea of a game between a proponent and an opponent (but not yet the idea of fully using game theory) came from discussions with Tarski.[22] It is difficult to make sense of the grafting of dialogical games over the conceptions set forth in 1955 book.[23]

At all events, in later presentations of his dialogical games, Lorenzen argues that his particle rules are abstracted from what he called our practical nonverbal activity (*die Praxis unseres sprachfreien Handelns*) or our prelogical speech practice (*vorlogische Redepraxis*) (Lorenzen, 1982, 29, 35, 1987, 83, 87), expressions which he obviously got from Dingler.[24] So a rational reconstruction of logic will have as a starting point the activities within a prelogical speech practice, from which one can eventually extract (after going through steps concerning predication, etc.) *the* particle and structural rules of the dialogical games for *the only operationally true logic*, intuitionistic logic. As Lorenzen writes:

> In the context of a specific practical activity any normal person can learn how to use sentences of, say, the form N [does] P or N [is] Q [...] We learn this kind of sentence and the words that appear in them exemplarily. In this way we have a speech practice that is justified within the context of practical activity. This is what Bühler called empractic justification. Only by participating in an activity do we acquire the speech appropriate to that activity. We learn by practice what it is to assert propositions or to contest the affirmation or denial of propositions (e.g., by nodding or shaking ones head.) We introduce a negator, ¬, where ¬*a* is used to express that we are contesting a proposition *a*. (Lorenzen, 1987, 83)

[21] See Prawitz (1977) or Dummett (1991, Chapter 9) for a more recent restatement.

[22] On this, see Lorenz (2001).

[23] Of course, these will not be discussed here. For a careful study of Lorenzen's earlier conceptions, see Peter Schroder-Heister's papers (Schroeder-Heister, 2007, 2008).

[24] There are of course many important differences between Dingler's and Lorenzen's programmes. See Janich (1985, Chapter 2) for a detailed presentation.

Lorenzen also presents his game rules as a 'normalisation' or 'regimentation' of conversational moves from the life-world. Now Hodges objected to this that Lorenzen's notions of 'attack' and 'defense' couldn't be said to be lifted from a 'prelogical speech practice'. In support, he gave three arguments, concerning particle rules, trying to show that what the opponent does in some cases cannot be really construed as an attack: sometimes it looks as he is helping instead.[25] His third argument (I shall not discuss the first two) has to do with the fact that in a dialogue, you can attack a claim either by arguing that it is not true or that it is useless for further deductions; Lorenzen has supposedly overlooked the latter case, which Hodges illustrates by quoting an exchange from Strindberg's *Miss Julie*:

> Jean: If you take my advice, you'll go to bed.
> Julie: Do you think I'm going to be ordered about by you? (Hodges, 2001, 24)

Here, Julie simply refuses to take Jean's advice, in other words, she replies to Jean's $\varphi \to \psi$ by rejecting φ. In view of this, for the opponent to concede φ is not truly to attack $\varphi \to \psi$ but somehow to help the proponent in her task of defending it. As a critique of Lorenzen's rule for $\varphi \to \psi$ this would be rather poor, as the rule is meant to capture the *semantic* content, for which the point raised in the example from *Miss Julie* is irrelevant. The point of the rule is that, when someone asserts ψ under the condition that φ—that is if φ is granted — there is no attack possible on the semantic content other than to grant φ and force the proponent to assert ψ and defend it.[26] I do not wish to argue for or against this point, but simply to note that the example from Miss Julie shows that there are in our prelogical speech practice many ways to deal with the conditional that are irrelevant to the particle rule itself, and this shows of itself that the rule is not simply 'lifted' from our 'prelogical speech practice', the element of 'regimentation' plays a crucial role. The very idea that the rules are somewhat extracted from the 'prelogical speech practice' thus becomes vague and ultimately unconvincing. (Recall that I am merely arguing here against the suggestion that the rules are to be arrived at from some analysis of the prelogical speech practice, not against the specific rules framed by Lorenzen.)

[25]There are suggestions that Lorenzen did not exactly see things the way Hodges portrays him, e.g., when Kamlah and him point out that players are "not discoursing against one another in order to carry their point, but rather with one another, so that in working together they may come up with true sentences", this being illustrated by the alleged move from the 'eristics' of the Sophists and the 'dialectics' of Socrates (Kamlah and Lorenzen, 1984, 142). This is a very interesting suggestion in itself but it is not clear if this is a correct representation of the difference between Socrates and the Sophists. At any rate, one should note that, assuming the distinction between the level of games and the level of strategies (Rückert, 2001, 175–177), any collaboration should occur at the level of strategies, while the games should remain fully agonistic. (I owe this point to Helge Rückert.)

[26]I owe this point to Helge Rückert.

The same goes for structural rules. As I already mentioned earlier, the difference between classical and intuitionistic logic hangs in Lorenzen's games merely on the difference between two structural rules. These are:

Intuitionistic rule: Each player can either attack a (complex) formula asserted by his adversary or defend herself against the last attack that has not yet been answered.

Classical rule: Each player can either attack a (complex) formula asserted by his adversary or defend herself against any attack, including those already defended (Rückert, 2001, 168).

Indeed, with help of this last rule, one can defend $\varphi \vee \neg\varphi$, but it is as easy to show that there is no defence when the first rule is applied. Here are two games, the left-hand one uses the intuitionistic rule:

$P:\ \varphi \vee \neg\varphi$	$P:\ \varphi \vee \neg\varphi$
$O:\ ?$	$O:\ ?$
$P:\ \neg\varphi$	$P:\ \neg\varphi$
$O:\ \varphi$	$O:\ \varphi$
	$P:\ \varphi$

Note that in both games, P answers the challenge choosing $\neg\varphi$ as she cannot assert an atomic formula. On the left-hand game, P ends up with no move and loses, so there is no defence of $\varphi \vee \neg\varphi$. On the right-hand side, P is now allowed to reply again to the first challenge by stating φ, as it is already asserted by O, who has then no more moves and loses.

But how can this first rule be convincingly said to be anchored in our 'prelogical speech practice', as opposed to the second? If the difference between classical and intuitionistic logic is the prohibition of repeated attacks, where is a justification for this to be found? The mere possibility of such objections shows at least this that one cannot so simply lift *one* set of rules from this hypothetical state of 'prelogical speech practice'. This much goes at least against Lorenzen's monism. (I shall come back to this point below.) But one could push this point further: it is not clear how can one define a *pre*logical (let alone pre*verbal*) state where conversations take place from which one could extract not just one specific set of rules but any set of rules. The notion of 'prelogical' conversations itself may very well be incoherent. The idea that logic is already included into the bargain, so to speak, with the ability to deploy any language or to controvert is tempting. But how one could cash it out in such foundational terms for logic is simply not obvious.

As it turns out, however, Lorenzen did not claim to lift so straightforwardly his game rules for the prelogical, empractic context of the life-world, his position is cleverer. He claimed instead that in protologic one learns rules of transition which would legitimate in turn the introduction of the particle rules:[27]

> Rules of transition [his example here is the Modus Ponens] in which we must affirm the conclusion if we have affirmed the premises are not logical rules. They are *prelogical*; they provide a set of practical linguistic activities, a set of linguistic practices, which, under rather complicated circumstances, justify the introduction of operators invented expressly for these linguistic practices, that is, logical operators. (Lorenzen, 1987, 83)

We are here back to the domain of protologic, where, as we saw, Lorenzen had accounted for the admissibility of rules in terms of eliminability. It appears therefore, that Lorenzen saw candidates for the particle rules as extracted from our 'prelogical speech practice' and then shown to be admissible by a procedure of elimination. In a paper entitled 'Protologic. A Contribution to the Foundation of Logic', Lorenzen illustrated how elimination is to proceed through a dialogue performed by Fritz and Hans, with Hans failing in his attempt to deduce a purely syntactic figure that Fritz could not deduce in turn from the first two of the three rules available to Hans (Lorenzen, 1987, 64–66). One may continue with our critical line of enquiry and ask in turn: Whence this exchange between Fritz and Hans? It immediately becomes clear that the antecedent proto- and pre-logical state in which Fritz and Hans play their eliminability game, from which one could learn which logical rules are admissible, is some sort of fiction or reconstruction. This is presumably, to use Lorenzen's own metaphor, what the rebuilding of our ship at sea amounts to. However, there is a definite air of circularity since, in order even to get to learn the schematic operations, Fritz and Hans do not start from scratch but already need logic, the same way that, to use an example to be discussed below, in order to learn the rule to be applied when ordering a coffee in a Viennese café, one already needs to master the use of assertions, or, to use a well-known example from Wittgenstein, in order to learn word by ostension, one must have already mastered the practice of definitions by ostension. Here, to 'reconstruct step by step' cannot amount to giving the expected sort of 'foundations'. Lorenzen was certainly aware of this kind of difficulty, as he dismissed circularity arguments such as the above as *verbalen Nebel* (Lorenzen, 1982, 29, 1987, 83). One wonders, however, where the fog truly remains. Construing logical inference in dynamic terms as action or operation is a pregnant idea, Lorenzen's approach of rules *via* admissibility and elimination is certainly one of his lasting achievements, but it does not have to stand and fall with this Dinglerian attempt at anchoring it in our life-world.

[27] This point seems to have been missed in Hodges (2001).

As with Hintikka games, the above does not amount to an argument against Lorenzen games as such but against the story he provided as a motivation for them, i.e., against a possible answer to Hodges' question that one could get from his writings. I would like to give further emphasis to this critique by a short digression concerning Karl Bühler's notion of empractic speech, alluded to by Lorenzen. Here is another allusion:

> You can play ball without using words. In this prelinguistic activity we can 'empractically'—as Bühler called it—define the use of simple words. [...] I trust you can easily imagine the sort of practical situations in which Leo would utter imperative sentences like the following:
> Throw!
> Throw ball!
> Mao! Throw ball!
> or *indicative sentences* like:
> Ball does fall
> not: Mao does throw
> [...] The sentence forms that have been 'empractically' justified to this point can be extended further in various ways before we introduce logical operators. (Lorenzen, 1987, 139–141)

In his 1934 book, *Theory of Language*, Bühler put forth an extended version of the 'context principle'. It is 'extended' because Bühler's notion of context (*Umfeld*) is not merely linguistic (as it would be for, say, Frege), it is also non-linguistic, in which case it is said to be either physical or behavioural; the latter is called 'empractic'. A typical case of empractic speech occurs when, sitting in a Viennese café, I see a waiter coming towards me and utter to him: '*einen schwarzen*', and he comes back a minute later with a black coffee (Bühler, 1990, 178). One must be careful in delineating Bühler's point here. In that passage, Bühler argues that in uttering '*einen schwarzen*' I do not mentally go through a sentence such as 'Please bring me a black coffee'.[28] (This is a point for which Wittgenstein is famous (Wittgenstein, 1997, §§ 19–20)). That one can always construct such a sentence does not prove anything (Bühler, 1990, 178). He had already argued, with help of similar examples, that not all language signs are 'symbols', some are 'signals' (Bühler, 1990, 122)—in Wittgenstein's words, they are a different tool in the tool-box of language (Wittgenstein, 1997, § 11)—and that such forms of speech are neither impoverished, nor incomplete (Bühler, 1990, 122). (This is also a point made by Wittgenstein (1997, § 18).) How could this relate to rules for Lorenzen games? Let us, for the sake of the argument, grant these points made by Bühler (and Wittgenstein). One will notice that they aim at showing that *some* utterances in

[28] Kevin Mulligan has shown in Mulligan (1997), that Bühler's theory is indeed of great help to understand properly the language-game of builders at the beginning of *Philosophical Investigations* (Wittgenstein, 1997, § 2).

some contexts are not assertions of elliptic versions of declarative sentences—although they might look like it—but of an altogether different nature from assertions. But logical games are about *assertions*, so there is strictly nothing that one would say about what is specific to 'empractic' forms of speech, which do not count as assertions, that could count as grounding rules for logical games in a prelogical state. Furthermore, even if one recognizes the validity of the Bühler–Wittgenstein point that, in uttering '*einen schwarzen*' I do not mentally go through a sentence such as 'Please bring me a black coffee', it remains that such uses of language can be seen as parasitic, because they rely on more fundamental ones such as the use of declarative sentences. Indeed it is not clear how one could claim that the use of '*einen schwarzen*' as 'signal' could stand on its own, without presupposing a convention for it, which was already established with help of assertions. In these conditions, can the convention of shaking one's head really be said to be primary, i.e., can it be used as the *ground* for the logical meaning or use of ¬? Lorenzen's idea of founding his game rules on an hypothetical prelogical or even preverbal state looks dangerously like a *hysteron proteron*.

Of course, these remarks do not settle the debate but, as Hodges said, each attempt at answering his question needs its own 'deconstruction'—to avoid the superfluous reference to Derrida: a simple critical examination—and Lorenzen's attempt at a 'foundation' in the life-world, with help of Bühler, does not appear a very promising way to pass that test. Before leaving the issue, a brief remark about logical monism. I have given reasons for being sceptical of Lorenzen's monism, but this is, again, not an argument against his games, only against his philosophy. Indeed, recent work on Lorenzen games has moved into this direction: Shahid Rahman and his collaborators[29] has shown that one could keep, on the one hand, the particle rules invariant and vary the structural rules, on the other, and obtain a formalisation of numerous known logics. This is known in the literature as the 'Dosen principle'. Alternatively, one may simply introduce new connectives, as one does in relevance or in linear logic; the principle here is sometimes known as 'Girard's principle'.[30] The distinction between 'particle' and 'structural' rules thus allows one to generate new logics by systematic variation and combination of both types of rules. One can thus see that, pace Lorenzen, the dialogical approach provides a framework for logical pluralism.[31]

*

[29] See, e.g., the papers collected in Rückert (2007), as well as Rahman and Keiff (2005).

[30] These two principles were framed in Rahman and Rückert (2001), see also Rahman and Keiff (2005, Section 1).

[31] Logical pluralism is already advocated in Rahman and Keiff (2005) and (Rückert, 2001, 2007). For another recent plea for logical pluralism, which is not from the standpoint of game semantics, see Beall and Restall (2005).

As Hodges also said: "If games don't occupy quite the roles that Lorenzen and Hintikka have sometimes claimed for them, then it behoves us to try to find what roles they do occupy" (Hodges, 2001, 25). Consequently, I shall now propose another answer (thus incurring the risk of a 'deconstruction') or at least suggest what a proper answer might look like. I shall get to my answer first by a brief historical detour, taking my cue from some other things Lorenzen said.

In his paper 'Logik and Agon' (Lorenzen, 1960, 187), as well as in a number of other places,[32] Lorenzen referred *en passant* to the practice of refutation or 'dialectics' in Ancient Greece as both the original motivation for the development of logic and as a source for dialogical logic. This suggestion, which looks merely like a rhetorical flourish, was not, as far as I know, followed by the scholarly investigation that it clearly deserves.[33] At all events, one should merely recall here a few facts. The Ancient Greeks had indeed developed a variety of sophisticated forms of question–answer dialogues in medical, legal, political, scientific and philosophical contexts. In philosophy, the Socratic method is well-known from Plato's dialogues, but 'dialectics' was already developed and used by Eleatic philosophers (e.g., Parmenides and Zeno) and the Sophists prior to Socrates. A set of rules was also described later by Aristotle in Book VIII of the *Topics*. In one particularly well-known variant, which fits Zeno's arguments, a designated proponent had to defend a given thesis φ, and the opponent's task was to lead the proponent to admit successively a number of claims $\psi_1, \psi_2, \ldots, \psi_i$ from which one could then force the proponent into an *elenchus*, i.e., to derive $\neg\varphi$ and contradict himself. The point is thus to refute φ by showing that it leads to a contradiction (or to an absurdity, or to a plainly false statement, in other variants). Assuming the principle of non-contradiction, one can devise an indirect proof: for any assertion φ, if propounding $\neg\varphi$ leads to a contradiction, then φ must be true. This method was used and the principle asserted, e.g., by Gorgias, well before Socrates and Plato.

Now, why would such disputes take place? There seems to be a natural answer to this, which is already stated in Aristotle's definition of 'dialectics' in *Topics*, I, § 10 as a dispute concerning assertions not known to be true or necessary—or as they were called, *hypotheses*. Assertions can be made that are directly verified or that are at least verifiable in principle. But in the case of metaphysical-cosmological truths or moral truths, as well as in mathematics, no such direct verification is even *in principle* possible. To take only one basic mathematical example, the observation that there are prime numbers spread

[32] E.g., Kamlah and Lorenzen (1984, 142) and Lorenzen (1987, 78).

[33] Indeed, numerous received views about the origins of logic are to be cast into doubt, and this should provide further grist to Lorenzen's mill. A programmatic presentation will be found in B. Castelnérac and M. Marion, "'Presocratic" philosophy and the Dialogical Origins of Logic', to appear.

throughout the natural number series as far as one can tell leads naturally to the question of their infinity, but the assertion 'there exists an infinity of prime numbers' is not verifiable by sifting through that infinite series, e.g., with help of the sieve of Eratosthenes. But one could ascertain it by use of an indirect proof, as Euclid famously did.[34] In the case of Eleatic philosophy, the situation was rendered even more acute by the proscription of appeals to verification by the senses (probably dictated by the wish to refute Heraclitus); thus, simply walking from one point to another could not count for an Eleatic as a refutation of Zeno's arguments against the reality of motion. (In Lorenzen's terms, Eleatic philosophers would only agree to play 'formal games'.) The point of playing these early forms of logical games was obviously to try and sort out good from bad assertions. If the proponent of φ was publicly driven into an *elenchus*, then φ would be dropped but if he successfully defended it, the result would not merely be that his skills would be admired by all present, it would also entitle them to adopt φ for themselves.[35] So, for example, assertions such as 'the "one" is indivisible', became accepted as true, while the hypothesis that the diagonal of a square is commensurable with its side was found to lead to a contradiction and dropped.

There is a lot more to say here, one could have included as forerunners the mediaeval practice on *obligationes*.[36] One should note that, pace Dingler and Lorenzen, these dialectical games were developed and used by the Greeks in very sophisticated, specialized debating contexts (this is even more obvious for *obligationes*); they cannot be said to have emerged from the life-world.[37] At all events, my point is merely to indicate that Greek dialectics already contain elements of an answer to Hodges' question. These elements can be systematized and given a more general foundation in the philosophy of language using the theory of assertion developed in the chapter on 'Assertion' in Sir Michael Dummetts' *Frege; Philosophy of Language* (Dummett, 1981, Chapter 10) and in Robert Brandom's paper on 'Asserting' (Brandom, 1983). I shall not summarize these here, but simply extract what seems to me the central point of the Dummett–Brandom theory, within the context of this paper. The key idea is that we *act* on assertions and that for this very reason they better be not just true but be backed up with some justification. Of course, some are directly verifiable from the context but the majority of our assertions aren't. However

[34] One should not forget here the wealth of arguments provided by Arpad Szabo in his controversial study *The Beginnings of Greek Mathematics*, devised to support the claim that Greek mathematics "grew out of the more ancient subject of dialectic" (Szabo, 1978, 245).

[35] The democratic nature of these dialogues was first recognized by British radicals in the nineteenth-century, George Grote and Henry Sidgwick. See, e.g., Mill (1991, 50).

[36] See Yrjönsuuri (1994, 11) and the recent study (Duthil Novaes, 2007, Part 3), which does for *obligationes* what should be done for Greek dialectics.

[37] Strangely enough, there are textual indications that Lorenzen also believed this, e.g., at Lorenzen (1987, 85).

that does not mean that assertions cannot be made on no basis whatever, only that they might require justification. According to Dummett:

> we do not of course learn to make statements on no basis whatever, and, if we did, such utterances would not constitute assertions [...], because there would not be such thing as acting on such statements. The process of learning to make assertions, and to understand those of others, involves learning what grounds, short of conclusive grounds, are regarded as justifying the making of an assertion, and learning also the procedure of asking for, and giving, the grounds on which an assertion is made. (Dummett, 1981, 355)

Robert Brandom also views expressing claims as "bringing them into the game of giving and asking for reasons" (Brandom, 2000, 57) and he has extended this analysis of assertions by introducing a distinction between the 'commitments' which a speaker takes on explicitly by making an assertion or by assenting to someone else claim, and the commitments a speaker is 'entitled' to.[38] Thus, according to Brandom,

> In asserting a claim one not only authorizes further assertions, but commits oneself to vindicate the original claim, showing that one is entitled to make it. Failure to defend one's entitlement to an assertion voids its social significance as inferential warrant for further assertions. It is only assertions one is entitled to make that can serve to entitle others to its inferential consequences. Endorsement is empty unless the commitment can be defended. (Brandom, 1983, 641)

Brandom on (Sellars on) Socratic method is also worth quoting in light of my above remarks about Greek dialectics:

> Socratic method is a way of bringing our practices under rational control by expressing them explicitly in a form in which they can be confronted with objections and alternatives, a form in which they can be exhibited as the conclusions of inferences seeking to justify them on the basis of premises advanced as reasons, and as premises in further inferences exploring the consequences of accepting them. [...]. *Expressing* [claims] is bringing them into the game of giving and asking for reasons [...] (Brandom, 2000, 56–57)

The central point of the Dummett–Brandom theory of assertions can thus be stated as follows: an act of asserting a statement brings with it a commitment to defend the assertion, if challenged, so *to make an assertion is to make a move in a game, in which one is asked for and must provide grounds or reasons justifying the making of that assertion*. In other words, the 'game of asking for and giving reasons' is embedded in the very nature of assertions. The notion of 'game' used here can be given precise logical content in terms of the dialogue games first proposed by Lorenzen. My point is thus that *the Dummett–Brandom theory of assertions provides conceptual foundations for*

[38] There are other innovations, such as the introduction of perspectival commitment stores or 'deontic scoreboards', into which we need not go here.

game semantics of the style first laid out by Lorenzen and, conversely, that game semantics can provide a logical precisification of this theory.

Of course, this proposal entails a considerable deviation for the projects of both Dummett and Brandom. Very briefly, in the former case, one may add, on the subject of the differences between Dummett–Prawitz semantics and dialogical semantics,[39] that, in the end, the particle rules of dialogical semantics give only, in Gentzen's terminology, *elimination* rules for the connectives; a wrong-headed approach according to Dummett, who emphasizes, following Gentzen and Prawitz, introduction rules (Dummett, 1981, 362, 1991, 280). As for Brandom, he couches indeed his argument for 'inferentialism' in terms of what he calls 'Dummett's Model' (Brandom, 1984, 116–118, 2000, 61–63), but he wishes in the end to do away with the requisite of 'harmony' (Brandom, 2000, 69–76), which means to do away with the inversion principle or admissibility, that are fundamental to, respectively, Dummett–Prawitz proof-theoretical semantics and to Lorenzen-style dialogical semantics.

Nevertheless, the avenue for this kindred proposal seems open, as, to speak in Brandom's jargon, *it is as a matter of fact through such games that we make our reasons explicit*. Indeed, 'dialogical' semantics can be reformulated in terms of the 'game of giving and asking for reasons', so that 'to attack' becomes 'to ask for reasons' and 'to defend' becomes 'to give reasons'.[40] The point is to win the game, which is the same as succeeding in 'making explicit' reasons for a given assertion. By playing these games against each other, we entitle ourselves to some assertions, as if, for example, we would play chess games in order to find out which claims we are entitled to. But, here 'to make explicit' must mean 'to construct': the ultimate aim is, through reflection on the games thus played, to construct a full justification, i.e., to provide ourselves, whenever possible, with constructive winning strategies.

The idea of providing an answer to Hodges' question along those lines seems so obvious, once stated, that one wonders if anybody had seen it before. A quick search of the literature shows, that Friedrich Kambartel, once member of the Erlangen School, had already made a very similar proposal more than 25 years ago, by providing an account of the particle rules of Lorenzen in terms of games of assertions (here in a paper with Hans-Julius Schneider):[41]

> The need for assertions arises in situations where language competence is developed to the degree that action depends on correctly performed elementary statements, and where the participants do not agree on the correctness of such

[39] About which, see Schroeder-Heister (2007, 2008).

[40] Incidentally, one should notice here that in the Erlangen school attacks are also considered as 'rights' and defences as 'duties' (Lorenz, 1981, 120); we are thus not far from Brandom's normative vocabulary since equivalences obtain between 'right to attack' and 'asking for reasons' on the one hand and 'duty to defend' and 'providing reasons' on the other.

[41] See also Kambartel (1979, 1981).

> a performance. In this case one can either give up common orientation as provided by elementary statements, or one can try to overcome private opinions by reaching a new level of transsubjectivity, by *argumentation*. With argumentation we mean here, quite simply, all attempts to settle differences on the basis of previously or newly established agreements. [...] Someone who now not only just states something, but *asserts* what he is stating, must be prepared to establish by argumentation a transsubjective agreement that his statement has been made correctly. Assertions are, on our everyday and scientific life, one of the language institutions, whereby we can rely on others in our orientations. [...] Trivially the reliability of assertions is undermined if people make assertions without having the corresponding justifications at hand. (Kambartel and Schneider, 1981, 169–170)

So, Kambartel formulated Lorenzen's particle rule in terms of attempting to reach intersubjective agreement about the validity of assertions that are needed for common orientation (Kambartel, 1979, 201–203, 1981, 406–408; Kambartel and Schneider, 1981, Section 7). This corresponds essentially to my proposal.[42] However, would 'assertion games' simply look like Lorenzen games, with another spin? This is a question that will have to be dealt with in another paper.[43]

References

Abramsky, S. (1997). Semantics of interaction: An introduction to game semantics. In Pitts, A. M. and Dybjer, P., editors, *Semantics and Logics of Computation*, pages 1–31. Cambridge University Press, Cambridge.

Abramsky, S. (2006). Socially responsive, environmentally friendly logic. *Acta Philosophica Fennica*, 78:17–45.

Abramsky, S. and Jagadeesan, R. (1994). Games and full completeness for multiplicative linear logic. *Journal of Symbolic Logic*, 59:543–574.

Albert, H. (1985). *Treatise on Critical Reason*. Princeton University Press, Princeton, NJ.

Barth, E. M. and Krabbe, E. C. W. (1982). *From Axiom to Dialogue*. De Gruyter, Berlin.

Beall, J. C. and Restall, G. (2005). *Logical Pluralism*. Clarendon, Oxford.

Blass, A. (1992). A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56: 183–220.

Brandom, R. (1983). Asserting. *Noûs*, 17:637–640.

Brandom, R. (1984). *Making It Explicit*. Harvard University Press, Cambridge, MA.

Brandom, R. (2000). *Articulating Reasons. An Introduction to Inferentialism*. Harvard University Press, Cambridge, MA.

Brandom, R. (2008). *Between Saying and Doing. Towards an Analytic Pragmatism*. Oxford University Press, Oxford.

[42] Papers collected in Kambartel (1989) show that he has, alas, moved away since from such ideas.

[43] I would like to thank Helge Rückert, Peter Schroeder-Heister, Frédérick Tremblay, and Tero Tulenheimo for their comments on an earlier version of this paper, which greatly improved it. (In particular, I owe much more to Helge Rückert's comments than explicitly acknowledged in the footnotes.) It goes without saying that they are not responsible for remaining errors, especially where I stubbornly persisted against their criticisms.

Bubner, R. (1981). *Modern German Philosophy*. Cambridge University Press, Cambridge.

Bühler, K. (1990). *Theory of Language*. John Benjamins, Amsterdam/Philadelphia, PA.

Butts, R. E. and Brown, J. R., editors (1989). *Constructivism and Science. Essays in Recent German Philosophy*. D. Reidel, Dordrecht.

Dingler, H. (1931). *Die Philosophie der Logik und Arithmetik*. Eidos Verlag, Munich.

Dingler, H. (1955). *Die Ergreifung des Wirklichen*. Eidos Verlag, Munich.

Dingler, H. (1964). *Aufbau der Exakten Fundamentalwissenschaften*. Eidos Verlag, Munich.

Dingler, H. (2004). *Hugo Dingler: Gesammelte Werke auf CD-ROM*. Wei, U., editor. Karsten Worm InfoSoftWare, Berlin.

Dummett, M. A. E. (1981). *Frege. Philosophy of Language*. Duckworth, London, second edition.

Dummett, M. A. E. (1991). *The Logical Basis of Metaphysics*. Harvard University Press, Cambridge, MA.

Duthil Novaes, C. (2007) *Formalizing Medieval Logical Theories. Suppositio, Consequentiae and Obligationes*. Springer, Dordrecht.

Felscher, W. (1985). Dialogues, strategies, and intuitionistic provability. *Annals of Pure and Applied Logic*, 28:217–254.

Felscher, W. (1986). Dialogues as a foundation for intuitionistic logic. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic*, volume III, pages 341–372. Kluwer, Dordrecht.

Gethmann, C. F. (1979). *Protologik. Untersuchungen zur formalin Pragmatik von Begründungsdiskursen*. Surhkamp, Frankfurt.

Gethman, C. F. and Siegwart, G. (1994). The constructivism of the "Erlanger Schule": Background, goals, and development. *Cogito*, 8:226–233.

Girard, J.-Y. (1987). Linear logic. *Theoretical Computer Science*, 50:1–102.

Girard, J.-Y. (1999). On the meaning of logical rules. I. syntax vs. semantics. In Berger, U. and Schwichtenberg, H., editors, *Computational Logic*, pages 215–272. Springer, Berlin.

Girard, J.-Y. (2001). Locus solum. *Mathematical Structures in Computer Science*, 11:301–506.

Henkin, L. (1961). Some remarks on infinitely long formulas. In *Infinitistic Methods. Proceedings of the Symposium on the Foundations of Mathematics, Warsaw*, 2–9 September 1959, pages 167–183, Pergamon, Oxford and PWN, Warsaw.

Hilpinen, R. (1982). On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics. *The Monist*, 62:182–189.

Hintikka, J. (1968). Language-games and quantifiers. In Rescher, N., editor, *Studies in Logical Theory*, pages 46–72. Blackwell, Oxford.

Hintikka, J. (1973). *Logic, Language-Games and Information. Kantian Themes in the Philosophy of Logic*. Clarendon, Oxford.

Hintikka, J. (1996). *The Principles of Mathematics Revisited*. Clarendon, Oxford.

Hintikka, J. (1998). *Selected Papers Volume 4. Paradigms for Language Theory and Other Essays*. Kluwer, Dordrecht.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Amsterdam.

Hodges, W. (1997). Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL*, 5:539–563.

Hodges, W. (2001). Dialogue foundations: A sceptical look. In *Proceedings of the Aristotelian Society*, volume Supplementary Volume LXXV, pages 17–32.

Hodges, W. (2004). Logic and games. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Available online at: `http://plato.stanford.edu/archives/win2004/entries/logic-games/`

Hodges, W. (2006). The logic of quantifiers. In Auxier, R. E. and Kahn, L. E., editors, *The Philosophy of Jaakko Hintikka*, pages 521–534. Open Court, Chicago and La Salle IL.

Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology*. Northwestern University Press, Evanston, IL.

Hyland, J. M. E. (1997). Game semantics. In Pitts, A. M. and Dybjer, P., editors, *Semantics and Logics of Computation*, pages 131–184. Cambridge University Press, Cambridge.

Hyland, J. M. E. and Ong, C.-H. L. (2000). On full abstraction for PCF: I, II, and III. *Information and Computation*, 163:285–408.

Janich, P. (1985). *Protophysics of Time*. D. Reidel, Dordrecht.

Japaridze, G. (1997). A constructive game semantics for the language of linear logic. *Annals of Pure and Applied Logic*, 85:87–156.

Japaridze, G. (2003). Introduction to computability logic. *Annals of Pure and Applied Logic*, 123:1–99.

Kambartel, F. (1979). Constructive pragmatics and semantics. In Bäuerle, R., Egli, U., and von Steckow, A., editors, *Semantics from Different Points of View*, pages 195–205. Springer, Berlin.

Kambartel, F. (1981). The pragmatic understanding of language and the argumentative function of logic. In Bouveresse, J. and Parret, H., editors, *Meaning and Understanding*, pages 402–410. De Gruyter, Berlin.

Kambartel, F. (1989). *Philosophie der humanen Welt.* Suhrkamp, Frankfurt.

Kambartel, F. and Schneider, H.-J. (1981). Constructing a pragmatic foundation for semantics. In Fløistad, G., editor, *Contemporary Philosophy. A New Survey*, volume 1, pages 155–178. Martinus Nijhoff, Hague.

Kamlah, W. and Lorenzen, P. (1984). *Logical Propaedeutic. Pre-school of Reasonable Discourse*. University Press of America, Lanham, MD.

Lorenz, K. (1968). Dialogspiele als semantische Grundlagen von Logikkalkülen. *Archiv für mathematische Logik und Grundlagenforschung*, 11:32–55, 73–100.

Lorenz, K. (1981). Dialogical logic. In Marciszewski, W., editor, *Dictionary of Logic as Applied in the Study of Language*, pages 117–125. Martinus Nijhoff, Hague.

Lorenz, K. (2001). Basic Objectives of Dialogue Logic in Historical Perspective. *Synthese*, 127:225–263.

Lorenzen, P. (1955). *Einführung in die operative Logik und Mathematik*. Springer, Berlin.

Lorenzen, P. (1960). Logik und Agon. In *Atti del XII Congresso Internazionale di Filosofia*, volume 4, pages 187–194, Sansoni Editore, Florence.

Lorenzen, P. (1961). Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods. Proceedings of the Symposium on the Foundations of Mathematics. Warsaw*, 2–9 September 1959, pages 193–200, Pergamon, Oxford and PWN, Warsaw.

Lorenzen, P. (1968). *Methodisches Denken*. Suhrkamp, Frankfurt.

Lorenzen, P. (1969). *Normative Logic and Ethics*. Bibliographisches Institut, Mannheim.

Lorenzen, P. (1971). *Differential and Integral. A Constructive Introduction to Classical Analysis*. University of Texas Press, Austin, TX.

Lorenzen, P. (1974). *Konstruktive Wissenschaftsstheorie*. Suhrkamp, Frankfurt.

Lorenzen, P. (1982). Die dialogische Begründung von logikkalkülen. In Barth, E. M. and Martens, J. L., editors, *Argumentation. Approaches to Theory Formation*, pages 23–54. John Benjamins, Amsterdam.

Lorenzen, P. (1987). *Constructive Philosophy*. University of Massachusetts Press, Amherst.

Lorenzen, P. and Lorenz, K. (1978). *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt.

Marion, M. (2006). Hintikka on Wittgenstein: From language-games to game semantics. *Acta Philosophica Fennica*, 78:255–274.

Mill, J. S. (1991). *On Liberty and Other Essays*. Oxford University Press, Oxford.

Mulligan, K. (1997). The essence of language: Wittgenstein's builders and Bühler's bricks. *Revue de Métaphysique et de Morale*, 102:193–216.

Pietarinen, A.-V. (2006). *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication*. Springer, Dordrecht.

Pietarinen, A.-V. and Snellman, L. (2006). On Peirce's late proof of pragmaticism. *Acta Philosophica Fennica*, 78:275–298.

Prawitz, D. (1965). *Natural Deduction. A Proof-Theoretical Study*. Almqvist & Wicksell, Stockholm.

Prawitz, D. (1977). Meaning and proofs: On the conflict between classical and intuitionistic logic. *Theoria*, 43:1–40.

Rahman, S. and Keiff, L. (2005). How to be a dialogician. In Vanderveken, D., editor, *Logic, Thought and Action*, pages 359–408. Springer, Dordrecht.

Rahman, S. and Rückert, H. (2001). Preface. *Synthese*, 127:1–6.

Rückert, H. (2001). Why dialogical logic? In Wansing, H., editor, *Essays on Non-classical Logic*, pages 165–185. World Scientific, Singapore.

Rückert, H. (2007). *Dialogues as Dynamic Framework for Logic*. Doctoral Dissertation, University of Leiden. Available online at: `http://hdl.handle.net/1887/12099`

Sandu, G. and Pietarinen, A.-V. (2003). Informationally independent connectives. In Mints, G. and Muskens, R., editors, *Games, Logic, and Constructive Sets*, pages 23–41. CSLI, Stanford.

Schroeder-Heister, P. (1981). Bibliographie Hugo Dingler (1881–1954). *Zeitschrift für philosophische Forschung*, 35:283–289.

Schroeder-Heister, P. (2007). Lorenzens operative Logik und moderne beweistheoretische Semantik in Mittelstraß, J., editor, *Der Konstruktivismus in der Philosophie im Ausgang von Wilhelm Kamlah und Paul Lorenzen*. Paderborn, Mentis, pages 167–196.

Schroeder-Heister, P. (2008). Lorenzen's operative justification of intuitionistic logic. In van Atten M., Boldini, P., Bourdeau, M., and Heinzmann G., editors, *One Hundred Years of Intuitionism (1907-2007)*. The Cerisy Conference, Basel, Birkhuser, pages 214–240.

Szabo, A. (1978). *The Beginnings of Greek Mathematics*. D. Reidel, Dodrecht.

Tennant, N. (1979). Language games and intuitionism. *Synthese*, 42:297–314.

van Benthem, J. (2003). Logic and game theory: Close encounters of the third kind. In Mints, G. and Muskens, R., editors, *Games, Logic, and Constructive Sets*, pages 1–22. CSLI, Stanford.

van Benthem, J. (2006). The epistemic logic of IF games. In Auxier, R. E. and Kahn, L. E., editors, *The Philosophy of Jaakko Hintikka*, pages 481–513. Open Court, Chicago, LaSalle, IL.

Wittgenstein, L. (1997). *Philosophical Investigations*. Blackwell, Oxford, second, re-issued edition.

Wolters, G. (1985). The first man who almost wholly understands me: Carnap, Dingler, and conventionalism. In Rescher, N., editor, *The Heritage of Logical Positivism*, pages 93–107. University Press of America, Lantham, MD.

Wolters, G. (1991). Dankschön Husserl!—eine Notiz zum Verhältnis von Dingler und Husserl. In Gethmann, C. F., editor, *Lebenswelt und Wissenschaft. Studien zum Verhältnis von Phänomenologie und Wissenschaftstheorie*, pages 13–27. Bouvier, Bonn.

Yrjönsuuri, M. (1994) *Obligationes. 14th Century Logic of Disputational Duties*. Acta Philosophica Fennica, 55.

Chapter 2

ON THE NARROW EPISTEMOLOGY OF GAME-THEORETIC AGENTS

Boudewijn de Bruin*
Faculty of Philosophy, University of Groningen
b.p.de.bruin@rug.nl

Abstract It is argued that game-theoretic explanations of human actions make implausible epistemological assumptions. A logical analysis of game-theoretic explanations shows that they do not conform to the belief-desire framework of action explanation. Epistemic characterization theorems (specifying sufficient conditions for game-theoretic solution concepts to obtain) are argued to be the canonical way to make game theory conform to that framework. The belief formation practices implicit in epistemic characterization theorems, however, disregard all information about players except what can be found in the game itself. Such a practice of belief formation is implausible.

The main claim of this paper is that the epistemological presuppositions non-cooperative game theory makes about players of games are unacceptably narrow. Here, with 'epistemological' I do not intend to refer to the assumptions about the players' beliefs (about the game and about each others' rationality) that may or may not be sufficient to ensure that the outcome of the game satisfies some solution concept.[1] Rather, I use the term 'epistemological' in its philosophical sense to refer to those aspects of the players that have to do with the way in which they use evidence to form beliefs. The claim is then that game theory makes unacceptable assumptions about how players form beliefs about their opponents' prospective choice of action. Here, I do not intend to refer to the assumptions about how the players will or would change their beliefs during the game on the basis of information about the behavior of their

*I am grateful to Johan van Benthem and Martin Stokhof for many inspiring discussions concerning the topic of this paper, and to the participants of the 2004 Prague colloquium on Logic, Games, and Philosophy: Foundational Perspectives for fruitful debate. Thanks, too, to two anonymous referees.

[1]The so-called 'epistemic characterization' results are, for instance, surveyed in (Battigalli and Bonanno, 1999).

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

opponents.[2] Rather, I wish to consider on the basis of what sorts of information the players form their beliefs and their belief revision policies. The claim is then that the evidence that the players are assumed to use to form their beliefs as well as their belief revision policies are of a peculiarly restricted and exclusive kind.

The structure of the paper is the following. First, I present a logical analysis of rational choice-theoretic and game-theoretic explanations of actions. It is then noted that game-theoretic explanations give rise to questions about the role of the beliefs of players in action explanation. I argue that epistemic characterization theorems are the only means to answer these questions adequately. I conclude by showing that it is precisely this kind of theorems that reveal the epistemological problems of game-theoretic agents. Throughout the paper, 'rational choice' theory is the theory of parametric interaction, of 'games against nature', and 'game theory' is the theory of strategic interaction.

Some preliminary logical analysis first. Suppose a rational choice theorist explains the action of some agent as maximizing expected utility. He—or, of course, she—would say something like:

> Agent S maximizes his expected utility in choice situation C by performing action a.

What precisely does he say? Not aspiring completeness of analysis here, I distinguish an 'existential' reading from a 'universal' one. According to the existential reading, the theorist claims the existence of some rational choice-theoretic model D of choice situation C. He claims that agent S was the owner of some utility function u and some probability function P, and that S solved the maximization problem corresponding to u and P by performing action a.[3] Or formally,

$$\begin{aligned}\exists D \exists u \exists \mathsf{P} \quad & (\mathrm{RCT}(D, C) \wedge \\ & \mathrm{Ut}(S, C, u) \wedge \mathrm{VNM}(u) \wedge \\ & \mathrm{Prob}(S, C, \mathsf{P}) \wedge \mathrm{Kolm}(\mathsf{P}) \wedge \\ & \mathrm{Perf}(S, C, a) \wedge \\ & \mathrm{Max}(D, u, \mathsf{P}, a)),\end{aligned}$$

with notation as in Table 2.1. According to the universal reading, no existence claims about models are being made. Only the hypothetical claim is being made that if some rational choice-theoretic model D (with utility function u and probability function P) is a model of C, then action a solves the corresponding maximization problem. Or formally,

[2] The so-called 'belief revision' policies are studied in, for instance, Stalnaker (1996, 1998).

[3] A distinction can be made between available actions and actions the agent knows to be available. But in order for a rational choice-theoretic or game-theoretic model to function properly, these two sets of actions have to coincide. I argued for this claim in de Bruin (2004). Cf., e.g., Hintikka (1996, 214–215).

Table 2.1: Abbreviations

RCT(D, C)	D is a rational choice-theoretic model for C
GT(Γ, C)	Γ is a game that models C
Ut(S, C, u)	u is S's utility function in C
VNM(u)	u satisfies the von Neumann and Morgenstern axioms
Prob(S, C, P)	P is S's expectations in C
Kolm(P)	P satisfies the Kolmogorov axioms
Perf(S, C, a)	S performed a in C
Max(D, u, P, a)	a solves the maximization problem of u and P in D
Nash(Γ, u, a)	a is part of a Nash equilibrium of Γ with utility function u

$$\begin{aligned} \forall D \forall u \forall \mathsf{P} \quad & ((\mathrm{RCT}(D, C) \wedge \\ & \mathrm{Ut}(S, C, u) \wedge \mathrm{VNM}(u) \wedge \\ & \mathrm{Prob}(S, C, \mathsf{P}) \wedge \mathrm{Kolm}(\mathsf{P}) \wedge \\ & \mathrm{Perf}(S, C, a)) \rightarrow \qquad \mathrm{Max}(D, u, \mathsf{P}, a)). \end{aligned}$$

I believe that the universal reading is hardly acceptable as a representation of what a rational choice theorist does in explaining human action. It entails that agents maximize their expected utility even in cases in which they are motivated by completely different kinds of reasons. In cases where they fail to have von Neumann and Morgenstern utilities and Kolmogorov probabilities they make the antecedent vacuously true. In other words, the universal reading makes the explanatory task of the theorist too easy. Although I will disregard the universal reading in the sequel, the arguments presented in this paper would work *mutatis mutandis* for the universal reading as well.

In a similar way we easily obtain a logical analysis of game-theoretic explanations. Suppose a game-theorist describes an action of some agent in some choice situation thus:

> Agent S performs action a in choice situation C according to game theory with the solution concept Nash.

Again an existential and a universal reading can be distinguished, and again the universal reading is too weak to be interesting. According to the existential reading, the game-theorist claims the existence of some game Γ that models C, and of some utility function u of which agent S is the owner. Mentioning the utility function separately is a bit superfluous here, as strictly speaking it is already contained in the game. But I will stick to this redundancy to make the comparison between rational choice theory and game theory more transparent.

Further, apart from the triviality that S really carried out action a, the theorist claims that it was part of a Nash equilibrium of Γ. Or formally,

$$\begin{aligned}\exists\Gamma\exists u \quad & (\mathrm{GT}(\Gamma, C)\wedge \\ & \mathrm{Ut}(S, C, u) \wedge \mathrm{VNM}(u)\wedge \\ & \mathrm{Perf}(S, C, a)\wedge \\ & \mathrm{Nash}(\Gamma, u, a)).\end{aligned}$$

2.1 Game-theoretic agents

Assuming the belief-desire framework of action explanation, a clear difference between rational choice and game-theoretic explanations of action emerges.[4] Rational choice-theoretic explanations provide beliefs and desires. The beliefs are the probability measures P; the desires, the utility functions u. Not so for game-theoretic explanations. Quite clearly we learn something about the desires of the players because the existence of some von Neumann and Morgenstern utility function u is claimed by the game-theorist. But we are not informed about the beliefs of the players. The question is whether this is a problem. Let me survey some possible answers.

(i) The theorist might admit that indeed it would be nice if beliefs could be specified. But, not having the means to accomplish that, we should be happy that in the form of the utility structure we at least have something. This is unacceptable because very often we do have information about beliefs.

(ii) The theorist might say that no extra reasons are needed because the situation he describes is one in which the players blunder into a Nash equilibrium. This is unacceptable unless all game-theoretic aspirations are given up. What would be the function of mentioning the solution concept if it only accidentally fits the outcome?

(iii) One may say that, apart from his utility function, the fact that his action is a Nash action forms a reason for the agent to perform it. But either that is silly, or it is elliptic for the expression of some propositional attitude. It is silly (and unacceptable) if taken literally, because the sole fact that something is a Nash action cannot play a motivational role for the agent. This is so because motivations require propositional attitudes (desires to change the world in certain respects, beliefs about what the world looks like in certain respects). It is elliptic (and acceptable if ellipsis be removed adequately) if one takes it to express that the agent believed that playing Nash is the best way to satisfy his desires, or something like that. But this would have to be made more precise, and I will make it more precise below using epistemic characterization theorems.

[4]Other ways of action explanation would be, for instance, an 'existential phenomenology' à la Merleau-Ponty (see Merleau-Ponty, 1945) or a method based on neurophysiology (cf. Bennett and Hacker, 2003).

(iv) The theorist might claim the existence of some dynamics and refer to some theorem from evolutionary game theory relating this dynamics to the Nash equilibrium. Although more sophisticated, this is again either silly, or elliptic. It is silly (and unacceptable) if taken literally, because the sole fact that some dynamics obtains cannot play a motivational role for the agent as motivations require propositional attitudes. It is elliptic (and still unacceptable) if one takes it to express that the agent believed that this dynamics obtained.

(v) The theorist would claim that evolution programs agents and that therefore no reference to reasons is needed. Agents as automata do not have reasons. They only have 'subpropositional' dispositions. This is consistent. But it is unacceptable if we wish to explain actions in terms of the reasons of the agents. One would have to doubt whether it is still actions that one explains. The difference between actions, reflexes, and so forth blurs.

(vi) The theorist could simply pick some probability distribution over the actions of S's opponents and make sure that action a maximizes expected utility (where the utility function is the one of which existence is being claimed). This is unacceptable because it is entirely ad hoc. All game-theoretic aspirations would be given up because the mentioning of an ad hoc probability distribution would not show why the Nash equilibrium (rather than another solution concept) figured in the explanation.

(vii) The theorist could pick some probability distribution over the actions of S's opponents and make sure that action a maximizes expected utility. If in addition the theorist could show that this probability distribution is not ad hoc, he would have given additional reasons for S's performing a in C. This is acceptable, but it has to be made more precise, and I will make it more precise below using epistemic characterization theorems.

(viii) The idea now is that a game-theorist who explains some action as an action that is part of a Nash equilibrium makes implicit reference to some beliefs of the players that are not ad hoc. How to avoid being ad hoc? By requiring some structural relation between the solution concept and the implicit beliefs.

A very obvious candidate for such a structural relation is presented by epistemic characterization theorems. They provide sufficient conditions for a solution concept to obtain. Epistemic characterization theorems are implications. The antecedent specifies conditions on the beliefs and desires of the players; the consequent states some conditions about the actions and the solution concept. For instance, the epistemic characterization of the Nash equilibrium is:

> If (i) all players are rational, (ii) all players know their own utility function, and (iii) all players know what their opponents are going to play, then they play a Nash equilibrium.

Another well known example characterizes common knowledge of rationality and utility structure as sufficient for iterated strict dominance (in two-person normal form games) and backward induction (in any extensive form game).[5]

The usual way to think about such theorems is that they can be used by the game-theorist to justify the use of some solution concept in a specific explanation. The theorist may defend, for example, his use of the concept of iterated strict dominance to explain the behavior of some agent by stating that among other things common knowledge of rationality and utility structure obtains in the choice situation. What I suggest is that game-theoretic explanations should be read in general as making such claims. Whenever a game-theorist uses some solution concept, he should be taken to make the claim that the epistemic conditions of the corresponding characterization theorem obtain. My argument is that there is no alternative way to distill from a game-theoretic explanation the right kind of reasons for the agents in a uniform way. All alternatives I discussed have some problems. Either something is wrong with the motivational force of the alleged reasons (the issue about the propositional attitude), or they are ad hoc and fail to account for the necessity of using a solution concept in the first place.

It is important to point out that it is feasible or consistent to require to use epistemic characterization results as the canonical way to the beliefs of the players. The most elegant way uses characterization theorems that are specified in terms of Stalnaker's 'game models.'[6] To use the example of iterated strict dominance, it cannot only be shown that common knowledge of rationality implies iterated strict dominance, but also that every outcome of iterated strict dominance of any game can result in a situation of common knowledge of rationality. In other words, given an arbitrary outcome of iterated strict dominance, the 'game models' approach enables us to sketch an epistemic setting in which (i) rationality and utility structure are common knowledge, and (ii) this very outcome is played. It is this converse direction that shows that the game-theorist is not committed to something infeasible. If he explains an action as, for instance, iteratively undominated, there is indeed a game playing situation in which there is common knowledge of rationality. To take the epistemic characterization results as the suppliers of beliefs then is a coherent assumption.

To sum up, the first claim is that the epistemic characterization theorems are the canonical way to the beliefs of the players in game-theoretic action explanation.

---

[5]The epistemic characterization of the Nash equilibrium is due to Aumann and Brandenburger (1995). Iterated strict dominance was dealt with, in various degrees of formality, by Bernheim (1984), Pearce (1984), and Spohn (1982). The *locus classicus* for backward induction is Aumann (1995). I am a bit sloppy in using common knowledge instead of common belief. See de Bruin (2004) for details.

[6]First introduced in Stalnaker (1996).

2.2 Epistemology

Game-theorists explain actions in terms of reasons. As reasons for some action of some agent they give von Neumann and Morgenstern utility functions and Kolmogorov probability measures. The former are given explicitly in the game-theoretic representation of the agent's choice situation. The latter are implicit, but can be obtained via epistemic characterization results. The logical analysis of game-theoretic explanations contains the element $\mathrm{Ut}(S, C, u)$ requiring that the utility function of which existence is claimed is in fact S's utility function in choice situation C; he has to be the owner of u so to speak. Of course, the same has to be true of the probability distribution implicitly referred to, and, of course, the action that has to be explained has to be a solution to the corresponding maximization problem. The agent has to maximize expected utility. Game-theoretic explanations and rational choice-theoretic explanations do not seem to be different then in principle. Both involve utility and probability, both involve maximization, and both involve the claims that the utility is the agent's utility, that the probability is the agent's probability, and that the agent solves a maximization problem. But these similarities are very deceptive.

Rational choice theorists and game-theorists, although they do need to bother addressing the ownership issue of the utility functions (the claim $\mathrm{Ut}(S, C, u)$, that is), do not need to explain such things as why S has the preferences he has (for instance by referring to S's education, or bourgeois background) or whether they are reasonable or not. Now compare in the same way rational choice theorists and game-theorists with respect to the agent's beliefs P. Rational choice theorists as well as game-theorists need to bother addressing the ownership of the probabilistic expectations (the claim $\mathrm{Prob}(S, C, \mathsf{P})$). Rational choice theorists do not need to explain such things as why S has the beliefs he has (for instance by referring to S's practices of belief formation, his critical or narrow mind) or whether these beliefs are reasonable. But game-theorists do need to bother thinking about these questions. In fact, referring to epistemic characterization results to explain the very probability measure P is to answer these questions. The epistemic characterization results say that S has formed his beliefs on the basis of inspection of the game structure and on the basis of rationality considerations. And on the basis of nothing else. Incidentally, this idea can be traced back to von Neumann and Morgenstern's *Theory of Games and Economic Behavior*:

> Every participant can determine the variables which describe his own actions but not those of the others. Nevertheless those 'alien' variables cannot, from his point of view, be described by statistical assumptions. This is because the others are guided, just as he himself, by rational principles.[7]

[7] von Neumann and Morgenstern (1944, 11).

To sum up, the second claim is that whenever the game theorist explains the behavior of some agent in truly game-theoretic terms, he is implicitly committed to the view that the agent, to form the beliefs necessary for his strategic deliberation, disregards all available information except what involves the game structure and the rationality of the players. Epistemic characterization theorems make this explicit.[8]

2.3 Narrowness

Yet the practice of discarding so much available information is an implausible, or simply bad, way of belief formation. It forbids players a large spectrum of possible evidence to base their beliefs on. It is not adequate as a description of how actual human beings reason, and it is even more inadequate as a theory of knowledge or scientific methodology. I will structure the argument by distinguishing these two cases, real and ideal agents.

The game-theorist's call to allow only truly game-theoretic information instead of exogenous or statistical data clearly puts a restriction on the possible sources of evidence players are allowed to invoke as reasons for their beliefs. The appeal of this call can be explained by looking at the abstract character of game-theoretic modeling. Indeed, if we abstract away from everything except the number of players, their possible actions, and their preferences, then there is not much exogenous or statistical information to be found. The game-theorist will not deny that the strategic choice situations he tries to model are concrete, and that real agents can actually use a large spectrum of concrete data for belief formation. Of course, all strategic choice situations game theory is concerned with are concrete and all game-theoretic models abstract. Abstraction is what happens everywhere in science, but nowhere in science too high a level of abstraction is good. The above logical analysis of game-theoretic explanations and the considerations about the specification of reasons for actions allow us to make precise statements about what it exactly is that gets lost in abstraction. By abstracting away from everything except the possible actions, the preferences, and the number of players, a game-theoretic model leaves unmodeled much of the evidence or data or information that real players will actually use to form beliefs about their opponents. This would not be a problem if, for instance, the origin of the beliefs did not matter. But, as we have seen, the origin of the beliefs matters crucially because without specification

[8] The epistemic characterization of the Nash equilibrium allows for exogenous information. The antecedent requires knowledge of the actions of the opponents, and it is not excluded that this knowledge is formed on the basis of, for instance, statistical information. But this observation does not save the Nash equilibrium, because it is now only right to use the solution concept in cases in which the players have knowledge, as opposed to mere and possibly mistaken belief, about their opponents. And the role of knowledge, as opposed to belief, in explanations of actions is highly disputed. See, for more details, de Bruin (2004).

of the origin of the beliefs in terms of the epistemic characterization results game-theoretic explanations of human actions do not provide the reasons for actions in a systematic manner.

Let us briefly take stock. I started from the assumption that game-theoretic explanations have to conform to the belief-desire framework. The desires are clear. What about the beliefs? There are many ways to sneak in beliefs, but I showed that only via epistemic characterization results a systematic commitment to particular beliefs can be obtained. That is, I have argued that if you start from the assumption that game-theoretic explanations have to conform to the belief-desire framework, then there is no way, except by using epistemic characterization results, to get the beliefs the theorist is committed to ascribe to the players. The point now is that this entails a very specific origin of the beliefs. For instance for iterated strict dominance: common knowledge of rationality and nothing else.

As long as it stays within the realm of the game-theoretical, the specification of the origin of the beliefs can only be phrased in terms of those aspects of the strategic choice situation that survive the abstraction process. Clearly this results in a distorted model of belief formation. Whereas there is no hope, then, of dealing with belief formation in game theory in a way that does justice to the concreteness of the evidence, it could still be the case that what game theory assumes about belief formation is plausible from the perspective of some theory of knowledge (for 'ideal' epistemic agents, so to speak). In fact, a rather strange sort of theory would be the result: an interpretation of game theory as descriptive (*ex post* or *ex ante*) of the actions of agents, but as prescriptive about their beliefs. But this sounds too far fetched.

A feature that distinguishes knowledge from belief is that knowledge is necessarily true, and belief not. Another, that knowledge meets very high evidential standards, and belief not. This is the point of a hierarchy of 'Gettier examples', but not dependent on such examples.[9] This does not mean, however, that anyone can believe anything without further qualifications. Senseless beliefs are no beliefs. If you say that you believe something, then you have to be able to give an answer to the question why you do so. In general people will try to answer such a question by presenting the interlocutor with what they think is good evidence for the belief. All in all, beliefs need reasons.

Applied to game theory, how should players (players who are ideal from the view point of some theory of knowledge) form beliefs? They should try to inspect their strategic choice situation in the most penetrating way possible; in particular, they should try to get as much information about their opponents as possible. They should be interested to hear something about the tradition in

[9]Gettier (1963); and many articles along similar lines.

which their opponents were raised or the training they have had. They should try to determine the reliability of hearsay evidence and reported observations, and to sort out how to weigh such evidence in relation to their own observations. If available, they should attempt to interpret statistical surveys and consider other available exogenous data, and determine their relevance for their purposes. And, of course, they should try to find out as much as possible about the way their opponents try to form their beliefs. One thing, however, they should not do: to disregard possible sources of information, to eschew statistical or exogenous data, to avoid going beyond what is immediate in the situation, to be narrow-minded and uncritical.

To sum up, the third claim is that by denying players access to any information except what is immediate from the game structure, game theory puts forward an epistemological claim that is inadequate as a description of real human beings, and implausible as a theory for epistemologically ideal agents.

References

Aumann, R. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19.

Aumann, R. and Brandenburger, A. (1995). Epistemic conditions for nash equilibrium. *Econometrica*, 63:1161–1180.

Battigalli, P. and Bonanno, G. (1999). Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53:149–225.

Bennett, M. and Hacker, P. (2003). *Philosophical Foundations of Neuroscience*. Blackwell, Malden, MA.

Bernheim, B. (1984). Rationalizable strategic behavior. *Econometrica*, 52:1007–1028.

de Bruin, B. (2004). *Explaining Games: On the Logic of Game Theoretic Explanations*. Ph.D. thesis, University of Amsterdam, Amsterdam.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23:121–123.

Hintikka, J. (1996). *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Librairie Gallimard, Paris.

Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52:1029–1050.

Spohn, W. (1982). How to make sense of game theory. In Balzer, W., Spohn, W., and Stegmüller, W., editors, *Studies in Contemporary Economics*, volume 2: Philosophy of Economics, pages 239–270. Springer, Berlin.

Stalnaker, R. (1996). Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12:133–163.

Stalnaker, R. (1998). Belief revision in games: Forward and backward induction. *Mathematical Social Sciences*, 36:31–56.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.

Chapter 3

INTERPRETATION, COORDINATION AND CONFORMITY

Hykel Hosni
Scuola Normale Superiore, Pisa
hykel.hosni@sns.it

Abstract The aim of this paper is to investigate a very general problem of (radical) interpretation in terms of a simple coordination game: the *conformity game*. We show how, within our mathematical framework, the solution concept for the conformity game does indeed provide an algorithmic procedure facilitating *triangulation*, in the sense of Davidson.

3.1 Introduction

Suppose that the robotic rovers *I* and *II* are conducting a joint operation on a terrain about which nothing was known to their designer (say the units are operating on Mars). Suppose further that communication among the units has been lost and that the only way *I* and *II* have to restore it is to meet on some location $\mathtt{l}$, chosen from a finite set of possibilities equally accessible to both. Assuming that any location is as good as any other, provided that *I* and *II* agree on it, how could the robots reason and act so as to facilitate their meeting? That is, how should they *choose* $\mathtt{l}$?

We see situations of this sort as instantiations of interpretation problems. After all, what *I* and *II* must do in order to restore communication is to (i) attach a certain meaning to the representation they have of their environment, (ii) form expectations about each other's behaviour, and (iii) act accordingly. More specifically, once the possible locations, say $\mathtt{l}_1, \ldots, \mathtt{l}_k$, are identified, given their common intention, agents must interpret each other relative to the 'external world'—the environment in which they happen to operate—so as to increase their chances of agreeing on the final choice of a location. Since *I* and *II* do not share a language, in fact they cannot communicate, the problem they face is one of *radical* interpretation.

At the same time, this situation is a clear example of strategic interaction: what corresponds to the 'rational' or 'commonsensical' or even 'logical' or

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

simply 'best' course of action for *I* depends on the course of action adopted by *II* (and the other way round). This quite naturally suggests that game theory might somehow provide us with precise and well-understood guidelines for the mathematical solution of our problem. As will be shortly illustrated, however, for the kind of strategic interaction that we shall be concerned with, the classical solution concepts studied in the theory of non-cooperative games are of no use whatsoever.

The framework of Rationality-as-Conformity, recently introduced by Jeff Paris and the present author (see Hosni and Paris, 2005; Hosni, 2005), attempts to define, within an abstract mathematical setting, 'rationality' in situations of strategic interaction of the sort mentioned above. It is the purpose of this paper to illustrate how such a mathematical characterization of rationality can be used to provide a solution concept for problems of (radical) interpretation, whenever the latter is considered in terms of games of (pure) coordination.

The paper is organized as follows. First (Sections 3.1.2–3.1.4), we isolate the fundamental aspects of radical interpretation problems in connection with the interactive choice problem considered in the Rationality-as-Conformity framework. Putting forward the intrinsic strategic nature of the problem of radical interpretation leads us to formulate it mathematically in terms of the *conformity game*, fully described in Section 3.2. Being a game of multiple (indiscernible) Nash-equilibria, the conformity game is indeed a (pure) *coordination game* and as such, it is generally regarded to be unsolvable within the traditional game-theoretical framework of non-cooperative games. We discuss in Section 3.2.1 the informal constraints that an adequate solution concept for the conformity game should satisfy and move on towards formalising the solution concept for the conformity game in Section 3.3. This is based on the Minimum Ambiguity Reason, introduced in Honsi and Paris (2005) as part of the Rationality-as-Conformity framework. We will then conclude by showing that this solution concept indeed provides an algorithmic solution for establishing communication—triangulating—in problems of radical interpretation.

Radical interpretation helps in clarifying the issues and the assumptions underlying a basic characterization of 'rationality' in communicationless scenarios yet without immediately providing any effective procedure to achieve it. Pure coordination games, on the other hand, help framing a variety of possible solution concepts based on *saliency*, which however seem to lack of a general formal structure allowing us to evaluate their 'rational' underpinnings. This paper attempts to unify the fundamental aspects of both frameworks by means of the mathematical abstraction provided by Rationality-as-Conformity.

Many connections between (linguistic) interpretation and (coordination) games have been explored, from the classic investigation by Lewis (1969) to the game-theoretic accounts of linguistic interpretation of Parikh (2000) and van Rooy (2004). Though Lewis considers the 'use of language' as a particular

kind of 'coordination problem' (Lewis, 1969), the present author has no knowledge of any attempt to relate mathematically the structure of *pure coordination* games with that of *radical* interpretation.

3.1.1 Why rationality-as-conformity?

As illustrated at length in Honsi and Paris (2005), we understand 'conformity' as the adoption of a choice process facilitating the selection of the same possible world (say a location 1 in the robotic rover example above) as another like-minded yet otherwise inaccessible agent.

Within frameworks of this sort, solid arguments can be put forward supporting the view that commonsensical agents not only happen to be generally able to conform, they should indeed aim at conforming if they are to be rational.

1. *The members of a society have a natural inclination to coordinate successfully*. This is a conclusion of the numerous empirical investigations that have been carried out during the last decades in the area of *behavioural game theory*, following Schelling's early intuitions about *coordination games* (Schelling, 1960) (see e.g. Mehta et al., 1994; Camerer, 2003). The common pattern of those investigations puts forward that, whenever, say, pairs of agents face a strategic choice problem in which they have a joint motivation (intention) to coordinate their solutions, they will be able to adopt certain kinds of choice processes facilitating this coordination. In other words, there are reasons to believe that principles, strategies and patterns of choice behaviour exist which, if adhered to, will result in agents having generally better chances to coordinate (and never strictly worse) as they would have, should they adopt random patterns of behaviour.

2. *Agents satisfying probabilistic 'commonsense' should end up assigning similar degrees of belief*. This is a consequence of a number of a contributions in the area of subjective probability logic. In the normative framework developed by Paris and Vencovská (1990, 2001) and Paris (1994) a small number of so-called commonsense principles are identified and it is shown that, if adhered to, those principles uniquely and completely determine any further assignment of probabilities, i.e. degrees of belief. This distribution of probabilities, the one with the largest possible entropy, is provably the only one jointly consistent with the (probabilistic) knowledge possessed by an agent and those principles. Hence, similar agents, possessing similar knowledge bases and applying the inference process identified with commonsense, all assign similar degrees of belief to the as yet undecided sentences.

3. *'Rationality is a social trait. Only communicators have it.'* This is the conclusion of Davidson (2001). The idea here is that a necessary condition for rationality is an adequate apparatus for communication, which in turn requires agents to be able to move from a condition of mutual inaccessibility (no shared language), to a condition in which communication is being enabled. This transition implies that agents are attaching similar meanings to the publicly accessible causes of their reciprocal choice behaviour. This aspect of Rationality-as-Conformity, which Donald Davidson calls *triangulation*, and its underlying structure are the main topic at focus in the rest of this paper.

3.1.2 Radical translation and the Principle of Charity

Put roughly, a problem of *radical translation* is one in which one agent—a linguist in the field—is trying to build up a 'translation manual' accounting for the utterances of a native speaker of a language about which the linguist has no knowledge whatsoever. This complete lack of information, together with the fact that the two agents are assumed not to share a third language, make the translation problem *radical*.

The radicalness of the situation induces Quine to observe that a hypothetical theory of radical translation should start by relating the native's linguistic behaviour to the one the translator would adopt, were she to be in the same 'observable situation' as the native.

In his classic example Quine, who was the first to introduce this problem in connection with the translation of logical constants (Quine, 1960, 2), imagines that the native speaker utters the expression GAVAGAI in correspondence of a rabbit passing by, causing—possibly on repetitions of similar events—the translator to conjecture that GAVAGAI translates into 'rabbit'.

There are many subtleties connected with this example, none of which being of particular interest for present purposes. Rather, two issues involved in the radical translation exercise are relevant for our present discussion:

1. What is it, if anything, that *justifies* (epistemologically) the translator in the above conjecture?

2. How far can the translator go in relying on this conjecture?

Those questions are clearly not unrelated. The former calls for the observation that a linguist may just introspect and conclude that "as a native speaker of English, I would utter RABBIT were that kind of animal to pass by". This subjunctive is clearly grounded on the assumption that the linguist and the native speaker, though lacking of a shared language, are nonetheless *like-minded* individuals and hence inclined to adopt similar linguistic behaviours under similar

(observable or conceivable) circumstances. Elevated to the status of a normative maxim, this is known as the *Principle of Charity*.

Any reasonable understanding of this principle, of course, asks for a clarification of what is meant by 'similar linguistic behaviour' as well as 'similar observable (conceivable) circumstances' and in the natural language case these are by no means trivial clarifications to do and many criticisms to the adoption of the principle seem to pivot on this difficulty (see e.g. Feldman, 1998; Wachbroit, 1987; McGinn, 1977 for the role of the principle in the explanation of rationality, and Nozick, 1993, 152–158; Glock, 2003, 194–199 for more forceful criticisms). It turns out, however, that in the abstract and simplified mathematical framework of Rationality-as-Conformity, correlated notions can be defined rigorously and put to work in the formal characterisation of rational choice behaviour in the absence of communication or learnt conventions.

The second crucial feature of radical translation problems relates to their fundamental indeterminacy. Quine argues that there cannot be a unique translation manual which the linguist in the field may be able to construct. Rather, there must be a plurality of manuals, all equally acceptable, that is to say, equally supported by the available evidence. The only attempt that the linguist can do to reduce this indeterminacy is the application of the Principle of Charity, leading her to *discard* all those possible translation choices that will make the native utterances systematically wrong (or incoherent), by the translator's lights. After this 'rational' refinement, the choice of a translation manual may simply be underdetermined by the empirical evidence available to the translator.

That 'rationality' might not always lead to a unique choice (without randomisation) is a feature captured by the Rationality-of-Conformity framework as well. Indeed, some problems might just be too hard to admit of a unique solution.

3.1.3 Radical interpretation and triangulation

The issues of radical translation and charity are taken a step further by Davidson's investigations on *radical interpretation*. For the purposes of the present discussion, the main points of departure of the situation described in the radical interpretation problem with respect to the one discussed in connection with radical translation can be outlined as follows. Davidson does not assume that agents are native speakers of distinct languages. He rather assumes that they do not have a shared language whatsoever and that their goal consists in establishing communication.

The Principle of Charity is thus sharpened and indeed assumed to be a necessary condition for the manifestation of rational behaviour tout court. Moreover, the interpretation problem is grounded on a fundamental symmetry which need not hold in the translation case, that is that both agents share a common

intention to communicate: the interpreter wants to understand the interpretee who, in turn, wants to be understood by the interpreter.

Differences in the formulation of the problem lead to differences in the proposed solutions. Quine's major problem is that of locating the common cause of the linguistic behaviour, which he identifies in the 'stimulus-meaning'. Davidson overcomes many of the difficulties related to this concept by introducing the metaphor of *triangulation*. While Davidson takes charity as a presumption of rationality upon which the possibility of interpretation and mutual understanding themselves rest, he acknowledges that it can only provide a 'negative' contribution, namely by guiding the interpreter towards *discarding* possible interpretations which would systematically make the interpretee wrong or incoherent to her own lights. Triangulation, on the other hand, is the recognition that the similarities observed in each other's linguistic behaviour find their common cause in the same portion of the external environment shared by the agents. It is the location of those causes that results in getting a clue about the other's meanings.

Davidson introduces triangulation by considering a 'primitive learning situation', in which a child learns to associate the expression "table" to the actual presence of a table in a room. The way the child can learn to do so, relies in her ability to generalise, to discover and exploit similarities among situations. Sharing similar generalisation patterns is what makes the child's response to the presence of a table—the utterance of the word "table"—meaningful to us. This is the rational structure that agents must have in order for communication to start.

> The child finds tables similar; we find tables similar; and we find the child's responses in the presence of tables similar. It now makes sense for us to call the responses of the child responses to tables. Given these three patterns of response we can assign a location to the stimuli that elicit the child's responses. The relevant stimuli are the objects or events we naturally find similar (tables) which are correlated with responses of the child we find similar. It is a form of triangulation: one line goes from the child in the direction of the table, one line goes from us in the direction of the table, and the third line goes between us and the child. Where the lines from child to table and us to table converge, 'the' stimulus is located. Given our view of child and world, we can pick out 'the' cause of the child's responses. It is the common cause of our response and the child's response. (Davidson, 2001, 119)

A fundamental aspect of the triangulation process, then, consists in the recognition of the role played by constraints imposed by the 'external world' on the interpretational choices. In particular, the interpreter should ascribe 'obvious beliefs' (e.g., the presence of a table) to the interpretee, and project onto her the likewise 'obvious' consequences (that she will behave accordingly). Suppose, for instance, that rover *I* in the initial example perceives the presence of a perfectly round crater. According to this way of reasoning, *I* should expect *II*

to be able to perceive the crater as a perfectly round one. At the same time *II* should expect *I* to expect that *II* itself would perceive the crater as a perfectly round one etc., and of course consider this as a relevant feature for the selection of the rendez-vous location 1. This 'like-mindedness' or 'common reasoning' of agents plays a fundamental role in the Rationality-as-Conformity framework and constitutes the main conceptual fulcrum on which the present analysis of interpretation, coordination and conformity pivots.

As for translation, in the case of interpreting natural language triangulation presents several difficulties mostly related to the rigorous explanation of what intervenes in the 'recognition of the common causes' of common linguistic behaviour. A recent comprehensive discussion on the topic can be found in Glock (2003). What is relevant for us here, however, is that the complication of considering the full case of interpreting natural language is surely one of the reasons why the theory of radical interpretation does not seem to allow for a clear-cut *procedure* by means of which agents can achieve, or at least facilitate, triangulation.

Within the mathematical framework of Rationality-as-Conformity we are able to provide one such effective procedure. It goes without saying that the structure therein considered (comparable to unary predicate languages) is much weaker than the one required by Davidson for the construction of a theory of meaning, namely the full first-order logic with equality. Our hope is, of course, that of eventually extending the results obtained in this initial framework to cover more 'realistic' situations.

3.1.4 Radical interpretation as coordination

Thomas Schelling is usually credited with the introduction of *coordination* problems in the game-theoretical literature. Roughly speaking, a tacit coordination game is a situation of interdependent, strategic choice characterised by the absence of communication among players who nonetheless aim at performing the same choice—i.e., coordinating. Schelling's example concerns a couple who get accidentally separated in a supermarket and want to rejoin.

Schelling calls this a problem of 'tacit coordination' with 'common interests' and notices that given the lack of communication—which indeed makes the coordination *tacit*—all that agents can rely on are the assumption of like-mindedness and the mutual expectations that this generates. What Schelling intends to discuss is the characterisation of 'rational rules' accounting for the ability humans have to coordinate in the complete absence of communication.

The situation described by Schelling is one of radical interpretation for which a triangulation-like solution is advocated. Indeed, after introducing the supermarket problem he goes on commenting as follows:

> What is necessary is to coordinate predictions, to read the same message in the common situation, to identify the one course of action that their expectation of

> each other can converge on. They must 'mutually recognize' some unique signal that coordinates their expectations of each other. We cannot be sure that they will meet, nor would all couples read the same signal; but the chances are certainly a great deal better than if they pursued a random course of search. (Schelling, 1960, 54)

The analogies with the solution proposed by Davidson for the radical interpretation problem stand out: both charity and triangulation appear clearly in Schelling's illustration of the fundamental features of the solution concepts adequate for tacit coordination games. Entirely analogous remarks can be made in relation to 'tacit agreement' as discussed by Lewis in his classic work on conventions (Lewis, 1969).

3.1.5 Towards a solution concept

What facilitates conformity in coordination problems of the sort introduced above is, according to the investigations initiated by Schelling, the selection of those possible options—strategies—that would be perceived by agents as *focal points*. Indeed, the many investigations that followed Schelling's original intuitions can be seen as attempts at providing an explanation for the ability that human agents have in exploiting focal points for the purpose of coordinating.

The intuition underlying the use of focal points is that these correspond to strategies which enjoy some degree of 'saliency' or 'conspicuousness', in Schelling's phraseology, which will lead agents to in fact focus on certain options instead of others. Distinctions are made then, on what saliency can be taken to be (see, e.g. Sugden, 1995; Kraus et al., 2000). For present purposes we will concentrate on salience as given by the identification of a *choice process* which an agent might adopt upon reflection about which choice process another like-minded agent with a common intention to coordinate might herself adopt. In the literature this is usually referred to as *Schelling's salience*.

The most distinctive feature of salience is the combination of *uniqueness* and *obviousness* of focal points. These are thought of as options which somehow *stand out* when considered in the context of the strategies available to the agents in a given coordination problem. So, for example, the robotic rovers of our initial example will base their choice on saliency if they will select a location $\mathtt{l}$ which stands out in the set $\{\mathtt{l}_1, \ldots, \mathtt{l}_k\}$. Naturally, if I can conclude that the location $\mathtt{l}_j$ does indeed stand out, the fact that II intends to conform to the choice it expects I to make will lead, together with the assumption that I and II are like-minded, to the conclusion that $\mathtt{l}_j$ is the *obvious* choice for this problem.

It is in this spirit that Schelling suggests that, in order for agents to coordinate successfully, they must 'mutually recognize a unique signal'. Intuitive as it may be, however, a lighthearted resort to 'uniqueness' can prove to be rather tricky. As it has been put forward by (Kraus et al., 2000), this becomes

a major concern once we take into account the limitations (i.e., bounded reasoning capabilities) of the agents. Moreover, there could be circumstances in which appeal to uniqueness may lead to undesirable conclusions, as we will have occasion to notice below.

In what follows, we will rather attempt at formalizing the notion of a focal point by characterising saliency in terms of the *minimisation of the ambiguity* of the options available to the agents. In order to do this we shall firstly provide a mathematical formalisation of the *context* within which focal points are to be discerned. This will enable us to study the corresponding *reasoning process*, that is to say an algorithm for the determination of the minimally ambiguous strategies within the context.

3.2 The conformity game

In the spirit of the Rationality-as-Conformity approach, we tackle the knowledge representation issue by considering the simple model in which options are the *possible worlds* generated by mapping a finite set A to the binary set $2 = \{0, 1\}$. Nothing else is assumed about the structure of the set A.

The domain of the game is $\wp^+(2^A)$, the set of non-empty subsets of 2^A which denotes the set of all possible worlds. We attach to elements $K \in \wp^+(2^A)$ an epistemic value, namely we take players to have common knowledge of the fact that the options they have to choose from are those in K, which includes the possible world which will be eventually selected. Intuitively, then, the the cardinality of K gives a quantitative measure of the agents' uncertainty about the other's actual choice.

The *conformity game* is a two-person, non-cooperative game whose normal form goes like this: Each player is to choose one strategy out of a set of possible choices, identical for both agents up to permutations of A and 2, where each strategy corresponds to one element of $K = \{s_1, \ldots, s_k\}$, say. Strategies are therefore represented in this game as finite binary strings. Players get a positive payoff p if they play the same strategy, and nothing otherwise, all this being common knowledge. (Figure 3.1 represents the conformity game for $k = 3$.)

		Player *II*		
		s_1	s_2	s_3
Player *I*	s_1	p, p	$0, 0$	$0, 0$
	s_2	$0, 0$	p, p	$0, 0$
	s_3	$0, 0$	$0, 0$	p, p

Figure 3.1: The conformity game

Note that, for present purposes, we limit ourselves to the case in which each identical pair of strategies yields a unique positive payoff p, so that any point in

the diagonal would be as good as any other as far as the agents are concerned: all that matters is that they conform on their world-view.

Being a game of multiple Nash-equilibria in which the players are assumed to be inaccessible to each other, the conformity game is a typical example of a (pure) *coordination game*, a kind of game which is generally considered to be unsolvable within the traditional theory of non-cooperative games. (See, e.g., Camerer, 2003 for a discussion on coordination problems other than 'pure'.)

Before going into any further details of the conformity game it will be useful to introduce some ideas concerning the selection of multiple Nash-equilibria in pure coordination games, and relate these to the intuitions underlying the conformity game.

3.2.1 Multiple Nash-equilibria and the conformity game

Traditional game theoretic solution concepts usually characterize distinguishability among options (strategies) in terms of the comparison of (ordinal) utilities, 'rationality' being defined in terms of utility maximization. As an immediate consequence of this, whenever options are perceived by an agent as being equally desirable—i.e., payoff-indistinguishable—the selection of strategies usually referred to as 'rational' turns out to be unhelpful as solution concept.

Here is where the concept of 'rationality' pursued in the Rationality-as-Conformity framework shows its most relevant point of departure from the game theoretic tradition. In the former, in fact, rationality is not defined in terms of maximisation of utility, but on the mutual expectations of agents sharing a common intention. Hence the conformity game is characterized by a complete symmetry with respect to both payoffs and players. Moreover, the possibility of considering 'extra structure' in the game by focusing on its presentation can be ruled out by means of appropriate mathematical devices, to be shortly introduced. Hence, in Schelling's terminology, the conformity game is a 'clueless', 'genius-proof' game.

To appreciate the point further, recall that the typical solution concept for non-cooperative games introduces a notion of distinguishability among strategy profiles—Nash-equilibrium—which is in fact weaker than simple pay-off dominance. If a Nash-equilibrium exists, yet is not unique, then a natural way of reducing the situation to the standard case would just involve selecting the equilibrium, if one exists, with the the highest possible payoff. In particular, it can happen that a strategic game admits of say two equilibria with distinct ordinal utilities, which nonetheless are, according to the theory of Nash-equilibrium, undistinguishable. Due to its wide applicability, a largely studied example is the following variant of the game known in the literature as the

Battle of the sexes (see, e.g., Osborne, 2004). Two players are to choose between a pair of options for a night at the concert hall (say, B and S, for Bach and Stravinsky) with the distinctive feature that whilst both players strictly prefer the same option (say B), they are still entitled to choose (S, S), a Nash-equilibrium of this game. The idea here being that although they both prefer going to the Bach concert, they still prefer going to the Stravinsky concert together rather than going to different concerts. In games of this sort, the theory of Nash-equilibrium gives agents exactly the same reasons for playing a payoff-dominated strategy as for playing a payoff-dominant one.

The conformity game, as any pure coordination game, pushes this limitation of the theory of Nash-equilibrium even further, given that the obvious refinement which would lead agents to select, among the Nash-equilibria, the one with the highest payoff (if this exists), cannot be applied due to the complete symmetry of the payoffs. Similar considerations apply to risk-dominance, the 'cautious' dual of payoff-dominance (Harsanyi and Selten, 1988).

It follows that traditional solution concepts are generally inadequate for the conformity game, and indeed for any other game of (pure) coordination. The general feeling on the matter can be illustrated by recalling Schelling's own words (1960):

> Poets might do better than logicians at this game, which is perhaps more like 'puns and anagrams' than like chess. (Schelling, 1960, 58)

An entirely similar attitude is shared (4 decades later) by Camerer, who indeed argues in favour of the empirical (behavioural) investigation on the way players choose among equilibria. As to the 'logical' approach, he remarks that

> [t]his *selection* problem is unsolved by analytical theory and will only be solved by observation. (Camerer, 2003)

Still, as noted by Schelling, players can generally do better than plain randomization in pure coordination games. The extensive empirical investigations that took place over the past decades (see, e.g., Mehta et al., 1994; Sugden, 1995; Janssen, 1998, as well as the results of computer simulations Kraus et al., 2000, strongly support Schelling's early insight that there are in fact choice processes that can facilitate conformity [i.e., that lead agents to coordinate their choice better than plain randomization]).

In the remainder of this paper we will provide a formalisation of a solution concept for the conformity game which is based on the considerations about salience and is underpinned by the principle of charity discussed in Section 3.1.3.

3.3 Solving the conformity game

Recall that the key element intervening in the representation of the conformity game is given by possible worlds, which in the present interpretation amount to the strategies available to the players. We clearly have two

possibilities: either worlds (strategies) in K have no structure other than being distinct elements of a set, or worlds in K do have some structure and in particular there are properties that might hold (be true) in (of) some worlds. In the former case we seem to be forced to accept that agents have no better way of playing the conformity game other than picking some world $f_i \in K$ at random (i.e., according to the uniform distribution). In the latter case, however, agents might use the information about the structure of the worlds in K to focus on some particularly 'distinguished' option to be taken as a focal point.

Consider, for example, the simple case in which worlds (strategies) are maps $f : 4 \longrightarrow 2$ and suppose $K = \{f_1, f_2, f_3, f_4, f_5\} \subseteq 2^4$ is presented as the matrix in Figure 3.2.

	0	1	2	3
f_1	0	0	0	1
f_2	0	1	0	0
f_3	0	1	1	0
f_4	1	1	1	1
f_5	0	0	1	0

Figure 3.2: A representation of the strategy set K

We know from the strategic representation of the conformity game that each pair of identical strategies yields the same utility, so players who intend to conform must look for salient properties to characterize some of the options as those which are likely to be selected by another agent. At the same time, however, we want to rule out the possibility that agents will take into account inessential properties of the set K as being salient, so our first goal is that of ensuring the complete symmetry of the representation. A way of achieving this consists in informing each agent that it is being presented with a matrix K (for instance the one illustrated in (2) which agrees to the one faced by the other player only up to permutations of A and permutations of 2, that is to say, only up to permutations of the columns (and of course rows) of the matrix as well as the uniform transposition of 0's and 1's.

On the assumption of like-mindedness, i.e. common reasoning, if one of those binary strings, say f_j should stand out as having some distinguished properties, agents will conclude that such properties are indeed intersubjectively accessible and hence select f_j. In this way players will go about producing a *reason* for selecting the option f_j. We now move on to formalize this notion.

3.3.1 Introducing asymmetries with reasons

Given the inapplicability of the payoff-dominance principle to the conformity game, the analogy with coordination games suggests that in order to facilitate triangulation we need to introduce some *asymmetries* among the strategies available to the players of the conformity game. We propose here

to formalise this by means of a choice process derived from the *Minimum Ambiguity Reason* introduced in Hosni and Paris (2005).

In a nutshell, the construction of this choice process, or Reason, takes place by means of identifying certain selection principles that players of the conformity game might come to tacitly agree upon, given the goal of the game and their common knowledge of it. This construction will adhere to the charity principle recalled above, in that it is pivoted on the idea that the only clue available to the players about each others' world view is that they share common reasoning.

We define a Reason R to be a choice function from the domain of the conformity game $\wp^+(2^A)$ to itself such that $R(K) \subseteq K$. The general intuition, as discussed in connection with radical interpretation, is that agents should apply Reasons to discard those possible strategies that will prevent them from conforming on their mutual expectations. Given the like-mindedness assumption and the fact that the size of K is proportional to the uncertainty of the players about each other's behaviour, it can be immediately appreciated that a *perfect reason* will be a choice function which always returns a singleton, a unique strategy. It is likewise immediate to see, however, that we cannot expect this to happen in general. As we learnt from radical translation and interpretation, there can be real indeterminacy in the choice problem at hand.

Hence, if after applying their Reason players are left with a plurality of strategies, they will conclude that the choice problem at hand is just underdetermined with respect to the information they possess (the structure of their binary matrix) and will go about to select at random from $R(K)$. In the worst possible case agents will find that $R(K) = K$. At this opposite extreme from the perfect reason, agents will just realize that the strategies from which the choice is to be made are—to their lights—absolutely undistinguishable.

The construction of the Minimum Ambiguity Reason, then, just amounts to constraining the choice process R in such a way as to facilitate the identification of focal points in the conformity game. This characterization will be provided by means of an effective procedure.

3.3.2 The minimum ambiguity reason

Our first goal is constraining R in a way that will provide an adequate formalisation of the symmetries among the players and the possible strategies. This will lead us to formulate the first requirement imposed on the algorithm for computing $R(K)$, namely that if f and g are, as elements of K, *indistinguishable*, then $R(K)$ should not contain one of them, f, say, without also containing the other, g. In other words, an agent should not give positive probability to picking one of them but zero probability to picking the other. The argument for this is that if they are 'indistinguishable' on the basis of K then another

agent could just as well be making a choice of $R(K)$ which included g but not f. Since agents are trying to make the same ultimate choice of element of K, taking that route may be worse, and will never be better, than avoiding it. Indeed, this requirement can be further motivated by direct reference to the radical interpretation problem. The ideal goal of translation as well as interpretation, consists in individuating systematically synonymy among linguistic expressions. In our abstract mathematical setting, synonymy can be understood as "undistinguishability" among possible worlds. It, therefore, follows that accepting in $R(K)$ only one of a pair of undistinguishable worlds amounts to admitting the systematic violation of synonymy, a most undesirable situation for any theory of interpretation.

The second requirement is that the players' choice of $R(K)$ should be as small as possible (in order to maximize the probability of randomly picking the same element as another agent) subject to the additional restriction that this way of thinking should not equally permit another like-minded agent (so also, globally, satisfying the first requirement) to make a different choice, since in that case any advantage of picking from the small set is lost.

The first consequence of this is that initially the agent should be looking to choose from those minimal subsets of K closed under indistinguishability, 'minimal' here in the sense that they do not have any proper non-empty subset closed under indistinguishability. Clearly, if this set has a unique smallest element then the elements of this set are the least ambiguous, most outstanding, in K and this would be a natural choice for $R(K)$. However, if there are two or more potential choices $X_1, X_2, \ldots, X_k$ at this stage with the same number of elements then the choice of one of these would be open to the obvious criticism that another 'like-minded agent' could make a different (in this case disjoint) choice. Faced with this revelation our agent would realise that the 'smallest' way open to reconcile these alternatives is to now permit $X_1 \cup X_2 \cup \cdots \cup X_k$ as a potential choice whilst dropping $X_1, X_2, \ldots, X_k$.

The agent now looks again for a smallest element from the current set of potential choices and carries on arguing and introspecting in this way until eventually at some stage a unique choice presents itself. We will understand this unique choice as the required focal point, the center of agents' triangulation.

In what follows, we shall give a formalisation of this procedure. All the results to follow have appeared (or are straightforward generalisations of those spelled out) in Hosni and Paris (2005) and Hosni (2005) and therefore the proofs are omitted here.

3.3.3 Transformations

We begin by formalising the intended notion of undistinguishability among worlds in K. In the current abstract mathematical framework this amounts to

providing a formalisation of synonymy among possible options—with respect to the radical interpretation problem—as well introducing a utility-free evaluation (pairwise comparison) of the strategies available to the agents in the conformity game.

The central concept is that of a *transformation* of possible worlds. The intuition to be formalised being that a transformation can act on a set of possible worlds by operating changes that agents should consider inessential to the choice problem they are facing. Hence the possibility of transforming (formally) one world into another one will lead agents to consider these to be indistiguishable.

We define a function $j : K \to 2^A$ a *transformation of* K if there is a permutation σ of A and a permutation δ of $\{0, 1\}$ such that $j(f) = \delta f \sigma$ for all $f \in K$. We shall say that a transformation j of K is *a transformation of* K *to itself* if $j(K) = K$.

The intuition here is that a transformation j of K to itself produces a copy of K—$j(K)$—in which the 'essential structure' of K is being preserved. To see this in practice, simply take the matrix introduced above in Section 3.3, from which the explicit mention of the set A and the labels of the binary strings are omitted, as illustrated in Figure 3.3:

$$\begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{array}$$

Figure 3.3: The matrix representing K

It can be easily seen that putting δ to be the identity function (id) and $\sigma = (1, 2)$ (the permutation transposing 1 and 2 in $\{0, 1, 2, 3\}$), we will obtain the transformation transposing the 'second' and 'third' column of the above matrix. Furthermore, by letting $\sigma' = id$ and $\delta' = (0, 1)$ we obtain a matrix with 0's and 1's exchanged. These can be represented as:

$$\begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{array} \quad \text{and} \quad \begin{array}{cccc} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{array}$$

let's say $j(K)$ and $j'(j(K))$, respectively.

Hence the requirement that the players' choices should be invariant under these 'inessential' transformations is captured by the following:

Transformation principle
Let $K \in \wp^+(2^A)$, and j be a transformation of K. Then

$$j(R(K)) = R(j(K)). \tag{Tr}$$

Intuitively, the Transformation principle states that applying some transformation j to the set of best elements (according to R) of K is just the same as choosing the R-best elements of the transformation of K by j.

The second step then in the construction of the Minimum Ambiguity Reason consists in the formalization of the 'ambiguity of worlds within K', so that agents, while satisfying the Transformation principle will go about selecting the most outstanding elements of K—the focal points. Notice that, as one would clearly expect from the discussion on triangulation and focal points, 'ambiguity' is being characterized as a contextual notion, relative in fact to the knowledge K.

So let $K \in \wp^+(2^A)$. Then for $f \in K$, *the ambiguity class of f within K at level m* is recursively defined by:

$$\mathbb{S}_0(K, f) = \{g \in K \mid \exists \text{ trans. } j \text{ of } K \text{ such that } j(K) = K \text{ and } j(f) = g\}$$

$$\mathbb{S}_{m+1}(K, f) = \begin{cases} \{g \in K \mid \ |\mathbb{S}_m(K, f)| = |\mathbb{S}_m(K, g)|\} & \text{if } |\mathbb{S}_m(K, f)| \leq m + 1; \\ \mathbb{S}_m(K, f) & \text{otherwise.} \end{cases}$$

The intuition of the base case is that of grouping together those possible worlds g which are in the range of a transformation j of K to itself taking f as argument, thus giving an initial measure of the ambiguity of f in K. The recursive step, on the other hand, causes worlds with the same ambiguity to be grouped in the same class, the purpose of the side condition being that of avoiding coalescing classes 'too quickly' (and hence possibly losing some 'natural' features of the relevant classes).

Define now, for $f, g \in K$, the relation

$$g \sim_m f \Leftrightarrow g \in \mathbb{S}_m(K, f).$$

Recall that one of the requirements of the algorithm is that agents should avoid selecting one but not both elements of a pair of undistinguishable options. Indeed the following proposition ensures that as f ranges over K, $\sim_m$ induces a partition on K.

Proposition 1. *$\sim_m$ is an equivalence relation and the sets $\mathbb{S}_m(K, f)$ are its equivalence classes.*

Moreover, this m-th partition is a refinement of the $m + 1$-st partition. In other words, the sets $\mathbb{S}_m(K, f)$ are increasing and so eventually constant fixed at some set which we shall call $\mathbb{S}(K, f)$.

We are now ready to introduce the *ambiguity of f within K*, which is formally defined by:

$$\mathbb{A}(K, f) =_{def} |\mathbb{S}(K, f)|.$$

Finally, we can define the *Minimum Ambiguity Reason* $R_{\mathbb{A}}(K)$ by letting:

$$R_{\mathbb{A}}(K) = \{f \in K \mid \forall g \in K,\ \mathbb{A}(K, f) \leq \mathbb{A}(K, g)\}. \tag{1}$$

As an immediate consequence of the definition of $R_{\mathbb{A}}$ we have the following result:

Proposition 2. *$R_{\mathbb{A}}(K) = \mathbb{S}(K, f)$, for any $f \in R_{\mathbb{A}}(K)$*

Recall that agents have to select a *unique* option from K, so as argued when introducing the informal procedure, whenever the size of $R_{\mathbb{A}}(K)$ is greater than 1, players will just randomize.

The following results show that the intuition that players of the conformity game should select the 'most distinguished' worlds from a set K while satisfying closure under undistinguishability is indeed captured by the minimum ambiguity reason.

Theorem 3. *$R_{\mathbb{A}}$ satisfies Transformation.*

Theorem 4. *A non-empty $K' \subseteq K$ is closed under transformations of K into itself if and only if there exists a Reason R satisfying Transformation such that $R(K) = K'$.*

The importance of these results is that in the construction of $R_{\mathbb{A}}(K)$ the choices $\mathbb{S}_m(K, f)$ which were eliminated (by coalescing) because of there currently being available an alternative choice of a $\mathbb{S}_m(K, g)$ of the same size are indeed equivalently being eliminated on the grounds that there is a like-minded agent, even one satisfying Transformation, who could pick $\mathbb{S}_m(K, g)$ in place of $\mathbb{S}_m(K, f)$. In other words it is not as if some of these choices are barred because no agent could make them whilst still satisfying Transformation. Once a level m is reached at which there is a unique smallest $S_m(K, f)$ this will be the choice for the informal procedure. It is also easy to see that this set will remain the unique smallest set amongst all the subsequent $S_n(K, g)$, and hence will qualify as $R_{\mathbb{A}}(K)$. In this sense then our formal procedure fulfills the intentions of the informal description given at the beginning of this section.

3.4 Concluding remarks

We conclude by evaluating the extent to which the Minimum Ambiguity Reason contributes towards providing a formalization of the problems arising in the process of triangulation and in the selection of multiple Nash-equilibria in pure coordination games.

$R_{\mathbb{A}}$ and triangulation. The distinct level of abstraction stands out in the comparison of the radical interpretation and the conformity game situations. While with the radical interpretation problem it is attempted to lay down a theory of interpretation *for natural languages*, the choice problem faced by the agents in the conformity game is based on the selection of otherwise meaningless binary strings. In both cases, however, agents should rationally aim at performing *disambiguating* choices and the framework of Rationality-as-Conformity provides agents with an algorithmic procedure to achieve this. It is a matter of future research to investigate the disambiguation of options arising in gradually more and more complicated structures.

Whilst the agents involved in the radical interpretation situation can appeal to actual *observations* of their own reciprocal (non linguistic) behaviour, the players of the conformity game can only *conjecture* about the expected behaviour of their fellows. Again, we see this as a difference of levels of abstraction, yet not of kind, as we concentrate on the 't_0' of the triangulation process, when the transition takes place from agents not sharing any communication devices, to conforming on the use of some. This is being paralleled by the controlled experiments in pure coordination games, as reported, e.g., in Mehta et al. (1994).

$R_{\mathbb{A}}$ and focal points. How far the Minimum Ambiguity Reason goes towards providing a solution to pure coordination games depends, in the first place, on whether the *uniqueness* of the selection is considered a necessary condition on the solution concept or not. Since the early investigations in focal points and salience, uniqueness has been given considerable importance. In some recent, computationally-oriented investigations on the subject, however, other properties of focal points have received attention, with the uniqueness requirement being considerably relaxed (see Kraus et al. (2000) for a comprehensive study). The construction of the Minimum Ambiguity Reason makes explicit the fact that certain coordination problems might be so nebulous that agents cannot rationally go beyond the selection of 'small' sets of options, the minimally ambiguous ones, if the closure under undistinguishability requirement is to be satisfied. The drawback for failing this being, as illustrated above, the possibility of systematically missing coordination.

Acknowledgments

I am greatly indebted to Jeff Paris for the formulation of the Minimum Ambiguity Reason and for many stimulating discussions on the topic. An early version of this paper was presented at *The 2004 Prague International Colloquium on Logic, Games and Philosophy: Foundational Perspectives*. I wish to thank the participants for many helpful remarks.

References

Camerer, C. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, Princeton, NJ.

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford.

Feldman, R. (1998). Principle of charity. In Craig, E., editor, *Routledge Encyclopedia of Philosophy*. Routledge, London.

Glock, H. (2003). *Quine and Davidson on Language, Thought and Reality*. Cambridge University Press, Cambridge.

Harsanyi, J. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. MIT, Cambridge, MA.

Hosni, H. (2005). *Rationality as Conformity.* Doctoral thesis, School of Mathematics, The University of Manchester, Manchester.

Hosni, H. and Paris, J. (2005). Rationality as conformity. *Knowledge Rationality and Action (Synthese)*, 144(2): 249–285.

Janssen, M. (1998). Focal points. In *New Palgrave Dictionary of Economics and the Law*. MacMillan, London.

Kraus, S., Rosenschein, J. S., and Fenster, M. (2000). Exploiting focal points among alternative solutions: Two approaches. *Annals of Mathematics and Artificial Intelligence*, 28(1–4):187–258.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA.

McGinn, C. (1977). Charity, interpretation, belief. *The Journal of Philosophy*, 74(9):521–535.

Mehta, J., Strarmer, C., and Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination. *The Americal Economic Review*, 84(3):658–673.

Nozick, R. (1993). *The Nature of Rationality*. Princeton University Press, Princeton, NJ.

Osborne, M. J. (2004). *An Introduction to Game Theory*. Oxford University Press, Oxford.

Parikh, P. (2000). Communication, meaning and interpretation. *Linguistic and Philosophy*, 23: 185–212.

Paris, J. B. (1994). *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge University Press, Cambridge.

Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximated Reasoning*, 4:183–224.

Paris, J. B. and Vencovská, A. (2001). Common sense and stochastic independence. In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*, pages 203–240. Kluwer, Dordrecht.

Quine, W. V. (1960). *Word and Object*. MIT Press, Cambridge, MA.

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.

Sugden, R. (1995). A theory of focal points. *The Economic Journal*, 105(430):533–550.

van Rooy, R. (2004). Evolution of conventional meaning and conversational principles. *Synthese*, 139(2):331–366.

Wachbroit, R. (1987). Theories of rationality and principles of charity. *British Journal for the Philosophy of Science*, 38:35–47.

Chapter 4

FALLACIES AS COGNITIVE VIRTUES

Dov M. Gabbay[1] and John Woods[2]

[1,2] *Department of Computer Science, King's College London*
dg@dcs.kcl.ac.uk

[2] *Department of Philosophy, University of British Columbia*
woodsj@dcs.kcl.ac.uk
jhwoods@interchange.ubc.ca

Sometimes you had to say Stuff Logic *and go with the flow.*
—Reginald Hill, *Good Morning Midnight*

Abstract In its recent attention to reasoning that is agent-based and target-driven, logic has re-taken the practical turn and recovered something of its historic mission. In so doing, it has taken on in a quite general way a game-theoretic character, precisely as it was with the theory of syllogistic refutation in the *Topics* and *On Sophistical Refutations*, where Aristotle develops winning strategies for disputations. The approach that the present authors take toward the logic of practical reasoning is one in which cognitive agency is inherently strategic in its orientation. In particular, as is typically the case, individual agents set cognitive targets for themselves opportunistically, that is, in such ways that the attainment of those targets can be met with resources currently or forseeably at their disposal. This not to say that human reasoning is so game-like as to be utterly tendentious. But it does make the point that the human player of the cognitive game has no general stake in accepting undertakings that he has no chance of making good on.

Throughout its long history, the traditional fallacies have been characterized as mistakes that are attractive, universal and incorrigible. In the present essay, we want to begin developing an alternative understanding of the fallacies. We will suggest that, when they are actually employed by beings like us, they are defensible strategies in game-theoretically describable pursuit of cognitive (and other) ends.

4.1 Introductory remarks

In its recent return to reasoning that is agent-based and target-driven, logic has recovered something of its historic mission. In so doing, it has taken on in a quite general way a game-theoretic character, precisely as it was with

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

Aristotle's theory of syllogistic refutation in the *Topics* and *On Sophistical Refutations*. Aristotle here presents winning strategies for disputations. They pivot on the refuter's exploitation of his opponent's concessions. While the opponent must believe his concessions, the refuter need not. The approach that the present authors take toward the logic of agent-based target-driven reasoning is one in which cognitive agency is inherently strategic in its orientation. In particular, as is typically the case, individual agents set cognitive targets for themselves opportunistically, that is, in such ways that the attainment of those targets can be met with resources currently or forseeably at their disposal. This not to say that human reasoning is so game-like as to be utterly tendentious. But it does make the point that the human player of the cognitive game has no general stake in accepting undertakings that he has no chance of making good on. Throughout its long history, the traditional fallacies have been characterized as mistakes that are attractive, universal and incorrigible. In the present essay, we want to begin developing an alternative understanding of the fallacies. We will suggest that, when they are actually employed by beings like us, they are defensible strategies in game-theoretically describable pursuit of cognitive (and other) ends. Needless to say, the generically game-theoretic approach has developed several more specialized tendrils. Some of these involve a re-writing of classical first order logic. Others are extensions or adaptations of the mathematical theory of games. Still others refine the generic notion into technically versatile models of dialogue. All of these are welcome developments, and many are of enduring importance. In some of our writings in progress, the more peculiarly game-theoretic aspects of practical reasoning are developed. But we continue to think that the generic notion, embodying the fundamental idea of strategies for the attainment of cognitive targets, is also of lasting importance. This is something that we shall attempt to demonstrate in this essay.

The present work is adapted from our book in progress, *Seductions and Shortcuts: Fallacies in the Cognitive Economy* (Gabbay and Woods, 2009). Our principal purpose here is to introduce readers to that work's founding assumption, and to identify some of the considerations that lend the idea support. We also have it in mind to attend to an important ancillary matter. It is the task of elucidating the role of what an agent is *capable of* in assessing whether his performance is faulty or defective. The essay is structured as follows. In Part I we discuss the question of cognitive agency. Part II illustrates our approach to fallacies.

PART I: PRACTICAL AGENCY

We begin with the so-called Gang of Eighteen, the name given to a loose confederacy of presumed errors that are discussed with a considerable regularity

in the contemporary literature on fallacies (Woods, 2004).[1] In one recent treatment (Woods et al., 2004), the Gang of Eighteen is represented by the following list.

ad baculum
ad hominem
ad misericordiam
ad populum
ad verecundiam
affirming the consequent
amphiboly
begging the question
biased statistics
complex question
composition and division
denying the antecedent
equivocation
faulty analogy
gambler's
hasty generalization
ignoratio elenchi
secundum quid

The Gang of Eighteen (*GOE,* for short) embeds a certain view of what it is to be a fallacy. It sees fallacies as mistakes of reasoning (or arguing) that are attractive, universal and incorrigible. So conceived of, fallacies retain a striking kinship with Aristotle's original definition, in which a fallacy is an argument (or a piece of reasoning) that appears to be good in a certain way, but is not in fact good in that way. It is easy to see that the first two marks of fallaciousness are expressly caught by Aristotle's definition. For a fallacy is not only an error but, because it appears not to *be* an error, is a mistake that has a certain attractiveness. It is also clear that Aristotle intends the attractiveness of fallacies to give them a kind of universal appeal: Fallacies are errors that people in general are disposed to make, not just the logically challenged or the haplessly inattentive. If their attractiveness grounds their general appeal, it also grounds their incorrigibility. To say that a fallacy is an incorrigible error is to say that, even when properly diagnosed, there is a general tendency to recidivize. The modern notion incorporates these interdependencies. Accordingly, we have it that

[1] We emphasize the looseness of the grouping. In Copi (1986) 17 fallacies are discussed; in Carney-Scheer (1980) the number is 18; Schipper-Schuh (1959) runs to 28; and Black (1946) limits itself to only 11. While all these lists are pairwise inequivalent, there is nonetheless a considerable overlap among them.

Proposition 1 (Fallacies)**.** *A fallacy is a generally attractive and comparatively incorrigible error of reasoning (or argument).*

The *negative thesis* we wish to propose is that the general idea of fallacy is correct but that there is something gravely defective about the Gang of Eighteen and any of its standard variations. As we shall attempt to show, there are two difficulties with these lists:

1. Some of their members aren't fallacies.
2. Those that are errors aren't usually mistakes committed by beings like us.[2]

Our *positive thesis* is that

3. Several of the *GOE* are actually cognitive virtues.

To make good on these theses requires that

(a) We identify the members of the *GOE* of which the theses are true.

(b) Establish in each case that the relevant thesis is indeed true.

(c) Give some account of how it came to be the case that by our lights, the defective inventory of fallacies took hold.

In proceeding with these tasks we want to make it clear at the beginning that it is not our view that people don't commit fallacies. Our view rather is that the *GOE* has not managed to capture any of them in wholly convincing ways. For either they are indeed fallacies which we happen not to commit, or we do commit them, but they are not fallacies.

In its most usual meaning, a fallacy is a *common misconception.* It is an attractive, widely held belief that happens to be untrue. In many cases, it is also a belief that people have difficulty letting go of, even, after its falsity has been acknowledged. So whereas the received idea among logicians has been that a fallacy is an *argument* that is defective in the traditionally recognized ways, the view of the layman is that it is a *belief* that has the requisitely counterpart features. We may wish to take note of the point that if our present theses about the Gang of Eighteen can be sustained, we will have shown that the logician's inventory of the fallacies is in the layman's sense itself a fallacy.[3] If this should

[2]Given one's tendency to apply the word "incorrigible" to practices (or practitioners) one disapproves of, this is very much the right word for the fallacies *as traditionally conceived of.* Since ours is a view of the fallacies that rejects the traditional conception, we shall replace "incorrigible" with the more neutral-sounding "irreversible".

[3]A theme sounded by two recent writers. See Grootendorst (1987), which is entitled "Some fallacies about fallacies", and Hintikka (1987), which is entitled "The fallacy of fallacies". For reservations see, in the first instance, Woods (2004, Chapter 9) in the second, Woods and Hansen (1997), and for a rejoinder (Hintikka, 1997).

prove to be the right sort of criticism to press against the *GOE* approach, then something like the following argument schema must itself be defective. Let us call it

The Fallacy of Fallacies Schema

1. Practice *P* is universal, attractive and incorrigible (irreversible).
2. Practice *P* lacks property *Q* (e.g. validity).
3. Therefore, practice *P* is a fallacy.

Our view is as follows. There are members of the Gang of Eighteen of which (1) and (2) are true, but (3) is false. There are other members of which (3) and (2) are true; and (1) is not true *of us.*

We are in no doubt about the burdens we have taken on in staking our case against the Fallacy of Fallacies Schema. Certainly, there is no realistic prospect of doing so in the space of a single chapter. So we shall proceed as best we can, beginning with some issues we believe it necessary to explore in some detail before moving on to the negative and positive theses about *GOE.* This will leave us space enough to test these claims against only one class of fallacies, known collectively as "hasty generalization". The complete case against *GOE* is the business of *Seductions and Shortcuts.*

4.2 Logic's cognitive orientation

Since its inception 2,500 years go, logic has been thought of as a science of reasoning. Aristotle held that the logic of syllogisms is the theoretical core of the wholly general theory of argument called for in the *Topics.* Even centuries later, when logic took its momentous turn toward the mathematical, the idea persisted that the canons of logic regulate at least mathematical reasoning which, in some versions, is reasoning at its best. One of the striking features of mainstream mathematical logic is the distance at which it stands from the behaviour of real-world reasoning agents. In its anti-psychologicism, context-independency and agent-indifference, it is hardly surprising that mathematical logic endorses principles which real-life reasoners do not, and often cannot, conform to. Rather than taking this as outright condemnation of reasoning as it actually occurs, mathematical logicians have sought a degree of mitigation in the idea that real-life reasoning is correct to the degree to which it *approximates* to conformity to these ideal canons of strictness.[4] Although the

[4]Cf. Matthen (2002, 344), whose mention of it is disapproving: "Human reasoning *tries* to instantiate logic, but, because of the regrettable necessity of making do in the real world, it falls somewhat short. In this it is something like human virtue as Aristotle describes it—a second-best life imposed on us by the exigencies of the human condition." The more nearly correct view is that "[o]ur capacity for reason is dictated by symbolic complexity required for tasks other than truth maximization" (Matthen, 2002).

approximation-to-the-ideal view has had its critics (e.g., Gabbay and Woods, 2003a), other reactions have been more constructive and conciliatory. They are reactions linked together by the common purpose of extending and adapting mainstream logic itself, so as to produce systems capable of modeling aspects of actual reasoning which the standard systems of mathematical logic leave out of account. Within the logic community these extensions or adaptations include modal logics and their epistemic and deontic variations (von Wright, 1951; Hintikka, 1962; Kripke, 1963; Gabbay, 1976; Lenzen, 1978; Chellas, 1980; Hilpinen, 1981; Gochet and Gribomont, 2005), probabilistic and abductive logics (Magnani, 2001; Williamson, 2002; Gabbay and Woods, 2009), dynamic logics (Harel, 1979; van Benthem, 1996; Gochet, 2002), situation logics (Barwise and Perry, 1983), game-theoretic logics (Hintikka and Sandu, 1997), temporal and tense logics (Prior, 1967; van Benthem, 1983), time and action logics (Gabbay et al., 1994), systems of belief dynamics (Alchourron et al., 1985; Gabbay et al., 2002, 2004a, b) practical logics (Gabbay and Woods, 2003b, 2005), and various attempts to float the programme of informal logic.[5] Work of considerable interest has also arisen in the computer science, AI and cognitive psychology communities, with important developments in defeasible, non-monotonic and autoepistemic reasoning, and logic programming (Sandewall, 1972; Kowalski, 1979; McCarthy, 1980; Reiter, 1980; Moore, 1985; Pereira, 2002; Schlechta, 2004).

The net result of these considerable efforts is a marked reorientation of logic to the ins-and-outs of reasoning as it actually occurs. It may be said that, if in the aftermath of the mathematical turn it were ever in doubt, logic has now to some extent reclaimed its historical mission of probing how human reasoning does (and should) work.

This is a significant development. If logic is once more a science of reasoning, it is well to pause and take some note of what reasoning is *for.* It is clear upon inspection that, in a rough and ready way, reasoning serves as an aid to *belief-change* and *decision.* Certainly it seems true to say that is these aspects of reasoning in which the new logic (if we might appropriate that term) seems most to concentrate on (Gabbay and Woods, 2001b). This being so, an answer to our present question becomes apparent. Reasoning is an aid to cognition. Accordingly,

Proposition 2 (The new logic)**.** *Logic investigates reasoning in its role as an aid to cognition. Or, as we might now say the new logic is an investigation of* (requisite aspects of) *cognitive systems.*

[5]The informal logic movement comprises three over-lapping orientations. One is argumentation theory (Johnson, 1996, 2000; Freeman, 1991; Woods, 2003). Another is fallacy theory (Hamblin, 1970; Woods and Walton, 1989; Hansen and Pinto, 1995; Walton, 1995; Woods, 2004). Completing the trio is dialogue-logic (Hamblin, 1970; Barth and Krabbe, 1982; Hintikka, 1981; MacKenzie, 1990; Walton and Krabbe, 1995; Gabbay and Woods, 2001a, c).

4.3 Practical agency

A consideration of agency is central to our task. Our view of agency is set out in a PLCS—*a practical logic of cognitive systems,* which can be sketched as follows:

- A cognitively sensitive logic is a principled description of certain aspects of the behaviour of a cognitive system, chiefly of those aspects that figure centrally in *belief and decision dynamics*.
- A cognitive system is a triple of an *agent C*, cognitive *resources R* and *cognitive tasks J* performed in real *time t*.
- A cognitive agent is an information-processing device capable, among other things, of *belief, inference* and *decision*.
- A cognitive agent is always an agent of a certain type, depending on where he or it sits under a partial order that we will call "commanding greater (cognitive) resources than".
- Such resources include, but are not exhausted by, *information, time* and *computational capacity*.
- A cognitive agent is a *practical* agent to the extent that it ranks low in this ordering.
- Accordingly, *practical reasoning* is the reasoning of a practical agent.
- A cognitive agent is a *theoretical* agent to the extent that it sits high in this same ordering.
- Accordingly, *theoretical reasoning* is the reasoning done by theoretical agents.
- Practical agents include *individuals*.
- Theoretical agents include *institutions*.
- It cannot in general be supposed that practical and theoretical reasoning are geared to the same *goals* or *targets* and subject to the same *performance standards*.
- Compared with what theoretical agency can achieve, practical reasoner's operate with *fewer resources*.
- Compared with what theoretical agency can achieve, practical agents set *more modest cognitive targets*.

Accordingly,

Proposition 3 (Practical agency)**.** *Practical agency is triangulated by two main factors. One is the factor of comparative resource-scantness. The other is the factor of comparative target-modesty.*

We accept that ours is a somewhat unusual use of the word "theoretical". In the account given by *PLCS,* when an individual is, for example, trying to simplify a proof of the completeness of modal logic in time to meet an editor's deadline, he is engaged in practical reasoning, even though, in one standard sense of the word, the completeness problem is a theoretical problem. In putting the word to our uses here, we intend neither rivalry nor imperiousness. Ours is but another sense of the word, which we've introduced as a technical term. Even so, the gap between our use and other uses typified by the theoretical status of the completeness problem is not as large as one might think. There are legions of theoretical problems (in the completeness-problem sense) that demand the resources and epistemic standards that characterize theoretical agency in our sense. Most of NASA's scientific problems are theoretical in the completeness-problem sense, and NASA is an exemplar of theoretical agency in our sense. All the same, it is well to note that the word "practical" has no wholly natural (non-negative) antonym in English. So any candidate we might select is bound to strike the ear somewhat oddly.[6]

4.4 Cognitive economies

Seen in this way, practical agents operate in a *cognitive economy.* They seek to attain their targets with the resources at hand and with due regard for what they are naturally unfitted for. An individual agent's resources are for the most part available to him in low finite quantities. Given the multiplicity of his cognitive ambitions and the sundry demands of maintaining his balance in a world of constant change, there is an inevitable competition for the resources needed for the advancement of cognitive agendas. In much of what he does, an agent is a zero-sum consumer of his own resources. In lots of cases, he can also seek to draw down his competitors' resources as well. The zero-sum harshness of resource-draw demands that in most cases an agent pay attention to costs and benefits. This is not to say that his cognitive *targets* are economic (not usually anyhow) but rather that, whatever they chance to be, handling them rationally requires that these economic factors be taken into account. This is true of agents both practical and theoretical. Resources are finite for each and ambitions frequently outrun what resources are able to handle. The rationality of

[6]Various candidates have been proposed. We find that none generalizes in quite the desired way: *specialized, alethic* (or *doxastic*), *formal, precise, strict, context-free, abstract* and, of course, *theoretical* (in the completeness proof sense). For further discussion, see Gabbay and Woods (2003a, 13–14).

cognitive agency takes this factor of comparative resource-scantness into deep account. In virtually all that they do as cognitive beings, agents of both stripes must learn to economize.

Given these resource limitations, we may postulate for practical reasoners various scant-resource *compensation strategies.*[7] Leading the list, hardly surprisingly, is *the setting of targets of comparative modesty,* itself an instance of the adjustment of goals to the means available for their effective realization. Other strategies include:

- A propensity for *hasty generalization*
- A facility with *generic inference,* and other forms of non-universal generalization
- Ready discernment of *natural kinds*
- A propensity for *default reasoning*
- A capacity *to evade irrelevance*
- A disposition toward belief-update and discourse *economies,* such as reliance *ad verecundiam* upon the assurance of others
- A facility with *conjecture* (or, in plainer English, *guessing*)[8]
- A talent for *risk aversion*
- An architecture for *inconscious* or *implicit* cognition.[9]

We emphasize that scantness of resources is a comparative matter. By and large individual agents have fewer of them than institutional agents such as NASA or MI5. It is sometimes the case, though not uniquely or invariably, that resource-paucity makes for resource-*scarcity.* But it would be quite wrong to leave the suggestion that individual agents are resource-strapped by definition, as it were.

There are two quite general attributes that are unique to the practical agent, and which give him a clear advantage in the cognitive economy. One is the *emotional make-up* of (human) practical agents—in particular their capacity to feel fear, which plays a pivotal role in risk-averse inference. The other is that, to a degree far greater than applies to institutional agents, practical agents are capable of a timely response to *feedback mechanisms.* This is standing occasion for the practical agent to correct damaging or potentially damaging

[7] See Gigerenzer and Selten (2001).
[8] See Peirce (1992, 1931–1958, 7.220).
[9] This on the analogy of implicit perception, concerning which see Rensink (2000).

errors before the harm they portend is done. It conduces toward what we might call "an efficiently corrigible fallibility". Institutional agents, on the other hand, are notorious for their feedback-laggardness. It is a laggardness that routinely compromises efficiency and often compromises correction.

It would appear that, on the face of it, the list of scarce-resource compensation strategies is rife with fallacy, what with its endorsement of hasty generalizations and *ad verecundiam* and reasoning. Should we not conclude, therefore, that practical agency and practical reasoning are intrinsically defective? It is the business of Part II to deal with this question, at least in part.

4.5 Cognitive targets

We should now say a word about cognitive targets.

Proposition 4 (Cognitive targets)**.** *A target T for an agent X is a cognitive target for him (or it) if and only if T is attainable only by way of a* cognitive state *of X.*

For example, if X wants to know whether Y will accompany him to the movies, his target is met when he knows that Y will accompany him to the movies. The desire to know whether is X 's target. X's knowledge—that enables X to hit the target. T, then, is a cognitive target for X.

Not all cognitive targets expressly embed the desire to know; that is, they are not always overt calls for knowledge. X may desire to make a decision between options O_1 and O_2. Upon discovery of some new information, X may now be in a state of knowledge in virtue of which he decides for O_1 rather than O_2. X's state of knowledge closed his decision-agenda. So his decisional target was a cognitive target in our sense. Perhaps it might be said that in his desire to decide between O_1 and O_2, X was implicitly calling for the knowledge that would enable him to turn the trick. There is little point in semantic wrangles over the purported equivalence between "wants to decide" and "wants knowledge that will enable a decision". A target is hit when X no longer has the desire or disposition in terms of which it was constituted in the first place. This can happen in one or other of two ways that can be regarded as cognitive. In one, X is in a state of knowledge that causes X's desire to be *satisfied* or his disposition to be *actualized.* In the other, X is in a state of knowledge that *kills* X's desire or *cancels* his disposition. In the one case, X may desire to know whether his companion will accompany him, to the movies and it may happen that in coming to know that his companion will indeed accompany him that his desire is fulfilled. It may also happen that X desires to know who is using Department copier for personal purposes, and on coming to know that there is some indication that the culprit is his brother, his desire may lapse and his enquiry may cease.

In what follows, we focus on the first kind of case. Accordingly,

Proposition 5 (Attainment)**.** *If T is a cognitive target, then T's attainment requires the satisfaction of the desire embodied in T (or the actualization of its embedded cognitive disposition).*

4.6 The logic of down–below

It is well to emphasize that this talk of cognitive desire is largely an expository device, as indeed is the idea of an agent's cognitive targets. Targets can be likened to *agendas,* to whose examination our (earlier) companion work, *Agenda Relevance,* devotes a number of pages (Gabbay and Woods, 2003a, 37–40). This is not the place to repeat that discussion in detail, but there is some advantage in touching briefly on a few of its principal claims. One is that agendas (hence targets too) need not be consciously held or set, and need not be attended by express recognition of the means of their attainment. Cognitive targets are better understood as cognitive dispositions to be in certain kinds of mental states. But this is much too general a description to capture them adequately. Any cognitive agent, structured in approximately the way we ourselves are, is at virtually all times causally primed to be in the states to which he (or it) is, then and there, susceptible. Suffice it here to say that something counts as a cognitive target when it is of a type that could be consciously held, openly desired and deliberately advanced upon. That targets need not be thus held and advanced is further indication of how much of our cognitive careers are set out and dealt with subconsciously and (probably) sublinguistically. A short way of saying this is that a good deal of human cognition occurs "down below".[10] Reasoning, like cognition itself, also occurs automatically, unconsciously, sublinguistically, hence "down below". But logic investigates reasoning in its role as an aid to cognition. If logic is to honour its pledge to reasoning, it must be prepared in turn to probe the reasoning of down below. Given the constraints, both ethical and mechanical, that inhibit the exposure of human subjects to the vicissitudes of the experimental method, the logician is left with little choice but to abduce and to analogize. Whereupon is surrendered the ancient conceit that logic is the most certain and epistemically privileged of the sciences.[11]

[10] Other characterizations that have been used to capture the idea of reasoning down below are: *unconscious, automatic, inattentive, involuntary, non-semantic* and *deep.* We note in passing the general inequivalence of these descriptors (Gabbay and Woods, 2003a, 37–40).

[11] The logic of down-below is very much in its infancy. But already various ideas of how it might go have started to stir rather attractively. For a connectionist approach, see Churchland (1989, 1995); a *RWR* (representation without rules) orientation is discussed in (Horgan-Tienson, 1999a, b) and (Guarini, 2001); offline anti-representationalism is discussed in (Wheeler, 2001); a semantic space orientation is developed by Bruza et al. (2004, 2006) and connectionist neural net approaches are to be found in (d'Avila Garcez et al., 2002; d'Avila Garcez and Lamb, 2004) and (Gabbay and Woods, 2005, Section 6.8). For a criticism of the idea that logic imposes universal constraints on rationality, see (Matthen, 2002). (Bermúdez, 2004) explores the cognitive wherewithal of young infants and animals.

4.7 Generic reasoning

The identification of a practical agent as someone (or something) that performs his (or its) cognitive tasks under conditions of resource-paucity in pursuit of comparatively modest cognitive targets is one that states a generic fact about practical agents. What is claimed is that it is *characteristic* of the cognitive actions of practical agents that they are performed under such conditions in relation to such targets. It would be a mistake to ignore the plain fact that there are specific cases in which practical agents complete a task without at all depleting the resources required for its wholly satisfactory transaction. Neither is it the case that, in his generic thrall to comparative resource-paucity, the practical agent is invariably at a disadvantage. Whether he is disadvantaged in this way, or not, depends on the cognitive goals it would be appropriate for him to set for himself and on the cognitive wherewithal available for achieving them.

Unlike the universally quantified conditional sentences that inductive logicians recognize as full-bore (or Hempelian) generalizations, generic generalizations (if the pleonasm might be forgiven) are sub-universal in their reach. There is a considerable difference here. The generalization, "For all x, if x is a tiger, x is four-legged", is *brittle.* It is overturned by a single true negative instance. But the generic claim, "Tigers are four-legged", is *elastic.* It can be true even in the face of true counterinstances.[12] This provides the practical agent with further occasion to economize. If he ventures the generic claim rather than the strictly universal claim, he can be wrong in particular without being wrong in general—a nice advantage. Generic generalizations are less precise than Hempelian generalizations; but what is lost in precision is made up for in elasticity. Genericity, in turn, hooks up with the concept of default.

Proposition 6 (Genericity and default inference)**.** *Given the generic claim that tigers are four-legged, together with the fact that Pussy is a tiger, the inference to "Pussy is four-legged" is an inference to a default. What makes it a default is precisely that "Pussy is a four-legged tiger" could be false without making it false that tigers are four-legged.*

Hasty generalization is intimately linked to genericity, which in turn is intimately linked to natural kinds. To see a tiger as the kind of thing it is involves having some grasp of properties it possesses as a thing of that kind. But this is knowing something about what is characteristic of tigers, hence true of them by and large. Seeing that Pussy is a tiger—that Pussy is of the tiger kind, rather than, say, of the James Bond villainess kind—involves an appreciation of what things of that kind are *like;* that things of that kind are by and large four-legged,

[12] For genericity Carlson and Pelletier (1995) is essential reading.

for example. It is doubtless an over-simplification, but something like the following holds true: that appreciating that this thing is of the tiger kind involves appreciating that various of this thing's properties are by and large properties of all things of that kind. So natural kind recognition involves hasty (generic) generalization of kind-properties.

The distinctive advantage of generic generalizations is that they can be retained without qualification even in the face of known counterinstances.

Hempelian generalizations are disabled by true counterinstances, and require, if not outright abandonment, nothing less than reformulation. There are four basic ways of achieving such reformulations, each problematic. One is to hit upon a principled means of exclusion, that is, a means that serves to exclude the requisite class of the true counterinstances that is stateable without making specific mention of them. Another is to restate the original generalization and append to it, one by one, classes of known exceptions. A virtue of the first approach is that it avoids the ad hocness of the second. A drawback is that it is often unknown as to what constitutes, with appropriate generality, the qualification that transforms a defeated generalization into a live one. Attesting to this difficulty is the liberal invocation of *ceteris paribus* considerations. A dubious evasion if ever there were one, retention of the original generalization is made possible only by the expedient of "paying in advance" for unspecified counterexamples. A fourth remedy is the hoary old device of approximation, in which a generalization, though defeated by counterinstance, is retained as approximately true.

Let us consider these in order, beginning with the base case.

- All tigers are four-legged.

Option one provides for something like

- All *properly made* tigers are four-legged.

This is troublesome. If "properly made" here entails "four-legged", the revision is vacuous. If it doesn't entail "four-legged", it is simply useless as things stand how "properly made" achieves the desired exclusions. Of course, various unpackings are possible. We might be invited to consider that properly made tigers are those with the wherewithal to preserve four-leggedness in the descendent class of tigers; but unfortunately this presupposes that all tigers (now) are four-legged, or that one or other of the very reformulations currently under review holds true of them, taking us again too close to circularity for comfort. But circularity aside, the present means of saving this low-order generalization also involves a considerable, and unwelcome, complexity.

A further option gives us

- All tigers, *except those with certain kinds of congenital effects or those injured in certain ways,* are four-legged.

This is also problematic. The trouble is the unspecificity of "certain kinds of" and "certain ways". Left unspecified, there is reason to doubt the generalization's truth. But if the intended specificity is presumed, the generalization is vacuously true. One way of achieving the exceptions without running foul of these difficulties is to list the exceptions, one by one, as in

- All tigers are four-legged *except Pussy, Fred, Baby and Monster.*

But this is hopeless. No one wanting to assert the generalization safely has the foggiest idea as to how the completed list goes.

The *ceteribus paribus* option gives us

- *Other things being equal,* all tigers are four-legged.

Here, too, the unspecificity of "other things being equal" threatens to falsify the generalization, and its specificity threatens to trivialize it. The same is true of

- It is approximately the case that all tigers are four-legged.

If "approximately" means "except those that aren't", we have triviality. If it means something less specific, it cannot be ruled out that it imposes the wrong qualification. It would be a mistake to leave the impression that this brief review of the options is decisive against the reformulation view of defeated Hempelian generalizations. But enough has been said to indicate how difficult and complex such repairs must prove to be. In plain English,

Proposition 7 (The economic advantage of genericity)**.** *Defeated Hempelian generalizations are* hard to fix. *Generic claims with true negative instances* don't have to be fixed.

4.8 Epistemology

Apart from its role in investigating reasoning in its role as an aid to cognition, logic has always carried epistemological presuppositions. Even in the comparatively small historical space of the century just past, one sees the passage from the apriorist, foundational, Platonized realism of Frege and Russell to the pragmatism of Quine, with a concomitant explosion of logical pluralism.[13] But once logic re-adopted *agents* as a central theoretical parameter, it became necessary to pay some degree of attention to what agents are like, to

[13]It may be more accurate to characterize Frege's realism as more Kantian than Platonic. Certainly Frege is not a realist about sets ("courses of values") in the way that Gödel is. Also, it must be acknowledged that as early as 1907 Russell on occasion was quite openly a pragmatist about the justification of "recondite" principles of logic. Strangely, this would later be a position taken up by Gödel. Concerning the first point we are indebted to Ori Simchen for helpful suggestions. Concerning the second, see Irvine (1989). Rodych examines whether Gödel's Platonic ontology is reconcilable with his pragmatic epistemology.

what their interests are and what they are capable of. If logic is to deal with reasoning that advances (or retards) an agent's cognitive agenda, it is necessary that it take note of what the agenda is and how it relates to the agent's wherewithal for advancing it. Any such observation will be incomplete until it is buttressed by an appreciation of the general conditions under which an agent achieves epistemic fulfillment.

If we re-examine various of the conceptual skeins of the new logic, especially in its emphases on defeasibility, non-monotonicity and defaultedness, it can be seen that at present the dominant epistemological presumption is *fallibilism.* Fallibilism is expressly endorsed in the present authors' multi-volume work, *A Practical Logic of Cognitive Systems.*[14] In the present chapter we re-establish that commitment. The idea that real-life cognizers are fallible agents has a certain clear attraction. It expressly embeds the idea of *error* or *mistake,* surely not an irrelevant circumstance for anyone writing about *fallacies.*

4.9 Fallibilism

Fallibilism is a philosophical thesis about error. Since fallacies are errors, it might well be expected that the philosophical thesis that fallibilism is would afford us some insight into the kind of error that fallacy is. Needless to say, the fruitfulness of the connection cannot be guaranteed in advance. It may turn out that there is less to it than we might have supposed. It cannot even be ruled out that there is nothing to it. But if that were so, it would be very odd; it would call out for an explanation.

In its most interesting form, fallibilism is a normative claim. It holds that

Proposition 8 (Fallibilism)**.**

(i) Not only do actual agents sometimes make errors; but

(ii) even when operating at optimal levels occasional error is unavoidable; and yet

(iii) it is wholly rational for a real-world cognitive agent to deploy cognitive strategies (including the adoption of rules of inference) that he (or it) knows in advance will on occasion lead him (or it) into error.

Examples abound. Deductive rules can lead us to false conclusions; inductive strategies can induce the acceptance of defective generalizations; abductive reasoning embodies the risk that attends conjecture; and on and on.

Clause (iii) encompasses two quite distinct notions of error; it is important to give each its due. To mark this difference it helps to take note of another one. It is the contrast between

[14] Of which volume 1 is Gabbay and Woods (2003a) and volume 2 is Gabbay and Woods (2005). Additional volumes will appear in due course.

(a) *Error-elimination* strategies

and

(b) *Error-susceptible* strategies

A good example of an error-susceptible strategy is a default inference from generic premisses. As we have said, a generic claim is a form of general proposition that remains true in the face of (certain classes) of true negative instances. Since a default is a conclusion of an inference in which the "major" premiss is generic, it imbibes this same feature, but in a particular way. Though some classes of negative instances of a true generic claim, *F*s *G*, are also true, it is *not an error* to claim that *F*s *G*, and it is not inconsistent to say that although some *F*s don't *G*, *F*s nevertheless *G*. But given that *F*s do indeed *G* and that *this* is an *F*, we have it as a default that this *G*s. The genericity of "*F*s *G*" allows that "This *F G*s" is false. If so, then the default that is our conclusion in this case *is* an error. This is important. Although, as we have it here, the premisses of the default inference are error-free, and the inference in question is correct, the inference is not of a kind as to preserve freedom from error. So in the absence of information to the contrary,

Proposition 9 (Default inference)**.** *It is reasonable to infer a default from a set of premisses, of which the major is a generic claim and the default an instance of it, notwithstanding that such inferences are not error-avoidance preserving, and that the reasoner is aware of this.*

4.10 Errors of logic

Standard approaches to deductive and inductive logic are wholly concerned with error-elimination strategies. If, as in the case of deductive logic, the error to avoid is invalidity,[15] that error *is* voided whenever the deductive protocols are applied properly. If, as in the case of inductive logic, the error to avoid is inductive weakness, that error *is* avoided whenever the probability rules are applied properly. This carries the suggestion that no such error is possible for any agent who deploys the requisite protocols correctly. Category (b) is different. Its protocols include those for generic inference, as well as various procedures for presumptive and default reasoning. Even if perfectly applied it cannot be guaranteed that they will hit their respective targets. They are, therefore, error-susceptible protocols. This bears on fallibilism in a twofold way. It provides that

(iv) Actual agents are prone (and know it) to applying both error-elimination and error-susceptible strategies incorrectly.

[15] For ease of exposition, we allow invalidity to stand in for the others: inconsistency and logical falsehood.

and it reminds us that

(v) It is insufficient for the cognitive agendas that agents actually have to deploy only strategies of type (a).

Accordingly, not only are actual agents destined to make *application errors,* they are also drawn to the use of strategies whose entirely correct application embodies the occasion of error; in other words, they are also prone to *susceptibility errors.*

It lies at the heart of the present conception of fallibilism that errors cannot be simply "wrong answers". In an extended sense, this is precisely the view that prevails in the error-avoidance precincts of standard logic. It allows us to characterize an argument (i.e., a sequence of propositions) as erroneous simply when it fails to be valid. It allows us to characterize an argument as erroneous simply when it fails to achieve a certain degree of inductive strength. This is plainly not the sense of error that fallibilism seeks to make something of, for then a considerable abundance of perfectly reasonable inferences would have to be classified as errors. What makes this so is that the great percentage of reasonable inferences actually drawn by real-life agents are neither valid (in the sense of deductive logic) nor inductively strong (in the sense of the calculus of probability).

What these conceptions of error lack is an aspect central to the fallibilist approach to the matter. It is the factor of *illusion, inapparency* or *agent-unawareness.* Accordingly,

Proposition 10 (Inapparency)**.** *It is fundamental to the conception of error that an error is a failure or a defect of which its committor is unaware.*

This, to be sure, is the common meaning of the term, as with its near-synonym "mistake". It is a conception that might well irritate those who believe that logic has no business investigating states of mind, but it can hardly be refused by a logic in which a central parameter is the real-life agent. Real-life agents come equipped with states of mind, like it or not. The idea of error as inapparent defectiveness is as old as logic itself. Aristotle expressly advances the notion in *On Sophistical Refutations.* He called them *fallacies.*

Aristotle held that the most general thing to be said about a fallacy is that it is an argument that appears to have a certain property which in fact does not have it. In *On Sophistical Refutations,* Aristotle was more narrowly focused. He wanted to characterize a certain kind of argument in which the notion of syllogism plays an integral role. Aristotle defined a *refutation* as a syllogism whose conclusion contradicts an opponent's thesis and whose premisses are drawn exclusively from the opponent's own concessions. Accordingly, a *sophistical* refutation is an argument thus conceived that seems to be a syllogism but it isn't.

In one of his first tasks as a logician, the founder of logic draws our attention to this phenomenon of false inapparency. In one place, he tells us that it is "the death of argument" (Woods, 2004, Prologue). *On Sophistical Refutations* takes up the task of classifying these bad arguments. Aristotle's list runs to thirteen, though there is reason to believe that he didn't think this an exhaustive inventory. Many pages of this little treatise are given over to brief examinations of where the fault of these bad arguments precisely lies. But no one, least of all Aristotle, thinks that these diagnoses are complete.

It is well to note that in *On Sophistical Refutations* comes close to sharing an assumption with modern formal logic. This is the assumption that the notion of error that these logics adumbrate is one of *deductive insufficiency.* In the case of modern logic, it is invalidity pure and simple. In the case of Aristotle, it is *either* invalidity pure and simple *or* the failure of one or other of the further conditions that Aristotle places on syllogisms. In other words, it is the error of syllogistic invalidity. When one tests this nearly-enough common assumption against actual argumentative practice, it is easy to see that there is something wrong with it. Taking modern logic as our example (it easily extends to fit the syllogistic case), it is no secret that validity is hardly ever an agent's cognitive target. Even in those relatively isolated instances in which a logician wants to know *whether* an argument is valid, producing an argument that *is* valid is neither necessary nor sufficient for the attainment of that target. To illustrate:

1. If an agent X wants to know whether $\langle\{P_1, \ldots, P_n\}, Q\rangle$ is valid, then producing the valid argument $\langle\{P\}, P\rangle$ doesn't hit that target.

2. Neither is it hit just by producing the *very argument* $\langle\{P_1, \ldots, P_n\}, Q\rangle$ (assuming it to be valid); for X may not know that it is valid.

3. X might hit the target by checking the Answers section in a logic textbook. But then he hasn't himself produced anything that is valid, and the answer itself might well consist of the single word "Valid".

Beyond these comparatively rare cases, an agent's cognitive target is not aimed at validity, even though validity may be the requisite *standard* that the attainment of that target may require. If an agent desires a proof of a proposition he will fail unless his reasoning meets the requisite standards, of which validity is one. Clearly, then, one's cognitive target might well be such that it will not be attained unless the validity standard is met. But it is misleading to say that validity is itself the agent's target.

Although the validity standard is sometimes necessary for target attainment, most cognitive targets neither require nor are advanced by fulfillment of the validity standard. We have it then, that invalidity is not, just so, an error, notwithstanding our assumption paragraphs ago that if modern logic had a concept of

error, it could only be invalidity. Invalidity is an error only in relation to cognitive targets for whose attainment the validity standard applies. In so saying, it may occur to us that this is not, in fact, contradicted by the presumptions of modern logic. Whatever its targets, mainstream deductive logic makes it a condition of attainment that the validity standard be met. If this is so, it is largely implicit. It is not much talked about by logicians.

Suppose that we were satisfied with the suggestion that the targets that (however tacitly) call for deductive reasoning require that the validity standard be met. This would be a good place to call attention to an impressive omission.

Proposition 11 (Accounting for error)**.** *Standard deductive logics embed a notion of error, but no such system gives an account of it.*

Why would this be so? Two reasons stand out. One is that, in its subscription to formal languages, standard systems of deductive logic seek to eliminate the linguistic confusions that give rise to fallacies (Frege, 1879; Peirce, 1992; Tarski, 1956; Quine, 1970). The other is that, in as much as deductive logic lacks the capacity to produce a formal theory of invalidity for natural languages, it may be thought that the concept of error lies beyond logic's theoretical embrace (Johnson, 1967; Massey, 1981). We take up these issues in (Gabbay and Woods, 2009).

Targets carry standards for their attainment. Something is an error if it fails to meet the required standard. Again, not just any valid argument will meet the validity standard of every cognitive target that embeds a validity standard. Speaking this way relativizes standards to targets and imposes the same relativity on the concept of error. One can only wonder whether these things might not be subject to further relativities. The answer is that they are.

An agent might wish to know the proof of the completeness of formal arithmetic. If so, he would have made an error. His target is defective in a quite particular way; it embodies a false presupposition. An agent might set himself the target of acquiring a Ph.D. in quantum computation. But if he is 92 years of age, a high school drop-out, and possessed of a modest I.Q., he too has made a mistake. It is not that the target of getting a Ph.D. in quantum computation is impossible to attain, but rather that it is impossible *for him* to attain. It was the wrong thing to aim for, given this agent's cognitive resources. Here, then, is another pair of factors that bear on the issue of error.

Proposition 12 (Error relativity)**.** *Something may be an error in relation to the standards required for target attainment, in relation to the legitimacy of the target itself, or in relation to the agent's cognitive wherewithal for attaining it.*

4.11 Parameters of the subpar

Let us tarry awhile with this idea of subpar cognitive performance. So again we ask: What is it to judge that someone's cognitive conduct is not up to snuff? It is to find fault with the action in the light of various criterial considerations. As we saw, one is what the agent's target is. Another is the standard that he needs to hit for that goal to be attained; in other words, the agent's means to that end. A third factor in judging an agent's cognitive performance is his *general competence.* In mentioning it, we reveal an interest in determining whether this is a goal whose satisfaction by hitting that standard is something that he is able to do. A fourth consideration has to do with *collateral considerations.* An agent may have the general capacity to achieve a certain goal in a certain way, but, owing to present particularities, not be able to achieve it or to achieve it in that way. In citing this factor, we are recognizing the importance, beyond general competence and means-end adroitness, of cognitive resource-contingencies such as (again) *information, time* and *computational capacity.*

Jointly, these factors give a blueprint of an agent's performance of a cognitive task. A cognitive target T is either attainable or not. (A proof of Fermat's Last Theorem is attainable; a proof of the joint consistency and completeness of Peano-arithmetic is not.) If a goal is attainable, then for any agent X, it falls within X's general competence or not. (A proof of the completeness of modal logic was within Ruth Barcan's reach but not, we may suppose, Hannah Arendt's.) If X has an attainable goal that lies within her general competence, the means she selects (or the standard she sets) may be appropriate for that goal or not. If, for example, X undertakes to show for some proposition P that P is something that might reasonably be believed, her standard may include an argument for P that meets the standard of validity. In her quest to justify a belief in P in this way, X would be at risk for two performance errors. Either validity may be an inappropriate way of achieving this goal, or it may be appropriate but beyond X's reach. X might not know how to construct valid arguments (perhaps she is a struggling student of First Year logic). If X has an attainable goal that is within her general competence, for which an appropriate means S is also within her grasp, X may lack additional resources R necessary for the completion of her task. She might not have information enough to command the desired means; or she may lack the time to achieve her objective in this way; or she may lack the computational power to do the calculations that her task requires of her. Alternatively, given the comparative scantness of such resources for real-life individuals in actual situations of cognitive effort, an agent may simply lack the means of achieving the goal. If, again, an agent's goal is to show that it is reasonable to believe that P, she may decide that an axiomatic proof of P is not a means for which she is adequately resourced at present; and she might try instead for a conditional proof relative to what is

widely held by experts (wherewith the potential for *ad verecundiam* error may present itself).

We may say, then, that

Proposition 13 (Further relativities)**.** *There are several basic ways in which an agent X's cognitive performance can go wrong:*

1. *X might set himself a simply unattainable target T.*
2. *X might set himself an attainable T that is not within his general competence.*
3. *X may set himself an attainable T for which he is generally competent, but his selected means (or goal-realization standard) S is either beyond his reach or inappropriate to the task at hand.*
4. *X may set himself an attainable T for which he is generally competent and set himself an appropriate S that lies within his reach, and yet he might lack necessary collateral resources R.*

When this last condition is met, we shall say that T is an attainable goal for which X is generally competent, that S is a realizable and appropriate means for X to set in relation to T, but that for lack of such things as information, time and fire-power, T sets a task that is *too big* for X.

Ed Koch, on his walking tours of New York when he was mayor, famously would ask, "How am I doing?" We daresay in inviting this assessment of his performance as chief magistrate, he was unaware of all the details of the template that structures a fair response. It is a template that calls for the assessment in terms of T, S, R. These are the structural elements necessary for a finding of "subpar" with respect to the ranges of cognitive performance that draw the attention of fallacy theorists. They apply to Ed Koch. And they apply to the rest of us as well.

4.12 Ought and can

No practical agent can be faulted for mismanaging a cognitive task that is too big for him, although he might well bear some responsibility for having acquiesced to such a task. Whatever we say about such (mis)performances, they are not fallacious. In *some* sense, a principle of "ought"-implies-"can" is at work here. There is, however, a certain confusion that we should try to avoid. In saying that a better performance is not possible for agents of type so-and-so, it is not always required that we deny its betterness. It is required only that we resist the inference that a possible performance that is less than better for agents of this type is *subpar for them.*

There is in these reflections occasion to consider a sister principle to "ought"-implies-"can". We could call it "can"-doesn't-imply-"ought"; it has

the virtue of being in general even more obviously true than its kin. It is not, however, trivially or vacuously true; for especially in enquiries into human cognitive performance, exceptions to it are expressly countenanced, some having the status of scientific postulates. In any account of human practice in which optimization is held to trump satisficization, and it is also assumed that it is always better to do one's best, that "can"-doesn't-imply-"ought" is conspicuously disregarded. Variations of its opposite, "can"-implies-"ought" flourish in standard accounts of belief dynamics and rational decision-making (Alchourron et al., 1985; Raiffa, 1968). However, it is well-attested in actual practice that practical reasoners often saticifice rather than optimize, even when optimization is available to them as an achievable goal. In such practice there is an important reciprocity between targets and standards. What a cognizer needs to know and how he sets about to know it is a matter of what the knowledge is wanted for. Peirce once quipped that we know who our parents are by hearsay. Given the documentary thoroughness of modern life, to say nothing of the identificatory capacities of DNA technology, one could know more of one's parentage—and know it more strictly—than the run-of the-mill offspring has (as the saying goes) "time for". It is not that this larger and more strictly realized knowledge exceeds his reach. In the general case it exceeds his cognitive goal (to know whom to call "Mum" and "Dad") and imposes a cognitive standard that he has no need of. For ranges of cases, "can" clearly does *not* imply "ought". When an agent pursues a target or a standard, or both, that is bigger than it need be, we shall say that their pursuit by that agent is a case of *overkill.*

Before leaving the suggestion that a version of "ought"-implies-"can" holds for the assessment of cognitive performance, care needs to be taken not to trample on the latitude underwritten by fallibilism. If fallibilism provides that there are cognitive procedures that it is rational to execute even in the knowledge that they are virtually certain to lead one to occasional error, and if it also holds that there is a sense in which such procedures can't be abandoned, then fallibilism allows for a conception of error that a reasoner can't help committing or can't help committing without cost to his procedural rationality. So we must not allow the sense in which "ought" implies "can" to trespass on this provision.

Consider now a real-life individual who has set himself the task of advancing his cognitive agendas—of living his cognitive life—on the model of NASA. Given his resources and the loftiness of his cognitive ambitions, his cognitive life is a guaranteed disaster. Cognizing on the model of NASA is too big a task for any individual. In one sense, it is quite right to "forgive" *X* his cognitive failures. One can't be expected to achieve what one hasn't the means of achieving. Even so, *X* didn't have to set his targets so high. It was well within his power to select his targets with a view to his ability to meet them. If this is

so, his massive failures are subject to disapproval of higher order. They were the inevitable outcome of unrealistic targets that he needn't (and shouldn't) have pledged to.

Be that as it may be, there still remains the utterly central question of whether, and to what extent, an agent—any agent of whatever type—can be held responsible for an *error,* given that an error is something that he cannot, then and there, see as such. Take a case. Let *X* now set his targets more realistically. Let us say that they are of a type for which he has the requisite competence and the necessary resources. They are not too big for him. Even so, we have it by the very idea of error that if *X* errs in his quest to attain *T*, his error is something inapparent to him. And we have it by the meaning of fallibilism that the best that is in *X* rationally to do involves him in cognitive procedures that will on occasion expose him to error, that *X* knows this; and that knowing it is no affront to his reasonableness in retaining those very procedures. Against this, there is a strong disposition to find fault with at least those errors that have acquired membership in *GOE.* As Douglas Walton has it, attributing such an error to *X* is one of the harshest criticisms that can be leveled at *X*'s performance (Walton, 1995). The literature also embeds the widely-held view that fallacies are errors of a kind made avoidable by due care. But, as we see, neither of these views rests well with any view on which errors are undetectable, especially when such a view is embedded in a fallibilist epistemology.

4.13 Inapparency

On the face of it, a theory of fallacy has a twofold task. Since a fallacy is an error, a theory of fallaciousness should embed an account of error. Since a fallacy is an inapparent error, a theory of fallaciousness should contain an account of the factor of inapparency. There is, to be sure, an element of redundancy in putting it this way, since inapparency is intrinsic to error. Accordingly, a theory of error would also have the task of dealing with inapparency. But there is no harm in listing the inapparency requirement as a separate theoretical responsibility, if only to lend it an emphasis to which the literature is largely inattentive.

Inapparency, then, is intrinsic to error. In committing an error, there is something its committor has *over-looked,* something that he has failed to *see.* It bears on this that in its most common meaning a fallacy is a "common misconception", a belief which, although false, is widely and confidently held. It is an attractive belief whose falsity has escaped the committor's attention. The psychological literature draws a useful distinction between *performance* and *competence* errors. A performance error arises from contingent factors such as fatigue, intoxication or intention. Competence errors spring from more structurally embedded kinds of inability. If a good night's sleep might arouse a

reasoner from yesterday's performance errors, it will do him no good on the score of incompetence. These are transgressions whose avoidance exceeds the very design of the committor's cognitive wherewithal. A particularly good example of a competence error is one that arises in the treatment of a problem whose solution requires an effort that exceeds the computational capacity of the type of agent in question. Competence errors are not, however, a particularly good example of the sort of inapparent misstep we are currently discussing. The reason for this is that

Proposition 14 (Abiding competence)**.** *It is a compensation strategy among beings like us to tend to avoid the employment of cognitive protocols that exceed their competence.*

A case in point: An exhaustive check of our present web of belief for truth functional consistency would involve us in a computational explosion vastly beyond the reach of what we are built for. But there isn't the slightest empirical evidence that, when beings like us *do* attempt to reconcile their beliefs to some standard of consistency, this involves anything like even an exhaustive search.

A further locus of inapparency has been held to be the argument (or piece of reasoning) itself. So seen, an argument (or inference) that we erroneously pledge to (or erroneously draw) is one whose defectiveness is inapparent even to a well-rested and competent cognitive agent, arising from a kind of camouflage or disguise. Needless to say, these are rather anthropomorphic metaphors, having a more literal application in cases of an interlocutor's intention to deceive his opponent. But the factor of disguise is, on this view, lodged not in the committor's malign intention but rather in his warp and woof of argument or inference. Of the many theorists who subscribe to such a view, perhaps it is Lawrence Powers who puts the point most clearly:

Proposition 15 (Powers' inapparency principle)**.** *The false inapparency of an erroneous argument or inference is an* objective feature *of the argument or inference, rather than an* interactive feature *of them with a cognitive agent (Powers, 1995).*

We leave it to Powers to identify those objective features. We ourselves are minded to look elsewhere—to the very structure of cognition itself—for an especially important and, in its way, objective, locus of false inapparency.

Let us observe that in one of its most common meanings the word "believe" (and its cognates) admits of a striking first-third person asymmetry. On this usage, when *Y* says of *X* that *X believes* that *p*, *X* would say of *himself* that *p*. Legions of philosophers have been right to observe that *self*-ascription of belief constitute a kind of attenuated or qualified subscription to the proposition at hand. But in the present meaning of the term, the *other*-ascriptions of a belief that *p* leave it entirely open that the person to whom the belief is attributed

holds to p (and is right to) assertively and without qualification. Accordingly, for the sense of "believes" in question,

Proposition 16 (Belief as knowledge-claim)**.** *Whenever it is true for Y to say of X that X believes that p, it is true that X takes himself as* knowing *p* to be true.

Proposition 16 is a blindspot context (Sorensen, 1988). Whenever it is true for Y to say of X that X believes that p, then for X to say of himself

- p, and I believe p

would constitute a *blind-spot.* That is to say, in the absence of further information, any person to whom the bulleted admission were directed would lack the means to ascertain just what the utterer's epistemic state toward p has been claimed to be. Is the utterer saying that he knows that p? Or is he saying that he (merely) thinks that p?

Consider now an agent X's cognitive target K. Suppose that K is such as to be attainable only when X is in an epistemic state k. Let k be the state in which it is true to say that X knows that p. X's target T is occasion of a kind of *cognitive irritation.*[16] X is so constituted and so related to T that he aspires to be in a state in which the irritation is relieved.[17] We have known at least since the presocratics that although being in k is the state that X is required to be in for T to be attained, it is *not* required for X's cognitive irritation to be relieved. Irritation-relief is one thing. Cognitive attainment is another. From the third-person perspective, this is not a difficult contrast to command. But from the first-person perspective, it is a contrast that collapses, and is recoverable if at all only in the person's own reflective aftermath. When that reflective aftermath is at hand, the first-person can now say what the third-person could have said all along: X only believed that p, rather than knowing it. We have it, then, that when X is in a state of belief that relieves the cognitive irritation occasioned by T, he is in a state which he takes to constitute attainment of T. Not only is that state, b, *not* the same as k, but X's being in k, carries no phenomenological markers over and above those carried by b. Accordingly,

Proposition 17 (Phenomenologically structured inapparency)**.** *By the phenomenological structure of individual cognitive agency, the difference between being in b and being in k is phenomenologically inapparent. So where one indeed is not in k, being in b disguises that fact.*

[16] We must take care with the metaphor of irritation. Not every irritation of the human system that is put right by the requisite causal adjustments is something the human agent is either conscious of or openly desirous of remedying. Given that cognition can be so deeply implicit, we require the same latitude be extended to the idea of cognitive irritants.

[17] Such aspirations flow from what St. Augustine calls "the *eros* of the mind". In Gabbay and Woods (2005) it is called "cognitive yearning".

If this is right, then the capacity for, indeed the likelihood of, false apparency is structured by the phenomenology of cognitive states. For one thing, it seems not so much to be a property of a given argument or a given piece of reasoning, but rather a factor intrinsic to the possession of *b*-states in relation to *T*'s that call for attainment by way of *k*-states. It bears repeating that cognitive *relief* is not, just so, cognitive *attainment;* it is rather the appearance of it. Certainly in our disposition to confuse relief with attainment, there need be not the slightest hint of fatigue or intoxication. In other words, our present confusion seems not to be, or to arise from, performance errors. Given that such confusions appear to be intrinsic to the phenomenological structure of cognitive states, it lies more in the ambit of the competence error, hence reflective of an objective fact about how individually cognitive agents are constructed.

4.14 Valuing validity and inductive strength

Let there be no doubt, when truth-preservation is indeed an agent's cognitive target, validity is a necessary part of the standard for its attainment. However since truth-preservation does not, just so, guarantee the proof of anything,[18] truth-preservation rarely achieves the status of cognitive target, and rightly. In realistic settings, truth-preservation is itself valued not as a target but as a standard. In other words, in realistic settings, truth-preservation and validity are the *same* standard.

Valuable though it is in some settings, it is easy to think too much of validity; at least this is so when validity is monotonic. Let T be a target that calls for a valid argument. Let V be such an argument. Let K be a proposition that contradicts V's conclusion and is not in V's premiss-set. Let us also put it that the discovery of K is a huge surprise for X. Let V^* arise from V by addition of K as premiss. Since V is valid, so is V^*. But it is clear that although V^* is valid, it is not of the slightest use to X. It is not of the slightest use notwithstanding that it is a valid argument retaining all the premisses of V, which, until the discovery of K, we may suppose to have been of considerable use to X. For it was a valid argument none of whose premisses is a proposition that X then had any reason to doubt. What we see, then, is that *validity-preservation* is not a realistic standard even for targets for which validity is a necessary standard.

Validity is unresponsive to new information. In this respect, it is natural to suppose that *inductive strength* is the more useful standard. Its usefulness is a matter of its *non*-monotonicity. Its non-monotonicity makes it responsive to new information. This is true but not especially availing. Let I be an argument whose conclusion C has a requisite degree of conditional probability given its

[18] Save for the corresponding conditional of the argument that the target's attainment standard requires to be valid.

premisses $P_1, \ldots, P_n$. I is an inductively strong argument. Suppose now that K is new information that falsifies C. Since K is new, it is not in I's premiss-set. Let I^* arise from I by addition of K as premiss. Notwithstanding that I is inductively strong, I^* is inductively impotent. It is clear that, even though new information can collapse inductive strength, there is an inductively strong argument available to X that is wholly untouched by the new information. This is argument I, and the reason that it is wholly untouched by K is that K is not in its premiss-set. What this tells us is that, even where inductive strength is part of a target's attainment standard, it is a smaller part than might have been supposed. As we now see, validity-*preservation* is not part of the standard of any target whose attainment calls for validity. The reason for this is that validity provides it automatically. Validity-preservation is a free-rider. But with induction we may say that the reverse is true. That is to say, given any target for which inductive strength is part of the attainment standard, preservation of inductive strength in the face of new information is also a requirement. It is easy to see that this latter imposes on an agent's inductive targets the weightier requirement that the inductions be made from up-to-date information, i.e., that they not admit any information that collapses inductive strength. In the inductive cases, falsifying new information matters *inductively.* In the deductive cases, falsifying new information does not matter *deductively.* In both cases, however, what matters more is the *state of the information* from which conclusions are drawn.

PART II: FALLACIOUS COGNITIVE VIRTUES

This would be a good place to restate our principal theses about the fallacies.

Proposition 18 (The no-fallacy thesis)**.** *Not all of the Gang of Eighteen are fallacies. Those that are are not characteristically committed by beings like us.*

Proposition 19 (The cognitive virtue thesis)**.** *Several of the Gang of Eighteen are cognitively virtuous scant-resource compensation strategies.*

In what remains of this essay, we shall attempt to vindicate these claims as they apply to hasty generalization.

Limiting the defence to just one might well strike the reader as favouring our cause with an artificially small sample. But the reason is the want of space.

4.15 Hasty generalization

Hasty generalization, also known as *thin-slicing*,[19] is an error when committed in response to a cognitive target T, whose attainment embeds the standard S of inductive strength. For example, T might be the goal of reaching a

[19] See Ambady and Rosenthal (1993) and Carrere and Gottman (1999).

generalization about some subject with scientific accuracy. In that case, it is reasonable to require that his (or its) reasoning rise to the standard of inductive strength. It is easy to see that it is comparatively rare for individual agents to set targets of such loftiness. If an agent is part of a drug assessment team for Health Canada, we would certainly expect him and his colleagues to set themselves such a T and bind themselves to such an S. But an indication of how comparatively rare this, even for this individual, is the comparatively generous command he enjoys of Health Canada's resources for T—time, information, computational power, money, infrastructural and cultural encouragement, and so on. To the extent that this is so, this person and his mates are not acting as practical agents. They have teamed together and they have attracted levels of support in ways that give their efforts the kick of theoretical agency. Most practical agents lack the rudiments of scientific method, whether knowledge of how to compose a stratified random sample or of how to calculate even low-level conditional probabilities. What is more, if they did know, it would in very large ranges of cases be beyond what they had either time or computational capacity for (Harman, 1986). There is a widely received view that all of this is true but beside the point. For even practical agents (it is said), limitations and all, are performing at their ampliative best when they strain against these limits and approximate to the behaviour called for by the methods of science.

This, of course, is *scientism.* Saying so doesn't take us much beyond name-calling. So something further must be said against the view that in matters ampliative it is best to conform one's reasoning to the requirements of induction. In preceding sections, we have given out part of what we take to be the correct treatment of hasty generalization. We have seen that when one generalizes hastily, one often generalizes to a generic proposition rather than to a universally quantified conditional proposition (full-bore Hempelian generalizations, as we called them). One of the chief virtues of proceeding in this way is that even when as instantiated default is false, it is necessary to forgo the instantiation but not to repair the generic generalization whence it sprang. There is a considerable economy in this, needless to say; and that alone vests it with an attractive advantage. A further point of importance—perhaps the fact of dominating significance here—is that even when we seek the lofty goals of scientifically pure induction, we tend to generalize hastily. In beings like us, hasty generalization is as natural as breathing. The compliant scientific methodist must struggle to stifle what his cognitive nature has already made him believe. Doing so takes effort (and often time); so costs are necessarily levied.

Generic summations do not exhaust the class of non-universal generalizations. Normalic generalizations, of which statistical generalizations are a particular case, also figure prominently in ampliative reasoning. Normalic generalizations are generalizations about what is the case nearly always, or for the most part. There is a use of the word "normally" which is a synonym of

"usually", which our term "normalic" draws upon. Unlike generic generalizations, normalic generalizations embed quantifiers. This is not everyone's understanding of genericity. But in light of the fact that some claims of the form "*F*s *G*" are true and "Most *F*s are *G*" is false (Carlson and Pelletier, 1995), we think it the correct understanding. Genericizations lack a quantificational organization precisely where normalic generalizations have it essentially. It is an important structural difference, carrying interesting semantic consequences. Whereas "This *F* doesn't *G*" can be true without "*F*s *G*" ceasing to be true, it remains the case that "This *F* doesn't *G*" is a negative instance of "*F*s *G*", albeit a true one. Yet "This *F* doesn't *G*" doesn't come close to being a negative instance of "Nearly all *F*s *G*". How to fill in these semantic differences is still an open question in the logic of general propositions. Interesting and important though the question is, we shall not press it here. It suffices to note that

Proposition 20 (Variable generality)**.** *Thin-slicing carries no* intrinsic *tie to types of generalization.*

Accordingly, one may hastily generalize to Hempelian generalizations, generic generalizations and normalic generalizations. We have pointed out the advantages of genericizing over Hempelianizing. Like advantages attach to normalicizing rather than Hempelianizing. In each case, the truth of propositions in the form "This *F* doesn't *G*" needn't disturb the truth of the respectively generalization. This leaves the question as to what would differentially motivate generic and normalic thin-slicing. The answer, broadly speaking, hinges on the element of defectiveness. Negative instances of generic thin-slicing are in some or other way defective cases of the subject term. There is no such assumption to be made in the case of normalic thin-slicing.

Normalic thin-slicing is but one example of judgements of non-universal quantification. If we allow that "Nearly all" as a quantifier, "Hardly any" cannot be denied the same recognition. The difference between "Nearly all" and "Hardly any" mimics the difference between "*n*% of" and *m*% of", where *n* is quite large and *m* is quite small. So statistical projections also have the general character of non-universal quantification.

We see in these similarities and differences an important moral.

Proposition 21 (Low non-universality)**.** *"Fs are hardly ever G" is as much a case of thin-slicing as is "Fs are G" or "Fs are nearly always G" when drawn from a small (enough) sample.*

Thin-slicing is largely automatic. To a considerable extent, it is part of what goes on down below. Hasty generalization is also a belief-forming device; and, as we have seen, belief from the inside perspective manifests itself as knowledge. This would be an epistemic disaster if the hasty generalizations we actually are drawn to make were always or frequently mistaken. If so, we would

be massively mistaken in what we are induced to think that we know. What is so striking about hasty generalizations, as they are drawn in real life by beings like us, is that they are by and large right, or right enough to allow us to survive and prosper, to contribute to the replication of our cognitive devices in the human descendent class, and occasionally to build great civilizations. So we may say that

Proposition 22 (The naturalness of hasty generalization)**.** *The hasty generalizations actually drawn by practical agents are cheap, irresistible and typically accurate enough to* fulfil *our interests.*

We may hypothesize that the capacity for generally accurate generalizational haste is something that is hard-wired into beings like us, or that, in any event, it is so primitive a skill that it must have been part of the yield of our earliest learning. It doesn't matter. Once the human individual is past his early infancy, his life is saturated with generalizations that are both hasty and accurate, and, when not accurate, efficiently corrigible. It is tempting to speculate that it all springs from the mechanisms of flight and fight. Perhaps this is so. But, again, what matters for the logicians are not the causes of such haste, but the cognitive utilities of it.

For this unfolding apologia to be defensible, it must be the case that

Proposition 23 (Practicality and haste)**.** *The extent to which an agent is operating practically, is not by and large appropriate that his targets be such as to impose the standard of inductive strength.*

Let us pause to consider the view that we are trying to dispel.

1. Cognitive rationality is the system of thought prescribed by the deductive and inductive logic and decision theory.

2. Human beings are naturally so constituted that they think in ways that closely approximate to the canons prescribed by these systems.

3. Accordingly, a theory of rationality should provide an account of how the state of affairs stated by (2) came to mirror the norm expressed by (1).

Our position is that the norm embodied in (1) is no norm and the fact expressed by (2) is no fact.[20] If we want to be right in our rejection of the norm purported by (1), we must discourage the idea that beliefs sanctioned by the standard of inductive strength constitute a kind of global maximum. But, as we have already pointed out, there are reasons to doubt any such claim. Unlike

[20]Matthen shares our scepticism about (2). He is rather more equivocal about (1). See Matthen (2002, Section 6).

(classical) validity, which is wholly impervious to new information, inductive strength is a veritable sitting duck. We can see this in an especially dramatic way when *C* is a generalization and *E* is a sample. Like the universally quantified conditional construal of generality, the property of inductive strength is highly brittle. Let a given such argument be as inductively strong as may be. If the next bit of information is a counterexample *N* to *C*, the original argument remains inductively strong and the result of supplementing its premisses by addition of *N* is an inductive disaster. What this shows is that the inductive strength of the original argument was no reason to think well of it, whereas the catastrophe engineered by the present argument invests over-heavily in freedom from counterexample in inductive contexts. Thus the norm embodied in (1) can't be relied upon unless accompanied by reasonable assurances of the non-existence of counterexamples. But this asks more from ampliative reasoning than it can possibly be expected to provide.

It is instructive to compare ampliative reasoning in an individual's hands and in NASA's. NASA's targets are such that it must pay for its counterexamples with disasters. When an *N* comes along that topples a *C*, all bets are off until, with considerable elaboration, *C* is reframed so as to tolerate *N* or *N* is reformulated to take the pressure off *C*, or *C* is abandoned and hopes for a happier successor are launched. In actual practice, these accommodations are often very difficult and very expensive. Individuals by and large simply aren't up to these levels of disaster-management. Accordingly, individuals do not typically repose their ampliative burdens on so fickle a standard as inductive strength. Rather they show their fondness for genericity and the like, which in turn is an invitation to make do with small samples. This makes a nonsense of inductive strength, needless to say. But it gives the practical reasoner a form of ampliation that serves him well and that he can afford. For, again, he is not typically wrong in the generic claims he wrests from small samples with such haste; and when he is wrong, i.e., when a true negative instance *N* does present itself, he is not, just so, faced with the burden and the cost of repairing *C*. As we said, *C* is elastic; it can remain true in the face of true negative instances.

4.16 Risk aversion

Hasty generalization genericizes or quantifies from small samples. Doing so would clearly be defective if the samples in question were unrepresentative. In the literature on inductive logic, it is common to require of an agent that, before he generalizes from a sample, he check it for, or otherwise assure himself of, its representativeness. This is true but unhelpful. In generalizing from a small sample, a reasoner implies that the sample is representative. To make it a condition on such generalizations that they be grounded in the conviction that the sample is representative is to require him to withhold his generalization until he thinks that he has reason to think it correct.

What counts here is that thin-slicers—that is, all of us—are adept at discerning representative samples among the very small. We have already made the point that our facility with sample representativeness is linked to our facility with natural kinds. Doubtless this is so, but it doesn't amount to much of an explanation. Better that we explore the link with our danger-recognition capacities. Hard-wired or not, one of the most primitive and successful of an individual agent's endowments is the wherewithal for the timely recognition of danger even in the face of utterly scant evidence of it. The attendant protocols of risk aversion are concomitantly *conservative.* They risk the effort of unnecessary evasions for the advantage of securing against the greater liabilities that attach to the contrary. The flight-fight mechanisms of beings like us are activated by factors of apprehensiveness; fear is the third 'f' in this trio. They are mechanisms that embed the fundamental structure of thin-slicing.

The fear factor is crucially important. When an individual runs from the unknown creature with large fangs, it is not at all necessary to attribute to him the tacit belief that such creatures are lethal biters but rather the anxiety that they *might* be. Risk aversion turns on epistemic estimates of comparatively low yield; not on the conviction that *F*s *G* but on the worry that *F*s might *G*. Behaviour is risk-averse in this conservative way precisely when it grounds non-trivial action on so slight and tentative an appreciation of what is the case. We may see thin-slicing as an adaptation of conservative rise-averse behaviour, in which the element of fear is replaced by that of belief and the estimate of mere possibility is upgraded somewhat. Even so, the basic structure is retained. When on the strength of a small sample one reasons that *F*s *G* or (most do or few do), one is tendering the projection with a requisite tentativeness. But if this is so, thin-slicing cannot be judged by the standard of inductive strength.

4.17 Probabilistic reasoning

Given that an argument is inductively strong to the extent that its conclusion is made more likely by the evidence cited in the premisses, a number of additional assumptions are the life's blood of mainstream inductive logic.

1. Likelihood is *probability.*

2. The relation of greater (or less) likelihood relative to a body of evidence is the relation of *conditional probability.*

3. The concepts of probability and conditional probability are accurately described by the theorems of the *probability calculus.*

4. Any set of premises that increases the conditional probability of a proposition also confers some positive degree of *confirmation* on it.

We have tried to make plain that the inductive strength standard is neither appropriate nor required for a practical agent's cognitive targets by and large. What would count against this claim? Here is a point that might give us pause. Everyone agrees that practical agents have an impressive command of probabilistic reasoning. Suppose it turned out that the present assumptions are true, and that actual probabilistic reasoning comported with them. If these things were so, our real-life probabilistic reasoning would satisfy conditions under which probabilistic success would indeed hit the standard of inductive strength. Clearly, we must say something about probabilistic reasoning.

If the behaviour of individual agents is anything to go one, then the standard accounts of inductive inference constitute significant distortion of the actual record. Can the same be said for the linked issue of probabilistic reasoning in the here-and-now? James Franklin sees in probability an interesting parallel with continuity and perspective (Franklin, 2001). All three of these things took a long time before yielding to mathematical formulation, and, before that happened, judgements of them tended to be unconscious and mistaken. We have a somewhat different version of this story. Sometimes a conceptually inchoate idea is cleaned up by a subsequent explication of it. Sometimes these clarifications are achieved by modelling the target notion mathematically. Sometimes the clarification could not have been achieved save for the mathematics. We may suppose that something like this proved to be the case with perspective and continuity. To the extent that this is so, anything we used to think of these things which didn't make its way into the mathematical model could be considered inessential if not just mistaken. It is interesting to reflect on how well this line of thought fits the case of probability.

In raising the matter, we are calling attention to two questions. (1) What was probability like before Pascal? (2) How do we now find it to be? Concerning the first of this pair of questions, We think that we may suppose that, in their judgements under conditions of uncertainty, people routinely smudged such distinctions as may have obtained between and among 'it is probable that', 'it is plausible that', and 'it is possible that'. If we run a strict version of that line over this trio, then not making it into the calculus of Probability leaves all that is left of these blurred idioms in a probabilistically defective state. There is a sense in which this is not the wrong thing to conclude, but it is a trivial one. For if what we sometimes intend by 'probability' fails to find a safe harbour in the probability calculus, then it is not a fact about probability that the probability calculus honours. But unlike what may have been the case with perspective and continuity, we must take care not to say without further ado that those inferences that don't make the Pascalian cut are mistakes of reason or even mistakes of probabilistic reason. In this we cast our lot with Cohen (1982) and Toulmin (1953) albeit for somewhat different reasons. With Cohen we agree that some of the Kahneman and Tversky (1974) experimental

results which show their subjects to have been bad Pascalians do so only if they had undertaken to be good Pascalians. The alternative, of course, is that, even though they were invited to be Pascalians and primed to make a workmanlike job of it, their sole mistake is that they slid unawares into a non-Pascalian disposition toward reasoning under conditions of uncertainty. Certainly had they been drawn to the task of compounding *plausibilities,* it is far from clear that the Kahneman-Tversky results show their efforts in a bad light.

We side with Toulmin in saying that not all judgements of probability, even when made by working scientists, express or attempt to express the concept of aleatory probability or to comport with its theorems. A similar moral can be drawn from the sheer semantic sprawl of the idioms of possibility.

Let us take it that, unlike perspective and continuity, idioms of probability (or probability/plausibility/possibility) that don't cut the Pascalian mustard leave residues of philosophically interesting usage. If this were so, there might well be philosophically important issues, the successful handling of which requires the wherewithal of this conceptual residue. Again, standard answers to Kahneman-Tversky questions don't cut the mustard of aleatory probability, but they do comport with conditions on plausible reasoning. What, then, are we to say? That these bright, well-educated subjects are Pascalian misfits or that they are more comfortably at home (though unconsciously) with a plausibility construal of their proferred tasks? If we say the second, we take on an onus we might be unable to discharge, or anyhow discharge at will. It is the task of certifying the conditions under which these non-Pascalian manoeuvres are well-justified. In lots of cases, we won't have much of a clue as to how to achieve these elucidations. Small wonder, then, that what we call the Can Do Principle beckons so attractively. This is the principle that bids the theorist who is trying to solve a problem *P* to stick with what he knows and to make a real effort to adapt what he knows to the requirements of *P*. One of the great attractions of Pascalian probability is that we know how to axiomatize it. Can Do is right to say that it would be advantageous if we could somehow bend the probability calculus to the task to hand. But sometimes, the connection just can't be made.

Bas van Fraassen is spot on in pointing out that there "has been a sort of subjective probability slum in philosophy, and its inhabitants, me included, have not convinced many other philosophers that what happens there is anything more than technical self-indulgence" (van Fraassen, 2005). This calls to mind our *Make* Do Principle, which is the degenerate case of Can Do. Make Do is just Can Do in circumstances in which the fit with *P* cannot be achieved satisfactorily. If *P* is the problem of avoiding "the naïveté and oversimplification inherent in much of traditional epistemology" (van Fraassen, 2005), then a decision to deploy the theory of probability by brute force would be a case of Make Do. It would capture the mood of the tasker who, not knowing what to

do about P, settles for he knows how to do about Q, and wholly ignores that it is all beside the point.

4.18 The link to abduction

If what we have been saying about thin-slicing is correct, hasty generalization bears a significant resemblance to abductive reasoning. Abductive reasoning is a response to an ignorance-problem. One has an ignorance-problem when one has a cognitive target that cannot be attained on the basis of what one currently knows. Ignorance problems trigger one or other of three responses. In the one case, one overcomes one's ignorance by attaining some additional knowledge. In the second instance, one yields to one's ignorance (at least for the time being). In the third instance, one abduces. The general form of an abductive inference can be set out as follows, putting T for the agent's target, K for his (or its) knowledge-base, K^* for an accessible successor-base of K,[21] R as the attainment relation relative to T, H as the agent's hypothesis; $K(H)$ as K's adaptation of H, and R^{pres} as the relation of presumptive attainment relative to T:

1.	$\neg R(K, T)$	[fact]
2.	$\neg R(K^*, T)$	[fact]
3.	$R^{pres}(K(H), T)$	[fact]
4.	Therefore, $C(H)$	[conclusion]
5.	Therefore, H^c	[conclusion]

What the schema tells us is this: T cannot be attained on the basis of Q. Neither can it be attained on the basis of any successor K^* of K that the agent knows then and there how to construct. H is an hypothesis such that when reconciled to K produces $K(H)$. H is such that if it were true, then $K(H)$ *would* attain T. But since H is only hypothesized, its truth is not assured. Accordingly we say that $K(H)$ *presumptively* attains T. That is, having *hypothesized* that H, the agent *presumes* that his target is now attained. But since presumptive attainment isn't attainment, the agent's abduction must be seen as preserving the ignorance that gave rise to his (or its) ignorance-problem in the first place. Accordingly, abduction is not a *solution* of an ignorance problem, but rather a *response* to it, in which the agent settles for presumptive attainment rather than attainment. $C(H)$ expresses the conclusion that it follows from the facts of the schema that H is a worthy object of conjecture. H^c denotes the decision

[21] K^* is an accessible successor to K to the degree that an agent has the know-how to construct it in a timely way; i.e., in ways that are of service in the attainment of targets aimed at K. For example if I want to know how to spell "accommodate", and have forgotten, then my target can't be hit on the basis of K what I now know. But I might go to my study and consult the dictionary. This is K^*. It solves a problem originally aimed at K.

to release H for further promissory work in the domain of enquiry in which the original ignorance-problem arose. The superscript is a label. It reminds us that H has been let loose on sufferance. (For an exhaustive discussion of abduction, see Gabbay and Woods, 2005.)

Abductions are a response to ignorance-problems intermediate between solving them and being defeated by them. Like the latter, successful abductions do not solve the ignorance-problems that give them rise. Like the former, abductions authorize (albeit defeasibly) subsequent actions that the agent may well have preferred to have seen grounded in a solution to his problem, rather than in an ignorance-preserving accommodation of it. Even so, abductions do license inferences, on which subsequent actions might reasonably be taken (albeit defeasibly, both times).

Thin-slicing resembles abduction in certain quite clear ways. Just as the relation between $K(H)$ and T is *presumptive* attainment for the abducer, so, for the thin-slicer, is the inference from his small sample and his generalization *presumptive.* Just as the abducer's inference of $C(H)$ itself only a *plausible inference,* the thin-slicer's instantiation of his generalization is a *default.*

What is less clear is whether it is invariably the case that whenever a thin-slicer slices thinly he has (however tacitly) set himself an abductive target. Certainly we may assume that when faced with a small sample, no hasty generalizer will take the view that this constitutes a knowledge-base that attains the cognitive target (if that's what it is) of *knowing* that the generalization is true. But it is another thing entirely as to whether we might also assume that in reaching his more qualified inference—presumptive generalization, as we might call it,—there is an hypothesis H which, when added to the sample, would indeed attain the generalization unqualifiedly. Of course, there is such an H. It says that the sample is *representative.* But it won't quite do for abduction, since the proposition that the sample of Fs that G is representative just is the generalization in question.

No doubt, these and other questions could be explored to advantage. But, unless we are mistaken, we have already seen enough of the similarity between thin-slicing and abduction to be able to emphasize what is essential to the making of hasty generalizations. They are made *presumptively,* and the instantiations they sanction are *defaults.* This gives us what we want.

Proposition 24 (Confirming our thesis)**.** *Individual thin-slicers characteristically do not take on the standard of inductive strength. Given that an inductive fallacy is one that fails the standard of inductive strength, the* GOE *fallacy of hasty generalization is characteristically not a fallacy committed by beings like us.*

What is more,

Proposition 25 (The virtue of haste)**.** *Given the generally good track record of individual thin-slicers and the considerable economies that thin-slicing achieves, the practice of hasty generalization possesses, for beings like us, the cognitive virtue of producing large stores of default propositions on which to ground, with due regard for the attendant risks, the appropriate actions. In other words, thin-slicing is a natural discouragement of paralyzing indecision.*

As we have remarked, the slenderness of our own sample might well leave the reasonable reader unmoved to accept our ambitious claim for all 18 members of *GOE*. Certainly, we will not be so brazen as to suggest that sample produced by thin-slicing is representative of all of *GOE*. Let us say it again. It is a small sample. Perhaps we would have been better advised to entitle our paper "Thin-slicing as a cognitive virtue". Even so, we do think that some quite general lessons can be drawn from our examinations of this sample. One is that a piece of reasoning is a fallacy only in relation to what the agent has in mind to achieve cognitively. So, at a minimum, before we can rightly accuse an individual agent of committing a *GOE*-fallacy, we must have independent reasons for supposing that his target T carries standards S that his reasoning violates. In light of the forgoing discussion, we take it as given that it is often far from obvious that such Ts and Ss are actually in play in the cognitive lives of beings like us.

We keep saying "beings like us". This is because it matters. Beings like us are individual agents. Individual agents tend to set for themselves moderate targets. Moderate targets are those that can be attained (or as we may now say, presumptively attained) by the deployment of scant-resources i.e., scant in comparison to what NASA and MI5 command. Agents whose resource-draw is greatly larger than ours certainly set themselves tougher targets governed by higher standards. We don't doubt for a minute that when NASA was in process of generalizing about O-ring integrity, it was clearly targeted on scientific certainty and clearly pledged to the standard of inductive strength. In such circumstances, thin-slicing would have been cognitively defective; worse, it would have been an ethical catastrophe. There is also a moral to be drawn from this.

Proposition 26 (Vindication of the tradition)**.** *On the traditional view, a fallacy is an inapparent error. Leaving aside the general point that all errors are inapparent, we see that hasty generalization conforms to the traditional view. For it* is *an error (when committed by NASA) and it* looks not to be *an error, because it is not an error (when committed by beings like us).*

Finally, we should make brief mention of the Principle of Charity.[22] The Principle of Charity bids us not to interpret our interlocutors in ways that convict them of error or irrationality; more carefully, we are not so to interpret them except in default of strong indications to the contrary. The Principle of Charity is itself hardly free from controversy, and we have no wish to rush to judgement on its behalf (see Woods, 2004, Chapter 14). Suffice it to say that if, when done by us, thin-slicing is indeed a fallacy, then beings like us are *massive inductive misfits.* There is, apart from the soundness of the Charity Principle, a further reason to doubt it. Suppose that we were indeed massive inductive misfits. It would hardly matter. For we get things more right than wrong. We survive, we prosper, and occasionally we build great civilizations. What this would tell us, given the present assumption, is that it is *not irrational* to be massive inductive misfits.

Acknowledgments

Our work is supported by the Engineering and Physical Sciences Research Council of the United Kingdom, the Social Sciences and Humanities Research Council of Canada, the Dean of Arts, University of British Columbia, the Dean of Arts and Science, University of Lethbridge, and the Head of Computer Science at King's College. We are deeply grateful for this assistance. We also extend our gratitude to Carol Woods for invaluable technical support. It is a pleasure to record our debt to the many persons with whom conversations and correspondence have helped us materially in forming our views about fallacies and the logic of cognitive systems: Atocha Aliseda, Peter Alward, Peter Bruza, Balasishnan Chandrasekaran, Artur Garcez, Hans Hansen, Jaakko Hintikka, David Hitchcock, Scott Jacobs, Erik Krabbe, Theo Kuipers, Henrike Jansen, Luis Lamb, Dom Lopes, Peter McBurney, Mohan Matthen, Lorenzo Magnani, Sami Paavola, Kent Peacock, Jeanne Peijenburg, Ahti-Veikko Pietarinen, Patrick Rysiew, Matti Sintonen, Patrick Suppes, Stephen Toulmin, Bas van Fraassen, Johan van Benthem, Paul Viminitz, Mark Weinstein, Joseph Wenzel.

References

Alchourron, C. A., Gärdenfors, P. G., and Makinson, D. (1985). On the logic of theory change; partial meet, contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530.

Ambady, N. and Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical atractiveness. *Journal of Personality and Social Psychology*, 64:431–441.

[22]Charity is way of distinguishing between rival analytical hypotheses, which is Quine's word for a translation manual for an alien's linguistic behaviour (Quine, 1970, Chapter 2) and concerning which Davidson holds that "[c]harity is forced upon us; whether we like it or not, if we want to understand others, we must count them right in most matters" (Davidson, 1984, xviii). Scriven writes to the same effect that the "Principle of Charity requires that we make the best, rather than the worst possible interpretation. . . ." (Scriven, 1976, 71).

Barth, E. M. and Krabbe, E. C. W. (1982). *From Axiom to Dialogue*. de Gruyter, Berlin, New York.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT, Cambridge, MA.

Bermúdez, J. L. (2004). *Thinking Without Words*. Oxford University Press, Oxford.

Black, M. (1946). *Critical Thinking*. Prentice-Hall, New York.

Bruza, P. D., Song, D., and McArthur, R. M. (2004). Abduction in semantic space: Towards a logic of discovery. *Logic Journal of the IGPL*, 12(2):97–109.

Bruza, P. D., Cole, R. J., Song, D., and Abdul Bari, K. (2006) Towards operational abduction from a cognitive perspective. *Logic Journal of the IGPL*, 14(2):161–177.

Carlson, G. N. and Pelletier, F. J., editors (1995). *The Generic Book*. Chicago University Press, Chicago, IL.

Carney, J. D. and Scheer, R. K. (1980). *Fundamentals of Logic*. Macmillan, New York, 3rd edition.

Carrere, S. and Gottman, J. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process*, 38(3):293–301.

Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge University Press, Cambridge.

Churchland, P. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. MIT, Cambridge MA.

Churchland, P. (1995). *The Engine of Reason, the Seat of the Soul*. MIT, Cambridge, MA.

Cohen, J. (1982). Are people programmed to commit fallacies: Further thoughts about the interpretation of experimental data and probability judgement. *Journal of Theory and Social Behavior*, 12:251–274.

Copi, I. M. (1986). *Introduction to Logic*. Macmillan, New York, 7th edition.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford University Press, Oxford, New York.

d'Avila Garcez, A. S. and Lamb, L. C. (2004). Reasoning about time and knowledge in neural-symbolic learning systems. In Thrum, S. and Schoelkopk, B., editors, *Proceedings of the NIPS 2003 Conference*, volume 16 of *Advances in Neural Information Processing Systems*, Vancouver, BC, MIT, Cambridge, MA.

d'Avila Garcez, A. S., Broda, K., and Gabbay, D. M. (2002). *Neural-Symbolic Learning Systems: Foundations and Applications*. Springer, Berlin.

Franklin, J. (2001). *The Science of Conjecture: Evidence and Probability Before Pascal*. The Johns Hopkins University Press, Baltimore, MD.

Freeman, J. B. (1991). *Dialectics and the Microstructure of Argument*. Foris, Dordrecht.

Frege, G. (1879). *Begriffschrift, a Formal Language, Modeled upon That of Arithmetic, for Pure Thought*. Harvard University Press, Cambridge, MA.

Gabbay, D. M. and Woods, J. (2001a). More on non-cooperation in dialogue logic. *Logic Journal of the IGPL*, 9:321–339.

Gabbay, D. M. and Woods, J. (2001b). The new logic. *Logic Journal of IGPL*, 9:157–190.

Gabbay, D. M. and Woods, J. (2001c). Non-cooperation in dialogue logic. *Synthese*, 127:161–186.

Gabbay, D. M. and Woods, J. (2003a). *Agenda Relevance: A Study in Formal Pragmatics*, volume 1 of *A Practical Logic of Cognitive Systems*. North Holland, Amsterdam.

Gabbay, D. M. and Woods, J. (2003b). Normative models of rationality: The theoretical disutility of some approaches. *Logic Journal of IGPL*, 11:597–613.

Gabbay, D. M. and Woods, J. (2005). *The Reach of Abduction: Insight and Trial*, volume 2 of A Practical Logic of Cognitive Systems. Elsevier, Amsterdam.

Gabbay, D. M. and Woods, J. (2009). Errors of logic. Forthcoming.

Gabbay, D. M. (1976). *Investigations in Modal and Tense Logics with Applications*. Reidel, Dordrecht, Boston, MA.

Gabbay, D. M., Hodkinson, I., and Reynolds, M. (1994). *Temporal Logic: Mathematical Foundation and Computational Aspects*, volume 1. Oxford University Press, Oxford.

Gabbay, D. M., Rodrigues, O., and Woods, J. (2002). Belief contraction, antiformulae and resourse-overdraft: Part I: Deletion in resources bounded logics. *The Logic Journal of the IGPL*, 10:601–652.

Gabbay, D. M., Pigozzi, G., and Woods, J. (2004a). Controlled revision: A preliminary account. *The Logic Journal of the IGPL*, 13:5–27.

Gabbay, D. M., Rodrigues, O., and Woods, J. (2004b). Belief contraction, antiformulae and resourse-overdraft: Part II: Deletion in resources unbounded logics. In Rahman, S., Symons, J., Gabbay, D. M., and van Bendegem, J. P., editors, *Logic Epistemology and the Unity of Science*, pages 291–326. Kluwer, Dordrecht, Boston, MA.

Gigerenzer, G. and Selten, R. (2001). Rethinking rationality. In *Bounded Rationality: The Adaptive Toolbox*, pages 1–12. MIT, Cambridge, MA.

Gochet, P. (2002). The dynamic turn in twentieth century logic. *Synthese*, 130:175–184.

Gochet, P. and Gribomont, P. (2005). Epistemic logic. In Gabbay, D. M. and Woods, J., editors, *Handbook of the History of Logic*, vol 7, pages 99–195. Elsevier, Oxford.

Grootendorst, R. (1987). Some fallacies about fallacies. In van Eemeren, F. H., Grootendorst, R., Blair, J. A., and Willard, C. A., editors, *Argumentation Across the Lines of Discipline*, pages 331–342. Foris, Dordrecht, Providence.

Guarini, M. (2001). A defence of connectionism against the "SYNTACTIC" argument. *Synthese*, 128:287–317.

Hamblin, C. L. (1970). *Fallacies*. Methuen, London.

Hansen, H. V. and Pinto, R. C., editors (1995). *Fallacies: Classical and Contemporary Readings*. Pennsylvania State University Press, University Park, PA.

Harel, D. (1979). *First-Order Dynamic Logic*. Springer, Berlin.

Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT, Cambridge, MA.

Hilpinen, R., editor (1981). *Deontic Logic: Introductory and Systematic Readings*. Reidel, Dordrecht.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca, NY.

Hintikka, J. (1981). *Modern Logic—A Survey*. Reidel, Boston, MA.

Hintikka, J. (1987). The fallacy of fallacies. *Argumentation*, 1:211–238.

Hintikka, J. (1997). What was Aristotle doing in his early logic, anyway? A reply to Woods and Hansen. *Synthese*, 113:241–249.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Amsterdam.

Horgan, T. and Tienson, J. (1999a). Authors' replies. *Acta Analytica*, 22:275–287.

Horgan, T. and Tienson, J. (1999b). Short précis of connectionism and the philosophy of psychology. *Acta Analytica*, 22:9–21.

Irvine, A. D. (1989). Epistemic logicism and Russell's regressive method. *Philosophical Studies*, 55:303–327.

Johnson, O. (1967). Begging the question. *Dialogue*, 6:135–160.

Johnson, R. H. (1996). *The Rise of Informal Logic*. Vale, Newport News, VA.

Johnson, R. H. (2000). *Manifest Rationality: A Pragmatic Theory of Argument*. Lawrence Erlbaum Associates, London.

Kahneman, D. and Tversky, A. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Kowalski, R. A. (1979). *Logic for Problem Solving*. Elsevier, New York.

Kripke, S. A. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 83–94.

Lenzen, W. (1978). Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:1–219.

MacKenzie, J. (1990). Four dialogue systems. *Studia Logica*, XLIX:567–583.

Magnani, L. (2001). *Abduction, Reason and Science: Processes of Discovery and Explanation*. Kluwer, Plenum, New York.

Massey, G. J. (1981). The fallacy behind fallacies. *Midwest Studies in Philosophy*, 6:489–500.

Matthen, M. (2002). Human rationality and the unique origin constraint. In Ariew, A., Cummins, R., and Perlman, M., editors, *Functions: New Readings in the Philosophy of Psychology and Biology*, pages 341–372. Oxford University Press, Oxford.

McCarthy, J. (1980). Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39.

Moore, R. (1985). Semantical considerations on non-monotonic logics. *Artificial Intelligence*, 25:75–94.

Peirce, C. S. (1992). *Reasoning and the Logic of Things: The Cambridge Conference Lectures of 1898*. Ketner, K. L., editor, introduction by Kenneth Laine Ketner and Hilary Putnam. Harvard University Press, Cambridge, MA.

Pereira, L. M. (2002). Philosophical incidence of logic programming. In Gabbay, D. M., Johnson, R. H., Ohlbach, H. J., and Woods, J., editors, *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, volume 1, pages 421–444. North Holland, Amsterdam.

Powers, L. (1995). Equivocation. In Hansen, H. V. and Pinto, R., editors, *Fallacies: Classical and Contemporary Readings*, pages 287–301. Pennsylvania State University Press, University Park, PA.

Prior, A. N. (1967). *Past Present and Future*. Oxford University Press, Oxford.

Quine, W. V. O. (1970). *Philosophy of Logic*. Prentice Hall, Englewood Cliffs, NJ.

Raiffa, H. (1968). *Decision Analysis*. Addison Wesley, Reading, MA.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 12:81–132.

Rensink, R. (2000). Visual sensing without seeing. *Psychological Science*, 15:27–32.

Sandewall, E. (1972). *An Approach to the Frame Problem and Its Implementation*. Edinburgh University Press, Edinburgh.

Schipper, E. W. and Schuh, E. (1959). *A First Course in Modern Logic*. Henry Holt, New York.

Schlechta, K. (2004). *Coherent Systems*. Elsevier, Amsterdam.

Scriven, M. (1976). *Reasoning*. McGraw-Hill, New York.

Sorensen, R. A. (1988). *Blindspots*. Clarendon, Oxford.

Tarski, A. (1956). The concept of truth in formalized languages. In *Logic Semantics, Metamathematics*, pages 152–278. Translated by J. H. Woodger. Clarendon, Oxford.

Toulmin, S. (1953). *The Philosophy of Science: An Introduction*. The Hutchinson University Library, London.

van Benthem, J. (1983). *The Logic of Time*. Reidel, Dordrecht.

van Benthem, J. (1996). *Exploring Logical Dynamics*. CSLI, Stanford.

van Fraassen, B. C. (2005). The day of the dolphins: Puzzling over epistemic partnership. In Peacock, K. A. and Irvine, A. D., editors, *Mistakes of Reason: Essays in Honour of John Woods*, pages 111–133. University of Toronto, Toronto.

von Wright, G. H. (1951). *An Essay in Modal Logic*. North Holland, Amsterdam.

Walton, D. (1995). *A Pragmatic Theory of Fallacy*. University of Alabama Press, Tuscaloosa, AL.

Walton, D. and Krabbe, E. C. W. (1995). *Commitment in Dialogue*. SUNY, Albany, NY.

Wheeler, M. (2001). Two threats to representation. *Synthese*, 129:211–231.

Williamson, J. (2002). Probability logic. In Gabbay, D. M., Johnson, R., Ohlbach, H. J., and Woods, J., editors, *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, pages 397–424. North Holland, Amsterdam.

Woods, J. (2003). *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*. Cambridge University Press, Cambridge.

Woods, J. (2004). *The Death of Argument: Fallacies in Agent-Based Reasoning*. Kluwer, Dordrecht, Boston.

Woods, J. and Hansen, H. V. (1997). Hintikka on Aristotle's fallacies. *Synthese*, 113:217–239.

Woods, J. and Walton, D. (1989). *Fallacies: Selected Papers 1972–1982*. Foris–de Gruyter, Berlin, New York.

Woods, J., Irvine, A., and Walton, D. (2004). *Argument: Critical Thinking, Logic and the Fallacies*. Prentice Hall, Toronto, 2nd edition.

Part II

GAME-THEORETIC SEMANTICS

Chapter 5

A STRATEGIC PERSPECTIVE ON IF GAMES

Merlijn Sevenster
Philips Research, Eindhoven
merlijn.sevenster@philips.com

Abstract Hintikka and Sandu's Independence-friendly logic (Hintikka, 1996; Hintikka and Sandu, 1997) has traditionally been associated with extensive games of imperfect information. In this paper we set up a strategic framework for the evaluation of IF logic à la Hintikka and Sandu. We show that the traditional semantic interpretation of IF logic can be characterized in terms of Nash equilibria. We note that moving to the strategic framework we get rid of IF semantic games that violate the principle of perfect recall. We explore the strategic framework by replacing the notion of Nash equilibrium by other solution concepts, that are inspired by weakly dominant strategies and iterated removal thereof, charting the expressive power of IF logic under the resulting semantics.

5.1 Introduction

Game theory has proven to be a tool capable of covering the essentials of established subjects in research areas such as logic, mathematics, linguistics and computer science. Game-theoretic concepts have also been proposed in cases where traditional machinery broke down. In this paper we will study the game theory that functions as a verificational framework for *independence-friendly* (IF) first-order logic, which is a generalization of standard first-order logic (FOL).

As a semantics used for evaluating FOL, Tarski semantics is well-known and widely agreed upon. Yet this semantics cannot be used to evaluate Hintikka and Sandu's IF first-order logic, see Cameron and Hodges (2001). IF logic abstracts away from the Fregean assumption that syntactical scope and semantical dependence of quantifier-variable pairs coincides. That is, in an IF logical formula, if '$\exists x$' is in the syntactical scope of '$\forall y$', the variable x can be made semantically independent of y by means of the slash operator. To evaluate IF logical formulae, Hintikka and Sandu (in Hintikka, 1996; Hintikka and

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

Sandu, 1997) introduce the notion of a *semantic evaluation game*. The independence of two variables expressible in IF logic is typically reflected by the corresponding semantic evaluation game being of *imperfect information*. This is in contrast to the evaluation games related to first-order formulae, they are of perfect information. Truth of an FOL or IF formula is defined in terms of its semantic evaluation game. This semantics was coined *game-theoretic semantics* (GTS) by Hintikka.

It has been noted in the literature (van Benthem, 2000, Dechesne, 2005) that some IF evaluation games violate the game-theoretic principle of *perfect recall*. In game theory, games without perfect recall have not been studied extensively, one of the reasons being that it is hard to understand what real-life situations they capture—put loosely, they are not '*playable*'. Thereby also the playability of IF games is called into question.

In this paper, we set up a strategic game-theoretic framework in which IF games can be defined. We will see that truth of IF under GTS can be characterized in terms of Nash equilibria in the strategic framework. We observe that the playability issues, concerning perfect recall, evaporate in the strategic framework, yet we get so-called *coordination problems* in return.

We explore the strategic framework by replacing the notion of Nash equilibrium by other solution concepts. That is, we also define truth for IF logic in terms of weakly dominant strategies and iterated removal thereof. Naturally, changing semantics affects the truth conditions of IF formulae, a phenomenon we study in terms of the expressive power of IF logic w.r.t. the resulting semantics.

Section 5.2 recalls the basics of IF logic and GTS. In Section 5.3, we define the strategic framework and establish the connection between GTS and truth in terms of Nash equilibria. Sections 5.4 and 5.5 explore the notions of truth that result after replacing the Nash equilibrium solution concept by different ones, that are inspired by the game-theoretic notions of weak dominance and iterated removal of strategies in strategic games.

The formal results are mostly given without proof. We hope to make an extended version of this paper, containing full proofs, available soon.

5.2 IF logic and game-theoretic semantics

The program of *quantifier independence*, as founded by Henkin (1959) and later Hintikka (1996), is concerned with abstracting away from the Fregean assumption that the syntactical scope and binding of quantifiers in first-order logic coincide. The syntax of *independence-friendly first-order logic* as proposed by Hintikka (1996) extends FOL, in the sense that, for example, if

$$\forall x_1 \exists x_2 \ldots \forall x_{n-1} \exists x_n \, R(x_1, \ldots, x_n)$$

is a FOL sentence containing the n-ary predicate R, then

$$(\forall x_1/X_1)(\exists x_2/X_2)\ldots(\forall x_{n-1}/X_{n-1})(\exists x_n/X_n)\, R(x_1,\ldots,x_n) \tag{1}$$

is an IF sentence, provided that $X_i \subseteq \{x_1,\ldots,x_{i-1}\}$. The variable x_i is intuitively meant to be *independent* of the variables in X_i, although it appears under their syntactical scope.

Definition 1. *In this paper* FOL *denotes the smallest set of first-order* sentences*, that are in prenex normal form and in which every variable is quantified exactly once. We will assume them being of the form*

$$Q_1 x_1 \ldots Q_n x_n\, R(x_1,\ldots,x_n), \tag{2}$$

where $Q_i \in \{\exists,\forall\}$*. If no confusion arises we will abbreviate any string of variables* $x_1, x_2, \ldots$ *using* $\bar{x}$*.*

The reader has noted that the language we call FOL is really a simple version of first-order logic. This simplification streamlines notation considerably when we define the IF language, without affecting the contention of this paper. Analogously to Hintikka (1996) we define the syntax of IF logic in terms of FOL, as follows.

Definition 2. *The language* IF *is obtained from* FOL *by repeating the following procedure a finite number of times: if* $\phi \in$ FOL*, then*

> *If '*$Q_i x_i\ \psi$*' occurs in* ϕ*, then it may be replaced by '*$(Q_i x_i/X_i)\ \psi$*', where* $Q_i \in \{\exists,\forall\}$ *and* $X_i \subseteq \{x_1,\ldots,x_{i-1}\}$*.*

Since sentences in FOL *are assumed to be as in* (2)*, sentences of* IF *will be of the form*

$$(Q_1 x_1/X_1)\ldots(Q_n x_n/X_n)\, R(x_1,\ldots,x_n), \tag{3}$$

writing '$Q_i x_i$*' instead of '*$(Q_i x_i/\emptyset)$*'.*

In $\phi \in$ FOL containing the strings '$Q_i x_i$' and '$Q_j x_j$', variable x_j depends on x_i iff $i < j$. In IF this linear ordering of dependency is given up—the quantifiers of IF sentences are *partially ordered*. The first partially ordered quantifier, also known as *Henkin quantifier*, appeared in Henkin (1959). For later usage, we formalize variable dependence by means of a binary relation. To this end let the set $Var(\phi) = \{x_1,\ldots,x_n\}$ denote the variables for the IF formula ϕ as in (3). Then, $B_\phi \subseteq Var(\phi) \times Var(\phi)$ is ϕ's *dependency relation*, such that for every $x_i, x_j \in Var(\phi)$

$$(x_i, x_j) \in B_\phi \quad \text{if} \quad i < j \ \text{ and } \ x_i \notin X_j.$$

Truth of an IF sentence is evaluated relative to a suitable *model* $M = (D, I, p)$ in which we distinguish a *domain* D of objects; an *interpretation function* I, that

determines the extension of relation symbols; and an *assignment p* that assigns an object from the domain D to each variable. Hintikka and Sandu (1997) associate with every $\phi \in$ IF and suitable model M a *semantic evaluation game* $g(\phi, M)$. The game is played by two players, called E and A, that control the existential and universal quantifiers in ϕ. In $g(\phi, M)$ the players and quantifiers are associated through the *player function* P, that is, the function such that $P(\exists) = E$ and $P(\forall) = A$. Intuitively, $g(\phi, M)$ proceeds as follows:

> $g((Q_i x_i/X_i)\ \psi, M)$ triggers player $P(Q_i)$ choosing an object $d_i \in D$; the game continues as $g(\psi, M)$.
>
> $g(R(\bar{x}), M)$ has no moves; E receives payoff 1 if $\bar{d} \in I(R)$, and -1 otherwise. A gets E's payoff times -1.

The above rules regulate the behavior of the game $g(\phi, M)$. Hintikka and Sandu (1997) do not provide a rigorous game-theoretic model for these games. However, the formal treatments provided in the literature all take an *extensive* stance towards these games, viz. van Benthem (2004); Pietarinen and Tulenheimo (2004); Dechesne (2005) and Sandu and Pietarinen (2003) for a propositional variant. In this paper the game $g(\phi, M)$—with a lower-case 'g'—denotes a Hintikka-Sandu style, extensive semantic game. In these games independence is modeled by means of *information sets* imposed on the *histories* of the game tree. We omit rigorous definitions, but illustrate the idea by means of the game tree of an IF sentence θ that reappears in our discussion below,

$$\theta = \exists x_1(\exists x_2/\{x_1\})\ [x_1 = x_2], \tag{4}$$

evaluated on the model $(\{a, b\}, =)$, depicted in Figure 5.1. From a game-theoretic perspective, every node in a game tree corresponds to a history, and every leaf to a complete history. On every complete history the players' utility functions are defined.

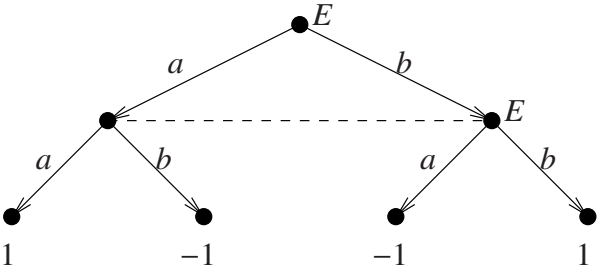


Figure 5.1: The game tree of $g(\theta, (\{a, b\}, =))$, containing seven histories. The top node corresponds to the empty history; the histories on the intermediate layer are denoted by $\langle a\rangle$, $\langle b\rangle$; and $\langle a, a\rangle, \langle a, b\rangle, \langle b, a\rangle, \langle b, b\rangle$ are the terminal nodes. The fact that $\langle a\rangle$, $\langle b\rangle$ sit in the same information set is reflected by the dashed line. The values 1 and -1 are payoffs for E

To say that two histories are in the same information set means that the player owning the set at hand cannot distinguish between the two histories while at it. As a consequence any *pure strategy* for this player prescribes only *one* action for all the histories in the information set.

We say that E has a *winning* strategy in $g(\phi, M)$ if there exists a strategy that guarantees an outcome of 1, against every strategy played by A; and a strategy is *uniform* with respect to the game's information sets, if it assigns to every information set in which E is to move exactly one object from the domain. Note that here and henceforth we consequently mean 'pure strategy' when speaking of 'strategy'. Truth under GTS is defined in terms of winning strategies.

Definition 3. *Let* $\phi \in$ IF *and let M be a suitable model. Then define truth under GTS as follows:*

> *ϕ is true under GTS on M, denoted by* $M \models_{\text{GTS}} \phi$*, if E has a winning strategy in* $g(\phi, M)$*.*
>
> *ϕ is false under GTS on M, if A has a winning strategy in* $g(\phi, M)$*.*
>
> *ϕ is undetermined under GTS on M, if neither E nor A has a winning strategy in* $g(\phi, M)$*.*

In the realm of IF semantic evaluation games, information sets only partition histories of equal length (cf. Sandu and Pietarinen, 2003). Pure strategies in IF semantic games therefore coincide with tuples of *Skolem functions*, as we know them from logic. We introduce Skolem functions by illustrative means. Let ϕ be as in (1), then its Skolemization looks like

$$\exists f_2 \ldots \exists f_n \forall x_1 \ldots \forall x_{n-1}\, R(\bar{x}, \bar{f}),$$

where f_i is a Skolem function, being a function of type $D^{\{x_1,\ldots,x_{i-1}\}\setminus X_i} \to D$.

Walkoe (1970) showed that the truth condition of every formula with partially ordered quantifiers can be expressed in the Σ^1_1 fragment of second-order logic. Later, the result, applied to IF, reappears in Sandu's and Hintikka's work (for references see Hintikka and Sandu, 1997) hinging on the fact that for ϕ as in (1)

$$M \models_{\text{GTS}} \phi \qquad \text{iff} \qquad M \models_{\text{Tarski}} \exists f_2 \ldots \exists f_n \forall x_1 \ldots \forall x_{n-1}\, R(\bar{x}, \bar{f}),$$

since any tuple $f_2, \ldots, f_n$ witnessing the truth of ϕ's Skolemization is a winning strategy for E in $g(\phi, M)$ and the other way around, assuming the Axiom of Choice. For Hintikka and Sandu (1997) it is the strategies that form the heart of the game-theoretic apparatus involved.

> What is essential [about game-theoretic conceptualizations] is not the idea of competition, winning and losing. ... What is essential is the notion of strategy. Game theory was born the moment John von Neumann formulated explicitly this notion.

Having read this, the thought occurs that defining IF evaluation games in a *strategic* way may be more in line with Sandu's and Hintikka's thinking. In this paper we will set up such a strategic framework; discuss the 'playability' of IF games in this context; and start exploring the framework.

The issue of playability of IF games, mentioned above, arises when we actually want to play games for IF sentences ϕ. In a game for ϕ, the turn-taking is governed by ϕ's quantifier prefix and the epistemic qualities of the agents by ϕ's slash sets. However, defining the IF language, we took no special care that our formulas would give rise to playable games. In fact, it has been observed that certain IF sentences yield games that require agents with odd epistemic features. That is, games that violate the game-theoretic principles of *perfect memory* and *action memory*. Roughly speaking a game of imperfect information has perfect memory if a player learning something (in our context: a previous move), implies it knowing this piece of information for the rest of the game; and, a game has action memory if every player recalls at least it's own moves. We refer the reader to Sevenster (2006) for an elaborate treatment of perfect recall and IF games.

For the sake of illustration, consider the extensive game $g(\theta, (\{a, b\}, =))$, with θ as in (4).[1] It is the case that $(\{a, b\}, =) \models_{\mathrm{GTS}} \theta$, since the tuple (play a, play a) is a winning strategy. But also we have it that the histories $\langle a \rangle$ and $\langle b \rangle$ are in E's information set indicating that these histories are *indistinguishable* for E. Thus, $g(\theta, (\{a, b\}, =))$ lacks both perfect memory and action memory.

The issue of the playability of $g(\theta, (\{a, b\}, =))$ evolves around the question *how E can understand that* (*play a, play a*) *is a winning strategy for E, despite the fact that she is uninformed at the intermediate stage.* That is, E seems to forget her own move right after playing it!

One explanation may be that E is allowed to decide beforehand on a strategy and consult it while playing the game, even if she is unsure about her own moves at the intermediary stage. (This explanation appears in van Benthem (2000).) In particular, that (play a, play a) is a winning strategy can then be understood as follows: First E picks a, thereafter she is uncertain about what history she is actually in: $\langle a \rangle$ or $\langle b \rangle$. By consulting here winning strategy, however, she derives that she actually is in $\langle a \rangle$ and not in $\langle b \rangle$. The imperfect information evaporates!

[1] The formula θ also appears in Janssen (2002), as an argument against Hintikka's claim of IF logic modeling quantifier independence. Janssen argues that, since θ holds on the domain, it must be the case that x_2 depends on x_1. However, in θ the choice for x_2 is independent of x_1, since $X_2 = \{x_1\}$. For more on IF logic and intuitions on independence, consult Janssen (2002).

This explanation requires more game-theoretic structure—i.e., consulting of one's strategy—than present in its description and would imply a non-game-theoretic understanding of having imperfect information *during* the game.

Another explanation may be that E is an *existential team*, hence associating with every existential variable a member of the team. This would make $g(\theta, (\{a, b\}, =))$ a two-player cooperative game. But then the very fact that θ holds on the model at stake suggests to be interpreted in such a way that the x_1-player and the x_2-player are allowed to settle on their strategies *before* the game. Again, no such event can be found in the definition of $g(\theta, (\{a, b\}, =))$ and it seems such an event would violate the game-theoretic understanding of information sets. Because, for instance in $g(\theta, (\{a, b\}, =))$ the second player in the E-team would really know the move of the first player.

Below we shall reduce the puzzle that arises with θ to the question how Nash equilibria are supposed to arise in strategic games. First we set up a strategic framework, in which the notion of Nash equilibrium and other solution concepts can be meaningfully employed.

5.3 Strategic framework for IF games

In this section we define IF games as strategic games. We characterize truth of IF under GTS in terms of Nash equilibria.

Definition 4. *Let* $\phi \in$ IF *and let* M *be a suitable model. Then, define the* strategic evaluation game of ϕ and M *as*

$$G(\phi, M) = \left(N_\phi, (S_{i,\phi})_{i \in N_\phi}, (u_{i,\phi,M})_{i \in N_\phi}\right).$$

N_ϕ *denotes the set of players,* $S_{i,\phi}$ *the set of strategies for player* i*, and* $u_{i,\phi,M}$ *is player* i*'s utility function. We also call* $G(\phi, M)$ *an* IF game.

Below we briefly introduce these ingredients componentwise and introduce some notation involved. Note that strings in IF are assumed to be as in (3). All definitions below are restricted to this assumption, but can be generalized without much ado.

Players. The set $N_\phi = \{i \mid x_i \in Var(\phi)\}$ contains the players. The set N_ϕ conveys the strong connection between variables in ϕ and players in $G(\phi, M)$. In fact, if $V \subseteq Var(\phi)$, then we will use $N(V) = \{i \mid x_i \in V\}$ to denote the set of players associated with the variables in V. Let E_ϕ (A_ϕ) be the set of existentially (universally) quantified variables in ϕ. We have adopted the *multi-player* view on IF games here, mainly because it is the framework that is most open to generalizations with respect to, for instance, the utility functions. Moreover, it allows for smoother terminology.

Strategies. For $x_i \in Var(\phi)$, define $U_{i,\phi} \subseteq Var(\phi)$ to be the set of variables on which x_i depends in ϕ. That is, $U_{i,\phi} = \{x_j \mid (x_j, x_i) \in B_\phi\}$. In the context of

the game and player i having control over x_i, we often say that i *sees* $U_{i,\phi}$. $S_{i,\phi}$ denotes the set of all player i's strategies in $G(\phi, M)$, being (Skolem) functions of type $s_i : D^{U_{i,\phi}} \to D$. If $U_{i,\phi}$ is empty, $S_{i,\phi}$ only contains *atomic strategy*.

Manipulating strategies. Define a *profile* s in $G(\phi, M)$ as an object in

$$\bigtimes_{i \in N'} S_{i,\phi},$$

for some $N' \subseteq N_\phi$. We call s *existential* (*universal*), if $N' \subseteq E_\phi$ (A_ϕ); otherwise we call it *mixed*. We call s *complete*, if $N' = N_\phi$; otherwise we call it *partial*. If $N' = N(E_\phi)$ ($N(A_\phi)$), we call the profile *completely existential (universal)*. If no confusion arises we will drop as many of the terms as possible.

If $s \in \bigtimes_{i \in N'} S_{i,\phi}$ for some $N' \subseteq N_\phi$ and $\{1, \ldots, j\} \in N'$, then $s_{1,\ldots,j}$ denotes the strategy profile s containing only player 1 to j's strategies. We will often discuss player j changing strategies with respect to a strategy profile s. We write (s_{-j}, t_j) to denote the profile that is the result of replacing s_j by t_j. If $s \in \bigtimes_{i \in N'} S_{i,\phi}$ and $s' \in \bigtimes_{i \in N''} S_{i,\phi}$ for disjoint $N', N'' \subseteq N_\phi$, then ss' is the result of concatenating s and s'. If s_i is a strategy of type $D^{\{y_1,\ldots,y_k\}} \to D$ and assignment p is defined over $\{y_1, \ldots, y_k\}$, then we will write $s_i(p)$ instead of $s_i(p(y_1), \ldots, p(y_k))$.

Finally, every profile $s \in \bigtimes_{i \in \{1,\ldots,j\}} S_{i,\phi}$ in $G(\phi, M)$ gives rise to an assignment $[s]$ that is defined over $\{x_1, \ldots, x_j\}$ as below. Note that s_1 is an atomic strategy.

$$\begin{aligned} [\mathrm{s}](x_1) &= s_1 \\ [\mathrm{s}](x_i) &= s_i([s_{1,\ldots,i-1}]). \end{aligned}$$

Utility functions. Let $i \in N_\phi$. Then, i's *utility function* in $G(\phi, (D, I, p))$ is defined over complete profiles s as follows:

$$u_{i,\phi,(D,I,p)}(s) = \begin{cases} c_i & \text{if } [s] \in I(R) \\ -c_i & \text{if } [s] \notin I(R), \end{cases}$$

where $c_i = 1$ if $i \in N(E_\phi)$, and $c_i = -1$ if $i \in N(A_\phi)$. As all utility functions of the players in $N(E_\phi)$ and $N(A_\phi)$, respectively, are equivalent and the models under consideration can be made up from the context we will simply denote them by u_E and u_A.

Now that we switched from extensive to strategic semantic games, observe that the notion of winning strategy in extensive games has a respectable strategic counterpart: *Nash equilibrium*. We say that the strategy profile $\hat{s}$ is a Nash equilibrium in the strategic game G, if none of the players i gains from unilateral deviation (see also Osborne and Rubinstein, 1994):

$$u_i((\hat{s}_{-i}, s_i)) \leq u_i(\hat{s}),$$

	Play a	Play b
Play a	1	-1
Play b	-1	1

Table 5.1: Every cell in the matrix corresponds to an assignment $[s]$ over $Var(\theta)$. We filled in the value $u_E([s])$ reflecting payoff for the members of the existential team

where s_i is any other strategy for player i and u_i is player i's utility function in G. The following lemma can also be understood as a proof of effective equivalence between $g(\phi, M)$ and $G(\phi, M)$.

Lemma 5. *Let $\phi \in$ IF and let M be a suitable model. Then, the following are equivalent:*

- $M \models_{\mathrm{GTS}} \phi$.
- *There exists a Nash equilibrium s in $G(\phi, M)$, such that $u_E(s) = 1$.*

Technically this lemma is not deep. Yet it shows us that strategic games can account for truth of IF logic. In the strategic framework the playability issues concerning perfect recall, encountered in extensive IF games, evaporate simply because the strategic games ignore the inner structure of games defined by consecutive moves by the agents. By ignoring the inner structure of the game, also the epistemic states of the agents—i.e., their information sets—are ignored.

But the issue of playability pops up in the strategic framework under a different guise. Revisit the game $G(\theta, (\{a, b\}, =))$. As is common usage in strategic games, we draw its payoff matrix, see Table 5.1. The puzzle induced by the truth of θ on $(\{a, b\}, =)$ in extensive contexts appears in the strategic context as a coordination problem. There are two equally profitable Nash equilibria, but which one to choose, without possibility to coordinate? How to understand Nash equilibria is a problem central in game theory, see Osborne and Rubinstein (1994).

In the upcoming two sections we explore semantic interpretations for IF logic that are motivated by solution concepts that are not subject to coordination problems.

5.4 Weak dominance semantics

In this section, we define a semantics based on the existence of *weakly dominant* strategies. Intuitively, a strategy is weakly dominant for a player if it outperforms any other strategy independently of the other players' strategic behavior.

Definition 6. *Fix some IF game $G(\phi, M)$. Then, $\hat{s}_i$ is a* weakly dominant *strategy in $G(\phi, M)$ for player i, if $\hat{s}_i \in S_{i,\phi}$ and for every complete mixed profile s it is the case that*

$$u_E((s_{-i}, \hat{s}_i)) \geq u_E(s).$$

We call $\hat{s}_i$ weakly dominant*, because possibly it is exactly as good as player i's original strategy in s. Dually, we define strategy $t_i \in S_{i,\phi}$ to be* strictly dominated *by $\hat{s}_i$ in $G(\phi, M)$, if for every complete mixed profile s it is the case that*

$$u_E((s_{-i}, \hat{s}_i)) \geq u_E((s_{-i}, t_i)) \qquad \textit{and} \qquad u_E((r_{-i}, \hat{s}_i)) > u_E((r_{-i}, t_i))$$

for at least one complete mixed profile r.

The notion of weak dominance we employ is weaker than the one usually adopted in game theory. For comparison we refer to Osborne and Rubinstein (1994). We now come to our definition of truth in terms of weak dominance.

Definition 7. *Let $\phi \in$* IF *and let M be a suitable model. Then we define truth of ϕ on M under* weak dominance semantics *(WDS) as follows*

> *$M \models_{\text{WDS}} \phi$ iff in $G(\phi, M)$ there exists a complete existential profile $\hat{s}$ such that $\hat{s}_i$ is a weakly dominant strategy for every $i \in N(E_\phi)$, and $u_E(\hat{s}t) = 1$, for any complete universal profile t.*

Falsity and undeterminedness under WDS are defined similarly.

The question remains, of course, what remains of IF logic evaluated under WDS. It becomes clear that GTS is less restrictive a semantics for IF logic than WDS, after reformulating truth under GTS in multi-player terms, since we may simply omit $\hat{s}_i$'s constraint of being weakly dominant:

> $M \models_{\text{GTS}} \phi$ iff in $G(\phi, M)$ there exists a complete existential profile s such that $u_E(st) = 1$, for any complete universal profile t.

Formally, our claim boils down to the claim that

$$M \models_{\text{WDS}} \phi \quad \text{implies} \quad M \models_{\text{GTS}} \phi, \tag{5}$$

but not the other way around. Note that $(\{a, b\}, =) \models_{\text{GTS}} \theta$, but θ does not hold on this domain under WDS, see Table 5.1. As an example of WDS, observe that, surprisingly, for any model M with more than one object in its domain it is the case that for $\tau = \exists x_1 \exists x_2\, [x_1 = x_2]$:

$$M \not\models_{\text{WDS}} \tau \quad \text{whereas} \quad M \models_{\text{Tarski}} \tau.$$

That τ is true under Tarski semantics is obvious. From Table 5.2 it becomes clear that τ is not true under WDS on the model with two objects $\{a, b\}$. Although player 2 has a weakly dominant strategy, player 1 has none.

	s_2^a	s_2^b	s_2^{copy}	s_2^{invert}
s_1^a	1	−1	1	−1
s_1^b	−1	1	1	−1

Table 5.2: s_i^d is the atomic strategy for player $i \in \{1,2\}$ assigning object $d \in \{a,b\}$. s_2^{copy} is player 2's strategy such that $s_2^{\text{copy}}(d) = d$, whereas s_2^{invert} switches the object chosen by player 1

In the remainder of this section we will characterize the truth-conditions of IF under WDS and see that very little is left the Σ_1^1-expressiveness *IF* enjoyed under GTS. We show in Theorem 10 that truth under WDS can be expressed in a fragment of FOL (evaluated under Tarski semantics). Before we come to a rigorous formulation, let us classify an IF sentence ϕ's variables and characterize one of the resulting classes.

Recall that we defined the dependency relation of ϕ's variables as a binary relation B_ϕ. The result of taking the transitive closure of B_ϕ we denote B_ϕ^*. That is, $(x_i, x_j) \in B_\phi^*$ iff there exists a chain $z_0, \ldots, z_m$ of variables in $Var(\phi)$ such that $z_0 = x_i$, $z_m = x_j$, and for every $t \in \{0, \ldots, m-1\}$ it is the case that $(z_t, z_{t+1}) \in B_\phi$. Such a chain of variables $z_0, \ldots, z_m$ we will call a B_ϕ*-chain*. Note that B_ϕ^* is irreflexive.

For every variable $x_i \in Var(\phi)$, partition $Var(\phi)\backslash\{x_i\}$ as follows:

$$U_{i,\phi} = \{x_j \mid (x_j, x_i) \in B_\phi\} \quad (6)$$

$$W_{i,\phi} = \{x_j \mid (x_i, x_j) \in B_\phi^*\} \quad (7)$$

$$V_{i,\phi} = Var(\phi)\backslash(U_{i,\phi} \cup \{x_i\} \cup W_{i,\phi}). \quad (8)$$

We encountered $U_{i,\phi}$ before, as it contains all variables seen by player i. $W_{i,\phi}$ contains the variables that can *(in)directly* see x_i. $V_{i,\phi}$ is the set of all other variables in ϕ not containing x_i. What is meant by 'seeing (in)directly' is pinpointed by the following lemma, that characterizes the variables in $W_{i,\phi}$.

Lemma 8. *Let $\phi \in$ IF be as in* (3) *and let M be a suitable model. Let $W_{i,\phi}$ be defined as in* (7) *for some sentence ϕ and $i \in N_\phi$. Then, $x_j \in W_{i,\phi}$ iff $i \neq j$ and in $G(\phi, M)$ there exists a complete strategy profile s and a strategy $t_i \in S_i$ such that $[s](x_j) \neq [(s_{-i}, t_i)](x_j)$.*

Intuitively, $W_{i,\phi}$ is the subset of $Var(\phi)$ consisting of variables that are sensitive to x_i changing assignments. The lemma, interpreted the other way around, teaches that, if x_j is not in $W_{i,\phi}$, for every strategy profile, player i changing strategies does affect the object assigned to x_j.

Theorem 9. *Let $\phi \in$ IF as in* (3) *and let M be a suitable model. The sets $U_{i,\phi}$, $W_{i,\phi}$, $V_{i,\phi}$ are defined as in* (6), (7) *and* (8)*, respectively. We also consider the*

set $W'_{i,\phi} = \{x' \mid x \in W_{i,\phi}\}$. *The strings of variables in these respective sets will be referred to by means of* $\bar{u}, \bar{v}, \bar{w}$, *and* $\bar{w}'$. *Then, in* $G(\phi, M)$ *player* $i \in N_\phi$ *has a weakly dominant strategy iff*

$$M \models_{\text{Tarski}} \forall\bar{u}\exists x'_i \forall x_i \forall\bar{v}\forall\bar{w}\forall\bar{w}' \quad [(i) \wedge (ii) \wedge (iii) \rightarrow (iv)], \tag{9}$$

where

$$\begin{aligned}
(i) &= R(\bar{u}, x_i, \bar{v}, \bar{w}) \\
(ii) &= x_i \neq x'_i \\
(iii) &= \bigwedge_{j \neq i \in N} \left(\left(\bigwedge_{x_k \in U_{j,\phi}} x_k = x^*_k \right) \rightarrow x_j = x^*_j \right) \\
(iv) &= R(\bar{u}, x'_i, \bar{v}, \bar{w}').
\end{aligned}$$

If $U_{j,\phi}$ *is empty, interpret* $(\bigwedge_{x_k \in U_{j,\phi}} x_k = x^*_k)$ *as* $\top$. *Note that* $Var((i)) = Var(\phi)$ *and that* $Var((iv)) = U_{i,\phi} \cup \{x'_i\} \cup V_{i,\phi} \cup W'_{i,\phi}$. *Furthermore,* $\cdot^*$ *is a mapping from* $Var((i))$ *to* $Var((iv))$, *as follows;*

$$y^* = \begin{cases} y & \text{if } y \in U_{i,\phi} \cup V_{i,\phi} \\ y' & \text{if } y \in \{x_i\} \cup W_{i,\phi}. \end{cases}$$

We will refer to the first-order formula in (9) *as* $\alpha_i(\phi)$.

Basically, $\alpha_i(\phi)$ states that *if* (i) there exists an assignment that satisfies R, (ii) player i changes the object assigned to x_i, but (iii) the other players j play according to a Skolem function that is uniform with respect to what they can see (i.e., the objects assigned to the variables in $U_{j,\phi}$), *then* (iv) there exists an object d_i to assign to x'_i that guarantees truth of R no matter what is played by the players that can (in)directly see to x_i. The strategy such that $\hat{s}_i(\bar{u}) = d_i$ for all $\bar{u}$ that satisfy (i) is a weakly dominant strategy. It is a weakly dominant strategy in $G(\phi, M)$ for player i, because $\hat{s}_i \in S_{i,\phi}$.

Theorem 9 characterizes the condition under which a player has a weakly dominant strategy. To be true under WDS, however, slightly more is required. The following theorem characterizes truth under WDS.

Theorem 10. *Let* $\phi \in$ IF *be as in* (3) *and let* M *be a suitable model. Let* E_ϕ *and* A_ϕ *partition* $Var(\phi)$ *in such a way that* E_ϕ *contains the existentially quantified variables in* ϕ. *We abbreviate the string of all variables in* E_ϕ *and* A_ϕ *using* $\bar{e}$ *and* $\bar{a}$. *Then,*

$$M \models_{\text{WDS}} \phi \quad \textit{iff} \quad M \models_{\text{Tarski}} \alpha(\phi) \wedge \beta(\phi),$$

where $\alpha(\phi) = \bigwedge_{i \in N(E_\phi)} \alpha_i(\phi)$ *and* $\beta(\phi) = \forall\bar{a}\exists\bar{e}\, R(\bar{a}, \bar{e})$.

Formula $\alpha(\phi)$ being true on M is equivalent to every existential player i having a weakly dominant strategy $\hat{s}_i$ in $G(\phi, M)$. Yet this does not guarantee that the existential players i playing according to $\hat{s}_i$ will always get 1. For instance, in $G(\psi, M)$ every existential player has a weakly dominant strategy, if ψ's relational symbol is false for every suitable tuple of objects from M's domain. However, playing according to it will always yield an outcome of -1. Truth of $\beta(\phi)$ is a sufficient and necessary condition for avoiding the latter situations.

For future comparison we conclude this section with a meta-statement about IF logic interpreted under WDS, that follows straightforwardly from Theorem 10.

Theorem 11. IF *under WDS has less than elementary expressive power.*

5.5 Beyond WDS

From Theorem 9 we learn that for a player to have a weakly dominant strategy it does not matter what is played by his team members. Even in the case all its team members leave him and join the other team, this would not make a difference with respect to him having a weakly dominant strategy. I.e., WDS ignores the opportunities that might come with the notion of a *team*. In this section we show by example that increasing the 'powers' of the involved players in IF games increases the expressive power of IF logic on the obtained semantics, Theorem 14 as opposed to Theorem 11.

Let us revisit the sentence $\tau = \exists x_1 \exists x_2\, [x_1 = x_2]$. We observed that τ is not true under WDS on any model M that has a domain with more than one object (see Table 5.2). On the assumption that player 1 knows 2 is rational, player 1 may infer that 2 plays s_2^{copy}, because playing this strategy is better for it than any other strategy. That is, s_2^{copy} is weakly dominant. After this inference, player 1 choosing a strategy in $G(\tau, M)$ then effectively boils down to it choosing a strategy in the game

$$G'(\tau, (\{a, b\}, =)) \quad = \quad \left(\{1, 2\}, (\{s_1^a, s_1^b\}, \{s_2^{\text{copy}}\}), \{u_E, u_A\}\right).$$

G''s trivial payoff matrix is depicted in Table 5.3.

In this spirit, the following definition hard-wires the procedure of players calculating what other players will play. As such it bears strong similarity to

	s_2^{copy}
s_1^a	1
s_1^b	1

Table 5.3: The payoff matrix of $G'(\tau, (\{a, b\}, =)) = (\{1, 2\}, (\{s_1^a, s_1^b\}, \{s_2^{\text{copy}}\}), \{u_E, u_A\})$

the game-theoretic literature on *iterated removal of dominated strategies*, see Osborne and Rubinstein (1994).[2] The result of this procedure $\mathcal{P}$ as applied to some IF game will be the game that is effectively played.

Definition 12. *Let* $\phi \in$ IF *as in* (3) *and let* M *be a suitable model. Then, define*

$$\begin{aligned} G^n(\phi, M) &= G(\phi, M) = (N, (S_1, \ldots, S_n), \{u_E, u_A\}) \\ G^{i-1}(\phi, M) &= \left(N, (S_1, \ldots, S_{i-1}, S_i^{\mathcal{P}}, S_{i+1}^{\mathcal{P}}, \ldots, S_n^{\mathcal{P}}), \{u_E, u_A\}\right) \end{aligned}$$

where $S_1, \ldots, S_{i-1}, S_{i+1}^{\mathcal{P}} \ldots, S_n^{\mathcal{P}}$ *are copied from* $G^i(\phi, M)$ *and*

$$S_i^{\mathcal{P}} = \{s_i \in S_i \mid s_i \text{ weakly dominant in } G^i(\phi, M)\}.$$

Finally, put the strategic evaluation game $G^{\mathcal{P}}(\phi, M) = G^0(\phi, M)$.

This vehicle we employ to define a semantics 'on top' of WDS.

Definition 13. *Let* $\phi \in$ IF *and let* M *be a suitable model. Then we define truth of* ϕ *on* M *under* weak dominance semantics plus $\mathcal{P}$ *as follows:*

> $M \models^{\mathcal{P}}_{\mathrm{WDS}} \phi$ *iff in* $G^{\mathcal{P}}(\phi, M)$ *for every complete profile* $\hat{s}$ *it is the case that* $u_E(\hat{s}) = 1$.

We thus state the truth of an IF sentence ϕ on M in terms of the outcome of playing the game $G^{\mathcal{P}}(\phi, M)$ by players that are empowered to reason according to the procedure $\mathcal{P}$. For instance, it is the case that $(\{a, b\}, =) \models^{\mathcal{P}}_{\mathrm{WDS}} \tau$.

First of all, note that, epistemically, player n needs to know nothing about the other players in order to pick a weakly dominated strategy, i.e., to act in accordance with $\mathcal{P}$. Now, player $n - 1$ needs to know that player n is indeed rational in order for it to be rational to consider game $G^n(\phi, M)$. In general, to explain why the players would execute $\mathcal{P}$, one has to assume that every player i is rational and i knows that $i + 1$ knows that ... knows that n is rational. Now, this is quite strong an assumption to make. Much stronger in any case than WDS's mere requirements that all the players are rational.

Secondly, we observe that for $\phi \in$ IF

$$M \models_{\mathrm{WDS}} \phi \text{ implies } M \models^{\mathcal{P}}_{\mathrm{WDS}} \phi \quad \text{and} \quad M \models^{\mathcal{P}}_{\mathrm{WDS}} \phi \text{ implies } M \models_{\mathrm{GTS}} \phi, \quad (10)$$

but the converses do not hold, witnessing τ and θ on $M = (\{a, b\}, =)$, respectively.

[2] It is tempting to clarify the inferences of the players by assuming *common knowledge of rationality.* (In fact a weaker concept of knowledge would do to trigger the procedure.) In this paper we consider the procedures simply as formal objects, leaving us space to define procedures that are not epistemologically justified (such as $\mathcal{ND}$, defined below). For much more on epistemological characterizations of game-theoretic solution concepts we refer to de Bruin (2004).

Thirdly, in Theorem 14 we observe that the expressive power increases when switching from $\models_{\mathrm{WDS}}$ to $\models^{\mathcal{P}}_{\mathrm{WDS}}$ with respect to FOL. Also, we draw the conclusion from this theorem that every FOL formula behaves under WDS plus $\mathcal{P}$ as it does under Tarski semantics. What the expressive power of IF logic is under WDS plus $\mathcal{P}$ is left open.

Theorem 14. *Let* $\phi \in$ FOL *and let* M *be a suitable model. Then,*

$$M \models_{\mathrm{Tarski}} \phi \quad \textit{iff} \quad M \models^{\mathcal{P}}_{\mathrm{WDS}} \phi.$$

The procedure $\mathcal{P}$ turns out to be the strategic counterpart of the *backwards induction algorithm* as applied to the extensive game tree of an FOL game. The proof of Theorem 14 boils down to showing that a tuple of Skolem functions $\bar{f}$ is a witness of $M \models_{\mathrm{Tarski}} \phi$ iff it is contained in $S^{\mathcal{P}}_1 \times \cdots \times S^{\mathcal{P}}_n$.

5.6 Conclusion

In this paper, we set up a strategic framework for IF semantic games, which are traditionally studied extensively. Naturally, by giving up the extensive structure that is traditionally given to IF games, we avoid conceptual issues that arise with the playability of IF games (i.e., lack of perfect recall). We observed that truth of IF logic under GTS can be characterized by the solution concept of Nash equilibrium. We saw that other issues arise in the strategic framework: how are players supposed to coordinate or, more eloquently, how are Nash equilibria supposed to arise?

We used the strategic framework to define two semantic interpretations for IF logic inspired by solution concepts related to weakly dominant strategies: $\models_{\mathrm{WDS}}$ and $\models^{\mathcal{P}}_{\mathrm{WDS}}$. The former does not require any of the involved players to know anything about the other players. We showed that under $\models_{\mathrm{WDS}}$ the expressive power of IF logic collapses to that of a fragment of first-order logic (under Tarski semantics). The epistemic demands of $\models^{\mathcal{P}}_{\mathrm{WDS}}$ were seen to be higher than that of $\models_{\mathrm{WDS}}$. We showed that the expressive power of FOL (under Tarski semantics) is left intact when evaluated under $\models^{\mathcal{P}}_{\mathrm{WDS}}$. Thus, all of IF logic (under $\models^{\mathcal{P}}_{\mathrm{WDS}}$) has expressive power of at least FOL (under Tarski semantics). Our findings can be summarized in the following table:

Solution concept $\mathcal{S}$	Expressive power $\models_{\mathcal{S}}$
Nash equilibrium	High ($=\Sigma^1_1$)
WDS + $\mathcal{P}$	Medium-high ($\geq$FOL)
WDS	Low ($<$FOL)

Further research will have to flesh out this table and determine what are the dependencies between solution concepts and the expressive power of IF logic evaluated under the associated solution concept. This enterprise would explore correlations between notions of agency and semantic interpretations of logical languages.

Acknowledgments

I gratefully acknowledge Peter van Emde Boas, Denis Bonnay, Theo Janssen, Sieuwert van Otterloo, Gabriel Sandu, Tero Tulenheimo, and the anonymous reviewers for valuable discussion and comments. An early version of this paper was presented at the *2004 Prague Colloquium* on *Logic, Games and Philosophy: Foundational Perspectives.* I thank the organizers for letting me present my research.

References

Cameron, P. J. and Hodges, W. (2001). Some combinatorics of imperfect information. *Journal of Symbolic Logic*, 66(2):673–684.

de Bruin, B. (2004). *Explaining Games, On the Logic of Game Theoretic Explanations*. Ph.D. thesis, ILLC, Universiteit van Amsterdam.

Dechesne, F. (2005). *Game, Set, Maths: Formal Investigations into Logic with Imperfect Information*. Ph.D. thesis, SOBU, Tilburg university and Technische Universiteit Eindhoven.

Henkin, L. (1959). Some remarks on infinitely long formulas. *Infinitistic Methods*, Proceedings of the Symposium on Foundations of Mathematics, Warsaw, 167–183.

Hintikka, J. (1996). *Principles of Mathematics Revisited*. Cambridge University Press, Cambridge.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. F. A. K. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–481. North Holland, Amsterdam.

Janssen, T. M. V. (2002). Independent choices and the interpretation of IF logic. *Journal of Logic, Language and Information*, 11:367–387.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT, Cambridge, MA.

Pietarinen, A. and Tulenheimo, T. (2004). An introduction to IF logic. Lecture notes for the 16th European Summer School in Logic, Language and Information.

Sandu, G. and Pietarinen, A. (2003). Informationally independent connectives. In Mints, G. and Muskens, R., editors, *Logic, Language and Computation*, volume 9, pages 23–41. CSLI, Stanford.

Sevenster, M. (2006). *Branches of Imperfect Information: Logic, Games, and Computation*. Ph.D. thesis, ILLC, University of Amsterdam.

van Benthem, J. F. A. K. (2000). Logic and games, lecture notes. Draft version.

van Benthem, J. F. A. K. (2004). Probabilistic features in logic games. In Kolak, D. and Symons, D., editors, *Quantifiers, Questions and Quantum Physics*, pages 189–194. Springer, Dordrecht.

Walkoe, W. (1970). Finite partially-ordered quantification. *Journal of Symbolic Logic*, 35: 535–555.

Chapter 6

TOWARDS EVALUATION GAMES FOR FUZZY LOGICS

Petr Cintula[1*] and Ondrej Majer[2†]
[1] *Institute of Computer Science, Academy of Sciences of the Czech Republic*
cintula@cs.cas.cz

[2] *Institute of Philosophy, Academy of Sciences of the Czech Republic*
majer@site.cas.cz

Abstract The article provides two kinds of game-theoretical semantics for fuzzy logics with special attention to Łukasiewicz logic. The first one is a generalization of the evaluation games for classical logic. It is shown that it provides an interesting contribution to the model theory of fuzzy logics as, unlike the standard semantics, it can deal with the so-called non-safe models. The second kind of semantics makes explicit the intuition about fuzzy logics as logics of partial truth and provides a semantics in the form of a bargaining game. Finally, a basic kind of logic of informational independence of a Hintikka-Sandu style is introduced.

6.1 Introduction

Both the areas of mathematical fuzzy logics and game-theoretical semantics have been extensively studied, but their intersection has received attention only quite recently. The literature on games in fuzzy logics concentrates on the proof-theoretical features of fuzzy logics and uses the framework of the Lorenzen-style dialogue games (e.g. Fermüller, 2003). Other applications of game-theory in fuzzy logics are Renyi-Ulam games used in Mundici (1993) to provide an alternative semantics for finite-valued Łukasiewicz logics. The aim of this article is to provide a game-theoretic semantics for the model theory of fuzzy logics.

Our aim is to explore the game semantics for a general fuzzy logic. However, we find it useful to restrict ourselves in this paper to a particular system:

*The work was supported by project 1M0545 of the Ministry of Education, Youth and Sports of the Czech Republic.
†Partial funding provided by grant GA401/04/0117. We wish to thank Christian Fermüller, Tero Tulenheimo and an anonymous referee for valuable comments on a previous version of this paper.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

the Łukasiewicz logic. Nice properties of this system illuminate motivations of the game semantics in the many-valued setting and provide an intuitive meaning to the rules of the game. The proposed semantics can be extended to the remaining basic fuzzy logics—product and Gödel, as well as for some other fuzzy logics. This extension is beyond the scope of the article and will be a part of the future work.

After the necessary preliminaries (Section 6.2), we introduce in the Section 6.3 an evaluation game of the Hintikka-Sandu style. It is a rather straightforward generalization of the evaluation games for classical logic, however, it has several interesting features. The game-theoretic interpretation of Łukasiewicz logic is in a certain sense more general than the usual Tarskian one (e.g. Hájek, 1998). In particular, the standard interpretation has to be limited to the so-called safe models in which all the suprema and infima (required by the standard interpretation of the existential and universal quantifiers) exist. The requirement of safeness—a crucial point of the standard interpretation—can be partially avoided by the proposed game semantics (this is discussed in Section 6.4). In Section 6.5 we introduce the notion of the bargaining fuzzy game—a non-zerosum version of the evaluation game that better captures some game intuitions about fuzzy logics as the logics of partial truth.

In Section 6.6 we explore the notion of imperfect information in the context of the bargaining fuzzy games. We do not discuss the full range of imperfection as studied in the framework of independence-friendly logics (see Hintikka and Sandu, 1989; Sandu, 1993). We restrict ourselves to the basic case of independence of existential and general quantifiers (the quantified variable might be independent of the quantifiers of the other kind to which it is syntactically subordinated). As in standard IF logics we obtain formulas lacking a truth value. These formulas have a range of truth degrees (a subinterval of [0,1]), such that none of the players has a winning strategy for the corresponding game (in a sense we have a formula which is neither partially true nor partially false).

6.2 Preliminaries

6.2.1 Łukasiewicz propositional logic

Łukasiewicz propositional logic was introduced in Łukasiewicz (1920). Here we survey its basic properties, for details about predicate logic see the Appendix. Its formulas are built from the propositional variables using the logical connectives in the usual way. There are various axiomatic systems for Łukasiewicz logic, we present the usual one:

(Ł1) $\varphi \to (\psi \to \varphi)$
(Ł2) $(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$
(Ł3) $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$
(Ł4) $((\varphi \to \psi) \to \psi) \to ((\psi \to \varphi) \to \varphi)$

In Hájek (1998) the Łukasiewicz logic was put into the context of fuzzy logics. It was shown that the Łukasiewicz logic is the extension of the so-called Basic Fuzzy Logic by one axiom (double negation law).

Now we introduce the algebraic semantics for this logic, for the sake of simplicity we present just the so-called standard one (on the interval $[0, 1]$). The semantics on more general algebraic structures will be given in the Appendix. We start with the following basic operations: binary $\oplus$, unary $\neg$, and nullary 0; their corresponding logical connectives are called strong disjunction, negation and bottom. Other operations (logical connectives[1]) are defined in the following way (we list them together with their standard semantics and the names of the corresponding logical connectives):

$x \otimes y$	is	$\neg(\neg x \oplus \neg y)$	$\max(0, x + y - 1)$	strong conjunction
$x \vee y$	is	$(x \ominus y) \oplus y$	$\max(x, y)$	weak-disjunction
$x \wedge y$	is	$\neg(\neg x \vee \neg y)$	$\min(x, y)$	weak-conjunction
$x \rightarrow y$	is	$\neg x \oplus y$	$\min(1, 1 - x + y)$	implication
1	is	$\neg 0$	1	top

Definition 1. *The standard* MV*-algebra (*$[0, 1]_S$*) has the domain* $[0, 1]$ *and the operations:*

$$\begin{aligned} x \oplus y &= \min(1, x + y) \\ \neg x &= 1 - x \\ 0 &= 0 \end{aligned}$$

We define the basic syntactical and semantical notions (theory, proof, provability, evaluation, model, and tautology) as usual. A completeness theorem with respect to the standard semantics can be proven.

Theorem 2 (Standard completeness theorem)**.** *Let φ a formula. Then φ is provable in Łukasiewicz logic iff it is a tautology in standard* MV*-algebra.*

6.2.2 Evaluation games for classical logic

Our point of departure is the standard evaluation game for the classical predicate logic as described in Hintikka and Sandu (1997) and Sandu (1993). Evaluation games provide an alternative semantics for first-order formulas: it is a zero-sum game of two players traditionally called Eloise and Abelard, who take up the roles of Verifier ($\mathcal{V}$) and Falsifier ($\mathcal{F}$) in the course of the game. The goal of Eloise (the initial Verifier) is to show validity of the initial formula in a fixed model $\mathbf{M}$ and an initial $\mathbf{M}$-evaluation v. A history is any sequence of moves starting from the initial position, and a strategy is a function from a set of histories to the set of admissible moves corresponding to the last position in

[1] In this paper we will use the same symbols both for logical connectives and operations in the corresponding algebraic semantics.

the history. Admissible moves in a current position of the game are given by the logical structure of the current subformula and evaluation:

Rules of the classical evaluation game—M-Game (φ, v).

- $\varphi = \psi_1 \vee \psi_2$: $\mathcal{V}$ chooses whether to play (ψ_1, v) or (ψ_2, v).
- $\varphi = \psi_1 \wedge \psi_2$: $\mathcal{F}$ chooses whether to play (ψ_1, v) or (ψ_2, v).
- $\varphi = \neg\psi$: *role switch*, the game continues as (ψ, v) with the roles reversed (current $\mathcal{V}$ becomes $\mathcal{F}$ and vice-versa).
- $\varphi = (\forall x)\psi$: $\mathcal{F}$ chooses $a \in M$, the game continues as $(\psi, v[x : a])$.
- $\varphi = (\exists x)\psi$: $\mathcal{V}$ chooses $a \in M$, the game continues as $(\psi, v[x : a])$.
- φ is an atomic formula: the end of the game—$\mathcal{V}$ wins if $\mathbf{M} \models \varphi[v]$, otherwise $\mathcal{F}$ wins.

The evaluation game is a zero-sum game of finite depth with perfect information, so it is determined according to the Zermelo's theorem. Thus in any evaluation game, either Eloise or Abelard has a winning strategy. We define the game-theoretical truth as the existence of a winning strategy for the Eloise. It is known that the game semantics corresponds with the standard Tarskian one.

Theorem 3. *Assuming the Axiom of Choice the following holds: Eloise has a winning strategy for the* $\mathbf{M}$*-Game* (φ, v) *iff* $\mathbf{M} \models \varphi[v]$.

6.3 Evaluation games for Łukasiewicz logic

Fuzzy logics can be seen as logics of partial truth—we are not just interested if the formula is true, but "how much" it is true. In particular, we want to know in which degree a formula is true in the given model (and evaluation). The fuzzy evaluation game will be a generalization of the standard evaluation game for the classical logic with an additional parameter representing the degree of truth of the formula in question. We shall start with the evaluation game for Łukasiewicz logic on the interval $[0, 1]$, but our results can be generalized to other fuzzy logics and more general structures.

We can give the fuzzy game a gambling motivation. Imagine the owner of your local casino introduced a new game. The bookmaker (=Abelard) lets you (=Eloise) bet on (the lower estimation of) the degree of truth of an assertion (expressed in a formal language of Łukasiewicz logic). Then you play with the bookmaker a game to justify your bet. If you win the play of the game, you get the corresponding proportion of the stake of 100 €. If you lose, you get nothing. What is the reasonable price for you to bet on the assertion? Obviously, your price should not be higher than your possible gain, i.e., $r \cdot 100$ €, where r is your estimation of its truth degree.

As in the classical evaluation game, the players take the roles of Verifier and Falsifier during the course of the game. The rules of the fuzzy game are a "conservative" extension of the rules of the classical evaluation game. It is obvious that we need to take care of the parameter r. It is less obvious that the procedural symmetry of the classical game is lost. In general, each round of the fuzzy game consists of two moves. In the first move $\mathcal{V}$ (possibly) modifies r, the second one is a choice for $\mathcal{F}$. For the classical connectives one of these moves is trivial, so the rules for them turn out to be the classical ones, i.e., the parameter r does not change.

Let us assume, more formally, that you play a game for a formula φ on a model $\mathbf{M}$ and an $\mathbf{M}$-evaluation v. In the beginning of the game you, as the initial Verifier, claim that $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v} \geq r$. To give an intuitive meaning to the rules of the game we shall continue our gambling motivation. As r represents the amount of money you are betting in the current subgame, we shall call it *stake*.

6.3.1 The rules of an evaluation game over $[0,1]_S$

We define rules of the game according to the logical form of φ. We could confine ourselves to the basic set of connectives of Łukasiewicz logic, but we prefer to have rules for all of them (except of $\rightarrow$, $\leftrightarrow$) in order to make game intuitions behind the fuzzy games more transparent. Let $\mathbf{M}$ be a $[0,1]_S$-model. $\mathbf{M}$-Game (φ, v, r) is given by a formula φ, an $\mathbf{M}$-evaluation v, and an element $r \in [0,1]$. The positions of this game are given by a triple (φ', v', r') where φ' is a subformula of φ.

Terminating rules. Atomic formulas are interpreted in the same way as in the classical games. They correspond to the terminating positions of a game, in which the corresponding action is a test of the current degree of truth of the (atomic) formula (which is determined by the parameters of the game). We have to take care of the zero value of the stake. As $\|\varphi\|_{\mathbf{M},v} \geq 0$ is always true for any formula φ, it is redundant to play the corresponding subgame. According to our motivation, zero stake means win for the (current) Verifier in the respective subgame. It is natural to consider this as a trivial win of $\mathcal{V}$.

(*at*) (ψ, v, r), where ψ is an atomic formula: the end of the game, if $\|\psi\|_{\mathbf{M},v} \geq r$ (the current) $\mathcal{V}$ wins, otherwise $\mathcal{F}$ wins.

(0) $(\varphi, v, 0)$: the end of the game, the current $\mathcal{V}$ wins.

Choice rules—disjunction. The moves for disjunction consist of the distribution of the current stake between the disjuncts. The difference between the classical (weak) and strong disjunction is that in the strong case $\mathcal{V}$ distributes the stake between the disjuncts and $\mathcal{F}$ chooses one of them while in the weak case $\mathcal{V}$ just moves all the stake to one of the disjuncts. $\mathcal{F}$ does not really have

a choice here as the choice of the disjunct with stake 0 means an immediate loss for him. So the choice of the disjunct is in fact made by Verifiers's move. The move for the weak disjunction is a special case of the strong disjunction: it corresponds to the fact that the classical connective is weaker ($\varphi \vee \psi \to \varphi \oplus \psi$ is a tautology of the Łukasiewicz logic).

($\oplus$) $(\psi_1 \oplus \psi_2, v, r)$: $\mathcal{V}$ chooses $r' \leq r$ and $\mathcal{F}$ chooses whether to play (ψ_1, v, r') or $(\psi_2, v, r - r')$.

($\vee$) $(\psi_1 \vee \psi_2, v, r)$: $\mathcal{V}$ chooses whether to play (ψ_1, v, r) or (ψ_2, v, r).

Negation. Łukasiewicz negation as in the classical case corresponds to the role switch. However, it includes a modification of the stake as well. This should be no surprise if we consider r and $1 - r$ as the stakes of $\mathcal{V}$ and $\mathcal{F}$ respectively. The same thing is to claim that $\|\psi\| = r$ and $\|\neg\psi\| = 1 - r$ for any ψ. If $\mathcal{F}$ disagrees that $\|\neg\varphi\| \geq r$, he has to claim not just that $\|\varphi\| \geq 1 - r$ but that $\|\varphi\| > 1 - r$, i.e., $\|\varphi\| \geq 1 - r + r'$ for some small r'. In gambling terms: if $\mathcal{F}$ wants to take the role of $\mathcal{V}$ for the currently negated formula, he has to increase his stake.

($\neg$) $(\neg\psi, v, r)$: $\mathcal{F}$ chooses $r', r \geq r' > 0$, *role switch*, game continues as $(\psi, v, (1 - r) + r')$

Conjunction. Playing the weak conjunction $\mathcal{V}$ just waits for the choice of the conjunct by $\mathcal{F}$ (as in the classical case) and the stake is unchanged. The move for the strong conjunction is more tricky: $\mathcal{V}$ divides the Falsifier's original stake (in fact Falsifier's bet on the negation of the conjunction) between two conjuncts, $\mathcal{F}$ chooses one of them and adds the original Verifier's stake to the stake on the chosen conjunct. This move is, in fact, a dual to the dividing stake by $\mathcal{V}$ in the strong disjunction move. The correspondence of the game moves and the logical meaning of two conjunctions is reflected by the fact that it is harder to play strong conjunction than the weak one as the player has to bet more.

($\otimes$) $(\psi_1 \otimes \psi_2, v, r)$: $\mathcal{V}$ chooses $r' \leq 1 - r$ and $\mathcal{F}$ chooses whether to play $(\psi_1, v, r + r')$ or $(\psi_2, v, r + (1 - r - r'))$.

($\wedge$) $(\psi_1 \wedge \psi_2, v, r)$: $\mathcal{F}$ chooses whether to play (ψ_1, v, r) or (ψ_2, v, r).

Quantifiers. As in the classical game the move for the existentially quantified formula $(\exists x)\psi$ consists of Verifier's choice of an element from the domain of the model witnessing the claim $\|(\exists x)\psi\|_v \geq r$, which is equivalent to $\sup(\|\psi\|_{v[x]}) \geq r$ (by $\sup(\|\psi\|_{v[x]})$ we mean the supremum of the set$\{\|\psi\|_{v[x:a]} \mid a \in M\}$). If the supremum is witnessed, i.e., there is an a' such that $\|\psi\|_{v[x:a']} =$

$\sup(\|\psi\|_{v[x]})$ (the supremum is in fact a maximum), we could leave the classical rule as it is. However, if the supremum is proper, i.e., $\|\psi\|_{v[x:a']} < \sup(\|\psi\|_{v[x]})$ for all $a' \in M$, $\mathcal{V}$ is, in principle, not able to find a witness for the existential claim. As we want to make this case a win for $\mathcal{V}$ to keep the correspondence with the standard semantics of fuzzy predicate logic, we have to do $\mathcal{V}$ a favor and make the winning condition for $\mathcal{V}$ weaker: it is not necessary to give a witness for r, it suffices to do it for any r' strictly smaller than r. In other words, we let $\mathcal{F}$ to decrease Verifier's stake (it is Falsifier's interest to decrease it as little as possible) and *then* $\mathcal{V}$ to find a witness in the domain to meet the weakened condition. We keep Falsifier's winning condition in the position $((\exists x)\psi, v, r)$ the same for r strictly greater than $\sup(\|\psi\|_{v[x]})$ ($\mathcal{F}$ can always win by choosing r' between the supremum and r).

(∃) $((\exists x)\psi, v, r)$: $\mathcal{F}$ chooses $r' < r$ and $\mathcal{V}$ chooses $a \in M$, the game continues as $(\psi, v[x : a], r')$.

For a witnessed structure we can keep the classical rule:

(∃′) $((\exists x)\psi, v, r)$: $\mathcal{V}$ chooses $a \in M$, game continues as $(\psi, v[x : a], r)$.

The position $((\forall x)\psi, v, r)$ corresponds to Verifier's claim that $\inf(\|\psi\|_{v[x]}) \geq r$. $\mathcal{F}$ is to move and he has to provide a counterexample, i.e., to find an a' such that $(\|\psi\|_{v[x:a']} < r)$. It is clear that the existence of the witnessing element does not influence Falsifier's choice, so we need not include a change of the stake in the rule for the universal quantifier and it turns out to be the same as in the classical case.

(∀) $((\forall x)\psi, v, r)$: $\mathcal{F}$ chooses $a \in M$, game continues as $(\psi, v[x : a], r)$.

The asymmetry of the quantifier rules might be explained when defining existential move from the general one using twice the negation rule.

6.3.2 The rules of an evaluation game over an MV-algebra

The rules are a direct analogy of the corresponding rules for **M**-game, we only replace operations on the unit interval with those in **L**. There is a small problem, that the MV-chain **L** could be atomic and the rule (∃) would fail. However, as we know that each structure over atomic MV-chain is witnessed we use the rule (∃′) instead.

Let $\mathbf{L} = (L, \oplus, \neg, 0)$ be an MV-chain and **M** an **L**-structure. $(\mathbf{M}, \mathbf{L})$-Game (φ, v, r) is defined relative to a formula φ, an **M**-evaluation v, and an element

$r \in L$. The following are the rules of $(\mathbf{M}, \mathbf{L})$-Game for non-atomic MV-chain $\mathbf{L}$:

(*at*) (φ, v, r), where φ is an atomic formula: $\mathcal{V}$ wins iff $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v} \geq r$;

(0) $(\varphi, v, 0)$: $\mathcal{V}$ wins;

($\oplus$) $(\psi_1 \oplus \psi_2, v, r)$: $\mathcal{V}$ chooses $r' \leq r$ and $\mathcal{F}$ chooses whether to play (ψ_1, v, r') or $(\psi_2, v, r \ominus r')$;

($\vee$) $(\psi_1 \vee \psi_2, v, r)$: $\mathcal{V}$ chooses whether to play (ψ_1, v, r) or (ψ_2, v, r);

($\otimes$) $(\psi_1 \otimes \psi_2, v, r)$: $\mathcal{V}$ chooses $r' \leq \neg r$ and $\mathcal{F}$ chooses whether to play $(\psi_1, v, r \oplus r')$ or $(\psi_2, v, r \oplus (\neg r \ominus r'))$;

($\wedge$) $(\psi_1 \wedge \psi_2, v, r)$: $\mathcal{F}$ chooses whether to play (ψ_1, v, r) or (ψ_2, v, r);

($\neg$) $(\neg\psi, v, r)$: $\mathcal{F}$ chooses $r', r \geq qr' > 0$, *role switch*, game continues as $(\psi, v, \neg r \oplus r')$;

($\forall$) $((\forall x)\psi, v, r)$: $\mathcal{F}$ chooses $a \in M$, game continues as $(\psi, v[x : a], r)$;

($\exists$) $((\exists x)\psi, v, r)$: $\mathcal{F}$ chooses $r' < r$ and $\mathcal{V}$ chooses $a \in M$, the game continues as $(\psi, v[x : a], r')$.

The rules of $(\mathbf{M}, \mathbf{L})$-Game for *atomic* MV-chain $\mathbf{L}$ are the some as above, we only replace the rule ($\exists$) by the rule ($\exists'$):

($\exists'$) $((\exists x)\psi, v, r)$: $((\exists x)\psi, v, r)$: $\mathcal{V}$ chooses $a \in M$, game continues as $(\psi, v[x : a], r)$.

It will be obvious from the proof of the correspondence theorem that in witnessed structures we can use the rule ($\exists'$) instead of the rule ($\exists$) and, on the other hand, in games for atomic MV-chains we *have to* use the new rule.

Notice that we do not need to assume that $\mathbf{M}$ is a *safe* $\mathbf{L}$-structure. The rules introduced in the previous section work also for non-safe structures. In particular, we do not have to change the existential rule. If a supremum does not exist (we can think of it as of a gap in the interval of truth values), it cannot be used neither by $\mathcal{V}$ nor by $\mathcal{F}$. We can stay with the understanding of Verifier's winning strategy as an arbitrarily close approximation of the sup.

6.3.3 The correspondence theorem

Fuzzy evaluation games are zero-sum games of a finite depth, so by Zermelo's theorem they are determined. We can prove the correspondence between the existence of winning strategies in a fuzzy game and the standard Tarskian truth.

Theorem 4. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be a safe* $\mathbf{L}$*-structure,* φ *a formula,* v *an* $\mathbf{M}$*-valuation, and* $r \in L$*. Then Eloise has a winning strategy for the* $(\mathbf{M}, \mathbf{L})$*-Game* (φ, v, r) *iff* $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v} \geq r$*.*

Proof. We prove the claim by dual induction over the complexity of φ. We have to prove both directions. As we have not showed any mutual interderivability of the rules of $(\mathbf{M}, \mathbf{L})$-Game, we have to prove it for all those rules. Step 0: If φ is an atomic formula the claim is trivial. Step $n \to n+1$,

• $\varphi = \psi_1 \vee \psi_2$: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi_1 \vee \psi_2, v, r)$, then he has a winning strategy σ_1 for the game (ψ_1, v, r) or a winning strategy σ_2 for the games (ψ_2, v, r). By induction property we get $\|\psi_1\|_v \geq r$ or $\|\psi_2\|_v \geq r$. Obviously, $\|\psi_1 \vee \psi_2\|_v = \|\psi_1\|_v \vee \|\psi_2\|_v \geq r$.

If $\|\psi_1 \vee \psi_2\|_v \geq r$, then $\|\psi_1\|_v \vee \|\psi_2\|_v \geq r$. Thus $\mathcal{V}$ can choose $i \in \{1, 2\}$ such that $\|\psi_i\|_v \geq r$. By induction property $\mathcal{V}$ has a winning strategy σ_i for $(\mathbf{M}, \mathbf{L})$-Game (ψ_i, v, r). So $\mathcal{V}$ has a winning strategy for (φ, v, r).

• $\varphi = \psi_1 \wedge \psi_2$: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi_1 \wedge \psi_2, v, r)$, then he has a winning strategy σ_i for the game (ψ_i, v, r) for $i \in \{1, 2\}$. By induction property we get $\|\psi_1\|_v \geq r$ and $\|\psi_2\|_v \geq r$. Obviously, $\|\psi_1 \wedge \psi_2\|_v = \|\psi_1\|_v \wedge \|\psi_2\|_v \geq r$.

If $\|\psi_1 \vee \psi_2\|_v \geq r$, then $\|\psi_1\|_v \vee \|\psi_2\|_v \geq r$. Thus for $i \in \{1, 2\}$ we have $\|\psi_i\|_v \geq r$. By induction property $\mathcal{V}$ has a winning strategy σ_i for $(\mathbf{M}, \mathbf{L})$-Game (ψ_i, v, r) for $i \in \{1, 2\}$. So $\mathcal{V}$ has a winning strategy for (φ, v, r).

• $\varphi = \psi_1 \oplus \psi_2$: the winning strategy σ of $\mathcal{V}$ for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi_1 \oplus \psi_2, v, r)$ consists of a choice of an $r' \in L$, $r' \leq r$ and a pair of strategies σ_1, σ_2 which are winning for the games (ψ_1, v, r'), $(\psi_2, v, r \ominus r')$ respectively. σ_1 is a response to Falsifier's choice of the left disjunct, by induction property we get $\|\psi_1\|_v \geq r'$, similarly σ_2 being winning strategy for the right conjunct implies $\|\psi_2\|_v \geq r \ominus r'$. Obviously, $\|\psi_1 \oplus \psi_2\|_v = \|\psi_1\|_v \oplus \|\psi_2\|_v \geq r' \oplus (r \ominus r') \geq r$ (the last inequality follows from the properties of MV-algebras).

If $\|\psi_1 \oplus \psi_2\|_v \geq r$, then $\|\psi_1\|_v \oplus \|\psi_2\|_v \geq r$. $\mathcal{V}$ chooses $r' = \|\psi_1\|_v$. Since $r' \oplus \|\psi_2\|_v \geq r$ we get that $\|\psi_2\|_v \geq r \ominus r'$. By induction property $\mathcal{V}$ has a pair of winning strategies σ_1, σ_2 for $(\mathbf{M}, \mathbf{L})$-Games (ψ_1, v, r') and $(\psi_1, v, r \ominus r')$ respectively. So $\mathcal{V}$ has a response for any choice of $\mathcal{F}$, thus he has a winning strategy for (φ, v, r).

• $\varphi = \psi_1 \otimes \psi_2$: analogous to the previous case.

• $\varphi = \neg\psi$: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\neg\psi, v, r)$, then according to the rules of the fuzzy game he has a winning strategy σ' as Falsifier in the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v, \neg r \oplus r')$ for any $r' \neq 0$. By induction property we get $\|\psi\|_v < \neg r \oplus r'$ for each $r' > 0$, thus $\|\psi\|_v \leq \neg r$. Finally, $\|\neg\psi\|_v \geq r$.

If $\|\neg\psi\|_v \geq r$, then $\|\psi\|_v \leq \neg r$. Thus $\|\psi\|_v < \neg r \oplus r'$ for each $r' > 0$. By the induction property there is a winning strategy σ for the Falsifier in the game

$(\psi, v, \neg r \oplus r')$, which implies existence of a winning strategy σ' for the Verifier in the game $(\neg\psi, v, r)$.

• $\varphi = \forall x\psi(x)$: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\forall x\psi(x), v, r)$ he must have for any $a \in M$ a winning strategy σ_a in the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a], r)$. Thus, we get $\|\psi\|_{v[x=a]} \geq r$ for each a (by induction property) and so $\|\varphi\|_v \geq r$.

If $\|\forall x\psi(x)\|_v \geq r$, then for each $a \in M$ we have $\|\psi\|_{v[x=a]} \geq r$. By induction property we know that $\mathcal{V}$ has a winning strategy σ_a for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a], r)$ for any $a \in M$. Thus he has a winning strategy for the $(\mathbf{M}, \mathbf{L})$-Game $(\forall x\psi(x), v, r)$.

• $\varphi = \exists x\psi(x)$ and $\mathbf{L}$ is not atomic: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\exists x\psi(x), v, r)$ he must be able to chose $a \in M$ for each $r' < r$ such that he has a winning strategy $\sigma_{r'}$ in the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a], r')$. Thus, for each $r' < r$ we get $\|\psi\|_{v[x=a]} \geq r'$ for some a (by induction property) and so $\|\varphi\|_v \geq r'$ for each $r' < r$. This implies (as $\mathbf{L}$ is not atomic and so r is a cumulative point) that $\|\varphi\|_v \geq r$.

If $\|\exists x\psi(x)\|_v \geq r$, then for each $r' < r$ there is $a_{r'} \in M$ such that $\|\psi\|_{v[x=a]} \geq r'$. By induction property we know that $\mathcal{V}$ has a winning strategy $\sigma_{r'}$ for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a_{r'}], r')$ for any $r' < r$. Thus he has a winning strategy for the $(\mathbf{M}, \mathbf{L})$-Game $(\exists x\psi(x), v, r)$.

• $\varphi = \exists x\psi(x)$ and $\mathbf{L}$ is atomic: if $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\exists x\psi(x), v, r)$ he must be able to chose $a \in M$ to have a winning strategy σ in the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a], r)$. Thus, we get $\|\psi\|_{v[x=a]} \geq r$ for some a (by induction property) and so $\|\varphi\|_v \geq r$.

If $\|\exists x\psi(x)\|_v \geq r$, then there is $a \in M$ such that $\|\psi\|_{v[x=a]} \geq r$ (as $(\mathbf{M}, \mathbf{L})$ is witnessed model). By induction property we know that $\mathcal{V}$ has a winning strategy σ for the $(\mathbf{M}, \mathbf{L})$-Game $(\psi, v[x = a], r)$. Thus he has a winning strategy for the $(\mathbf{M}, \mathbf{L})$-Game $(\exists x\psi(x), v, r)$. □

It immediately follows that tautologies of Łukasiewicz logic correspond to the strategies for r=1 in the respective game.

Corollary 5. *Let φ be a formula,* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be a safe* $\mathbf{L}$*-structure, and an* $\mathbf{M}$*-evaluation v. Then Eloise has a winning strategy for* $(\mathbf{M}, \mathbf{L})$*-Game* $(\varphi, v, 1)$ *iff* $(\mathbf{M}, \mathbf{L}) \models \varphi[v]$.

As we have seen, winning strategies are based on players' knowledge of the values of subformulas and the operations in the corresponding algebras (as they do in the classical games). This does not undermine our gambling motivation, however. Players must know the meaning of the connectives, i.e., the operations in the corresponding algebra. But they might not be able to calculate the exact value of the subformulas (e.g., because of their complexity). Playing the game allows a player *to estimate* the value of the formula in question.

6.4 Winning strategies and safe structures

In games over safe structures (defined in the previous sections) we identified the (game-theoretical) value of a formula with the existence of Eloise's winning strategy for a certain value of a formula. We define two sets induced by (non-)existence of winning strategies and use them for analyzing safeness of fuzzy predicate structures.

Definition 6. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be an* $\mathbf{L}$*-structure,* φ *a formula, and* v *an* $\mathbf{M}$*-valuation. We define:*

$\mathcal{WS}^+(\mathbf{M},\mathbf{L},v,\varphi) =_{\mathrm{df}} \{r \mid$ *Eloise has a winning strategy for the* $(\mathbf{M},\mathbf{L})$*-Game* $(\varphi,v,r)\}$.

$\mathcal{WS}^-(\mathbf{M},\mathbf{L},v,\varphi) =_{\mathrm{df}} \{r \mid$ *Abelard has for any* $r' > r$ *a winning strategy for the* $(\mathbf{M},\mathbf{L})$*-Game* $(\varphi,v,r')\}$.

If $\mathbf{L}$ and $\mathbf{M}$ are clear from the context we will write $\mathcal{WS}^+(v,\varphi)$ instead of $\mathcal{WS}^+(\mathbf{M},\mathbf{L},v,\varphi)$. Furthermore, if v is clear from the context we will write only $\mathcal{WS}^+(\varphi)$ (analogously for $\mathcal{WS}^-(\mathbf{M},\mathbf{L},v,\varphi)$). The definition of $\mathcal{WS}^-$ seems to be slightly more complicated than necessary; we need it to obtain a duality of the operations over the $\mathcal{WS}^+, \mathcal{WS}^-$ defined below. Let us note that there is no game (φ,v,r') for $r' > r = 1$, so it is always the case that $1 \in \mathcal{WS}^-(\varphi)$.

Lemma 7. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be an* $\mathbf{L}$*-structure,* φ *a formula, and* v *an* $\mathbf{M}$*-valuation. Then:*

1. $\mathcal{WS}^+(\varphi)$ *is a lower set*[2]*;*
2. $\mathcal{WS}^-(\varphi)$ *is an upper set;*
3. $\mathcal{WS}^-(\varphi) \cup \mathcal{WS}^+(\varphi) = L$*;*
4. $\|\mathcal{WS}^-(\varphi) \cap \mathcal{WS}^+(\varphi)\| \leq 1$*;*
5. *For* safe $\mathbf{M}$*:* $\mathcal{WS}^+(\mathbf{M},\mathbf{L},v,\varphi) \cap \mathcal{WS}^-(\mathbf{M},\mathbf{L},v,\varphi) = \{\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v}\}$.

These claims are all rather trivial consequences of the definition of an $(\mathbf{M},\mathbf{L})$-Game (φ,v,r); the last one is a direct consequence of Theorem 4. The following theorem gives a game-theoretic characterization of the notion of a safe structure.

Theorem 8. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be an* $\mathbf{L}$*-structure. Then* $\mathbf{M}$ *is safe iff* $\mathcal{WS}^+(\mathbf{M},\mathbf{L},v,\varphi) \cap \mathcal{WS}^-(\mathbf{M},\mathbf{L},v,\varphi) \neq \emptyset$ *for each* v *and* φ*.*

Proof. One direction is the last claim of the previous lemma. The reverse one is proven by induction over the complexity of the formula. □

[2]Namely, if $r \in \mathcal{WS}^+(\varphi)$ and $r' \leq r$, then $r' \in \mathcal{WS}^+(\varphi)$; analogously for the upper set.

If $\mathcal{WS}^+(\mathbf{M}, \mathbf{L}, v, \varphi) \cap \mathcal{WS}^-(\mathbf{M}, \mathbf{L}, v, \varphi) = \{a\}$, we can call a the truth value of the formula φ (in the $\mathbf{L}$-structure $\mathbf{M}$ and $\mathbf{M}$-evaluation e). Observe that in *safe* structures this coincides with the usual terminology (by the above theorem). However, it "works well" in non-safe structures as well, as demonstrated by Example 11.

Definition 9. *Let* $\mathbf{L}$ *be an* MV*-chain. We define the set operator* $Clo : \mathcal{P}(L) \to \mathcal{P}(L)$ *in the following way:*

$$Clo(X) \quad = \quad \begin{cases} X \cup \{\sup(X)\} \cup \{\inf(X)\} & \text{if } \sup(X), \inf(X) \text{ exist}, \\ X & \text{otherwise}. \end{cases}$$

The following lemma shows how the sets $\mathcal{WS}^+, \mathcal{WS}^-$ of a compound formula can be expressed by $\mathcal{WS}^+, \mathcal{WS}^-$ of its subformulas.

Lemma 10. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *be an* $\mathbf{L}$*-structure,* φ *a formula, and* v *an* $\mathbf{M}$*-valuation. Then:*

1. $\mathcal{WS}^+(\varphi \oplus \psi) = \{r \oplus s \mid r \in \mathcal{WS}^+(\varphi) \text{ and } s \in \mathcal{WS}^+(\psi)\}$; $\mathcal{WS}^-(\varphi \oplus \psi) = \{r \oplus s \mid r \in \mathcal{WS}^-(\varphi) \text{ and } s \in \mathcal{WS}^-(\psi)\}$;
2. $\mathcal{WS}^+(\varphi \otimes \psi) = \{r \otimes s \mid r \in \mathcal{WS}^+(\varphi) \text{ and } s \in \mathcal{WS}^+(\psi)\}$; $\mathcal{WS}^-(\varphi \otimes \psi) = \{r \otimes s \mid r \in \mathcal{WS}^-(\varphi) \text{ and } s \in \mathcal{WS}^-(\psi)\}$;
3. $\mathcal{WS}^+(\varphi \vee \psi) = \mathcal{WS}^+(\varphi) \cup \mathcal{WS}^+(\psi)$; $\mathcal{WS}^-(\varphi \vee \psi) = \mathcal{WS}^-(\varphi) \cap \mathcal{WS}^-(\psi)$;
4. $\mathcal{WS}^+(\varphi \wedge \psi) = \mathcal{WS}^+(\varphi) \cap \mathcal{WS}^+(\psi)$; $\mathcal{WS}^-(\varphi \wedge \psi) = \mathcal{WS}^-(\varphi) \cup \mathcal{WS}^-(\psi)$;
5. $\mathcal{WS}^+(\neg\varphi) = \{\neg r \mid r \in \mathcal{WS}^-(\varphi)\}$; $\mathcal{WS}^-(\neg\varphi) = \{\neg r \mid r \in \mathcal{WS}^+(\varphi)\}$;
6. $\mathcal{WS}^+((\forall x)\varphi) = \bigcap_{a \in M} \mathcal{WS}^+(v[x = a], \varphi)$; $\mathcal{WS}^-((\forall x)\varphi) = Clo(\bigcup_{a \in M} \mathcal{WS}^-(v[x = a], \varphi))$;
7. $\mathcal{WS}^+((\exists x)\varphi) = Clo(\bigcup_{a \in M} \mathcal{WS}^+(v[x = a], \varphi))$; $\mathcal{WS}^-((\exists x)\varphi) = \bigcup_{a \in M} \mathcal{WS}^-(v[x = a], \varphi)$.

Proof. These claims are rather trivial consequences of the definition of an $(\mathbf{M}, \mathbf{L})$-Game (φ, v, r). □

The next example illustrates the behavior of the sets $\mathcal{WS}^+$ and $\mathcal{WS}^-$ in a non-safe structure.

Example 11. Let $\mathbf{L}$ be the subalgebra of the standard MV-algebra with the domain $[0, 1] \cap Q$. Let q be an irrational number greater than $\frac{1}{2}$ and let a_i be a

sequence of rationals descending to q. Let $\mathbf{M}$ be the $\mathbf{L}$-structure of a predicate language with one unary predicate P, where the domain of $\mathbf{M}$ is the set of natural numbers and $P_{\mathbf{M}}(i) = a_i$. Obviously $\mathbf{M}$ is not a safe $\mathbf{L}$-structure—the truth value of $\forall x P(x)$ is undefined. Observe that $\mathcal{WS}^+(\forall x P(x)) = [0, q)$. The truth value of $\varphi = (\forall x)P(x) \oplus (\forall x)P(x)$ is also undefined. However, from the previous lemma we know that, $\mathcal{WS}^+(\varphi) = [0, 1]$ and $\mathcal{WS}^-(\varphi) = \{1\}$. Thus $\mathcal{WS}^+(\varphi) \cap \mathcal{WS}^-(\varphi) = \{1\}$. Thus it seems reasonable to claim that the formula φ *holds* (its truth value is 1). This issue (evaluation games on non-safe structures) will be elaborated in future work.

6.5 Bargaining games for Łukasiewicz logic

A classical logical game can be seen as a quarrel about the 'truth', in particular about the full and indivisible 'truth'. The player who has a winning strategy is able to win the 'truth' in any play of the game and the formula is correspondingly taken to be True or False. Fuzzy logics can be seen as logics of partial 'truth'. It seems quite natural to see a logical game from the fuzzy perspective as bargaining the 'truth' between Eloise and Abelard. If the value of a formula in a game is r then we can consider r as Eloise's part of the 'truth' and the rest $1-r$ as belonging to Abelard. Obviously, each of them wants to get as much 'truth' as possible. But in general none of them has a winning strategy for the whole 'truth'. The fuzzy game defined in the previous section is not about bargaining 'truth'. It is a standard zero-sum game, since the partiality of 'truth' is hidden in one of the parameters of the game. In this section we introduce another kind of the fuzzy game which reflects the bargaining intuition more straightforwardly.

Let us remember our gambling motivation for the standard fuzzy evaluation game. Eloise bets on a value of the formula in question before the start of the game. The stakes in the rest of the game (the current values of subformulas) have just a technical auxiliary meaning, the payoff in the end of the game is given by the initial bet. From the point of view of the bargaining motivation it is more intuitive to let the players *negotiate* the value of the formula. The aim of Eloise is to push the value of the formula up, while the aim of the Abelard is the opposite one. Assuming the value of the formula is not determined from the very beginning, the only kinds of moves where the players substantially influence it are the quantifier moves. The moves for the other connectives just modify the stake in a minimal way and redistribute it between subformulas. If a game proceeds with an existential move for the formula $\varphi = (\exists x)\psi$, Eloise has to choose an object $a \in M$ and the game continues as $(\psi, v[x : a])$. As she wants to keep the value of the formula as high as possible, she shall choose a so that $\|\psi\|_{v[x:a]} = max_{a' \in M}\|\psi\|_{v[x:a']}$, which is exactly the interpretation of the existential quantifier in fuzzy logic (for witnessed models where the maxima

exist). Similarly, Abelard minimizes y in the general move for the formula $\varphi = (\forall y)\psi$ to keep the value of φ down.

If the formula in question is in prenex form or can be transformed into one (this is the case of Łukasiewicz logic), we can modify the fuzzy evaluation game in the following way: the value of the formula is not determined in the beginning of the game; it starts with the quantifier moves consisting of choices of witnesses in order to maximize (minimize) the value of the formula. Playing quantifier prefix can be considered as a "negotiation" part of the game. After the prefix part the value of the current (quantifier-free) formula is calculated (by a judge or an oracle) and the game continues as the standard fuzzy evaluation game without quantifiers moves. Let us for the sake of simplicity skip the second part and compute the "payoffs" directly. Then the game reduces to the play of the quantifier prefix. We shall call it *bargaining fuzzy game*. The term *bargaining game* is used in game theory in a different meaning (see Osborne and Rubinstein, 1994), we hope this will not lead to a clash of intuitions.

6.5.1 The rules of the bargaining fuzzy game

Assume that **L** is an MV-chain and **M** is a *witnessed* **L**-structure.

Definition 12. *Let φ be a formula in a prenex form and v an* **M***-evaluation. The rules of the bargaining* $(\mathbf{M}, \mathbf{L})^{pr}$*-Game* (φ, v) *for two players are defined as follows:*

$(\forall x)$ $((\forall x)\psi, v)$*: Abelard chooses $a \in M$, game continues as* $(\psi, v[x{:}a])$.

$(\exists x)$ $((\exists x)\psi, v)$*: Eloise chooses $a' \in M$, game continues as* $(\psi, v[x{:}a'])$.

term *φ is open formula: Eloise wins* $\|\varphi\|_{\mathbf{M},v}$ *and Abelard wins* $1 \ominus \|\varphi\|_{\mathbf{M},v}$.

In general, none of the players wins "the whole truth", so we cannot speak about winning strategies, just about the strategies guaranteeing some payoff.

6.5.2 Properties of bargaining games

Definition 13 (strategies in prenex games)**.** *We say that σ is a player's strategy of the level r (r-strategy) for the* $(\mathbf{M}, \mathbf{L})^{pr}$*-game* (φ, v) *iff σ guarantees him a win at least r in this game.*

The following lemma is a consequence of the previous two definitions.

Lemma 14. *Abelard has r-strategy iff Eloise has no r'-strategy for any $r' > 1 \ominus r$.*

As one would expect, there is a straightforward way to relate bargaining games to fuzzy evaluation games.

Lemma 15. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *a witnessed* $\mathbf{L}$*-structure, and* φ *a formula in a prenex normal form. Then Eloise has a winning strategy for the* $(\mathbf{M}, \mathbf{L})$*-game* (φ, v, r) *iff Eloise has an r-strategy for* $(\mathbf{M}, \mathbf{L})^{pr}$*-game* (φ, v).

Proof. Let $\varphi = Q(x_1, \ldots, x_n)\psi$, where Q is a string of quantifiers and ψ is a quantifier-free formula.

"$\rightarrow$" Any winning strategy σ for the game $(\mathbf{M}, \mathbf{L})$-Game (φ, v, r) starts with strategy σ_Q containing all the quantifier moves. We show that σ_Q is Eloise's r-strategy for $(\mathbf{M}, \mathbf{L})^{pr}$-game (φ, v).

The strategy σ_Q consists of a sequence of choices $x_{i1} := a_1 \ldots, x_{ik} = a_k$ such that $x_{i1}, \ldots, x_{ik}$ are among $x_1, \ldots, x_n$ and Eloise has a winning strategy for the $(\mathbf{M}, \mathbf{L})$-game (ψ, v', r), where $v' = v[x_{i1} := a_1, \ldots, x_{ik} := a_k]$. According to the Theorem 4 we get $\|\psi\|^{\mathbf{L}}_{\mathbf{M},v'} \geq r$.

"$\leftarrow$" Eloise has an r-strategy σ for the $(\mathbf{M}, \mathbf{L})^{pr}$-game (φ, v). It consists of a sequence of choices $x_{i1} := a_1 \ldots, x_{ik} = a_k$ such that $x_{i1}, \ldots, x_{ik}$ are among $x_1, \ldots, x_n$ and $\|\psi\|^{\mathbf{L}}_{\mathbf{M},v'} \geq r$, where $v' = v[x_{i1} := a_1, \ldots, x_{ik} := a_k]$. According to the Theorem 4, Eloise has a winning strategy $\sigma_{v'}$ in the $(\mathbf{M}, \mathbf{L})$-game (ψ, v', r).

Eloise's strategy σ' for $(\mathbf{M}, \mathbf{L})$-game (φ, v, r) is as follows: start by strategy σ, it yields an evaluation v'. Then follow strategy $\sigma_{v'}$. Obviously, σ is a winning strategy. □

We make use of our notions of $\mathcal{WS}^+$ and $\mathcal{WS}^-$ (introduced in Section 6.4) for a convenient reformulation of the previous lemma. On the first view "peculiar" definition of $\mathcal{WS}^-$ allows us to formulate it for Abelard's strategies in a nice dual way.

Theorem 16. *Let* $\mathbf{L}$ *be an* MV*-chain,* $\mathbf{M}$ *a witnessed* $\mathbf{L}$*-structure, and* φ *a formula in a prenex normal form. Then:*

- $\mathcal{WS}^+(\varphi) = \{r \mid$ *Eloise has an r-strategy for* $(\mathbf{M}, \mathbf{L})^{pr}$*-game* $(\varphi, v)\}$.
- $\mathcal{WS}^-(\varphi) = \{1 \ominus r \mid$ *Abelard has an r-strategy for* $(\mathbf{M}, \mathbf{L})^{pr}$*-game* $(\varphi, v)\}$.

Proof. The first claim is a trivial consequence of the previous lemma and the definition of $\mathcal{WS}^+$. To prove the second claim we write a chain of equivalent claims: Abelard has an r-strategy for $(\mathbf{M}, \mathbf{L})^{pr}$-game (φ, v) IFF Eloise has no r'-strategy for any $r' > 1 \ominus r$ IFF Eloise has no winning strategy for $(\mathbf{M}, \mathbf{L})$-game (φ, v, r') for any $r' > 1 \ominus r$ IFF Abelard has a winning strategy for $(\mathbf{M}, \mathbf{L})$-game (φ, v, r') for any $r' > 1 \ominus r$ IFF $1 \ominus r \in \mathcal{WS}^-(\varphi)$. □

Theorem 17. *Let* σ *be Eloise's strategy of the level* $\|\varphi\|_{\mathbf{M},v}$ *and* γ *be Abelard's strategy of the level* $1 \ominus \|\varphi\|_{\mathbf{M},v}$ *for the* $(\mathbf{M}, \mathbf{L})^{pr}$*-game* (φ, v)*. Then* $< \sigma, \gamma >$ *is a Nash equilibrium of the* $(\mathbf{M}, \mathbf{L})^{pr}$*-Game* (φ, v).

The proof is evident—according to the previous claims none of the players can improve his payoff by deviating from the strategies σ, γ respectively. Observe that if we formulated bargaining game for non-witnessed structures we

would lose both the previous theorem and Lemma 15. In fact, we could find a setting (formula and evaluation) such that for each player's r-strategy there is an r'-strategy for that player for some $r < r'$.

We shall illustrate the fuzzy bargaining game with a simple example.

Example 18. Let us have a formula $\varphi = \exists x \forall y P(x, y)$, such that the predicate P takes only finitely many values according to the following table:

$$P(x, y) = \begin{pmatrix} 0.2 & 0.6 & 0.1 \\ 0.6 & 0.7 & 0.2 \\ 0.4 & 0.5 & 0.6 \end{pmatrix}$$

The goal of Eloise in the bargaining game is to choose a row m to maximize the value of φ, the goal of Abelard is to choose a column n to minimize it. As Eloise moves first, she chooses m such that $min_j P(m, j) = max_i min_j P(i, j)$, then Abelard replies by n such that $P(m, n) = min_j P(m, j)$. Hence the solution of the game is a tuple $< m, n >$ such that $P(m, n) = max_i min_j P(i, j)$. In our example this holds for $m = 3, n = 1$, the value of the game is $P(3, 1) = 0.4$, hence $\mathcal{WS}^+(\varphi) = [0, 0.4]$ and $\mathcal{WS}^-(\varphi) = [0.4, 1]$.

6.6 Fuzzy games with imperfect information

Games of imperfect information have been a standard part of game theory since the very beginning (see von Neumann and Morgenstern, 1994). Their introduction in logic by Jaakko Hintikka and Gabriel Sandu (see Hintikka and Sandu, 1997; Sandu, 1993) led to the birth of (the phenomenon of) independence-friendly (IF) logics. In a classical evaluation game both players have full information about all the previous moves at each position of the game. In logical games with imperfect information, the full history of the game might not be available at some positions.

We shall concentrate here on the prototypical case in which the independence concerns quantifier moves: a player has to choose an element from the domain independently of (i.e., without knowledge of) some of the previous choices of the other player. The independence is in IF logics syntactically denoted by the slash notation. For example, in the formula $\exists x \forall y_{/x} \psi$ the choice of y is independent of that of x. One might see an introduction of imperfect information as an obvious move when creating game semantics. In fuzzy logic we can find an independent motivation, however. There are formulas which are neither true nor false in standard IF logics. It is possible to consider these formulas as having the third truth value 'Undefined' and so to move from classical bivalent logic to a three valued logic (see Sandu and Pietarinen, 2001). If our point of departure is a many valued logic, what will be the result of introducing independence into its game semantics? This section provides basic steps toward answering this question.

There are several options to introduce independence into the fuzzy games. One possibility would be to use the definition of a two-move round in the fuzzy evaluation game: we could make the choice made by Abelard independent of the distribution of the stake made by Eloise. We can also introduce the usual quantifier independence into the fuzzy evaluation game. In this section we shall examine the most interesting case (in our opinion): the independence of the quantifier moves in the bargaining game defined in the previous section. We shall give a semantics for the formulas of Łukasiewicz logic in prenex normal form, where the prefix possibly contains slashed quantifiers.

The introduction of imperfect information requires obviously a modification of the game, but as we confine ourselves to a special kind of informational independence we do not have to introduce the full machinery of standard games with imperfect information.

The game tree is of the same kind as in the bargaining fuzzy game (i.e., nodes are labelled by a subformula and a valuation). The difference is that some positions are now indistinguishable from the point of view of some of the players. Such a set of positions is called an *information set*. Indistinguishable positions correspond to the same subformula and they may differ in the valuation of the slashed variables (if the leftmost quantifier is slashed).

Definition 19. *Two positions* (φ, v) *and* (φ, v') *of the* $(\mathbf{M}, \mathbf{L})^{pr}$*-Game* (φ, v) *are in the same information set for Eloise (Abelard) iff*

- $\varphi = \exists y_{/x_1,\ldots,x_n}\psi$ *(or* $\varphi = \forall y_{/x_1,\ldots,x_n}\psi$*) for some* $n \geq 0$*;*
- v *and* v' *agree on all variables with the possible exception of those from the set* $x_1, \ldots, x_n$*.*

The rules of the bargaining game remain the same: a player has to pick an object from the domain to fix the value of the quantified variable. The difference is in the information he can use, so we have to change the definition of a strategy.

One option is to make strategy a function on sets of histories (those which end in the same information set) rather than on a single history as in the previous case. As we want to let the strategies be of the same kind in both the games of perfect and imperfect information, we shall use the second option—the requirement of the *uniformity of strategies*. The uniform strategies must agree on the histories which end in the same information set.

Definition 20 (Uniformity of strategies)**.** *Let* σ *be a strategy of the player* $I \in \{$*Abelard,Eloise*$\}$ *in the game* $(\mathbf{M}, \mathbf{L})^{pr}$*-Game* (φ, v)*. Then* σ *is uniform iff for any two histories* h_1*,* h_2 *such that their last positions are in the same information set belonging to I,* $\sigma(h_1) = \sigma(h_2)$*.*

The rules of the games with imperfect information.

Definition 21. *Let φ be a formula in prenex form containing slashed quantifiers of the form $\forall x_{/y_1,\ldots,y_m}$ ($\exists y_{/x_1,\ldots,x_n}$) such that $y_1,\ldots,y_m$ ($x_1,\ldots,x_n$) are existentially (universally) quantified variables preceding x (y) respectively. Let v be an* **M***-evaluation. The rules of the bargaining $(\mathbf{M},\mathbf{L})^{pr}$-Game (φ, v) with imperfect information are as follows:*

($\forall x/$) ($\forall x_{/y_1,\ldots,y_m}\psi$)*: Abelard chooses $a \in M$, game continues as $(\psi, v[x{:}a])$.*

($\exists y/$) ($\exists y_{/x_1,\ldots,x_n}\psi$)*: Eloise chooses $a \in M$, game continues as $(\psi, v[x{:}a])$.*

We shall introduce the notion of a uniform *r*-strategy.

Definition 22 (Uniform strategies in fuzzy bargaining games)**.** *We say that σ is a player's uniform strategy of the level r (r-strategy) for the $(\mathbf{M},\mathbf{L})^{pr}$-game (φ, v) iff σ is uniform and guarantees him a win of at least r in this game.*

It is evident that the fuzzy bargaining game defined in the previous section is a special case of that with imperfect information—all the information sets are singletons and hence any *r*-strategy is uniform. Thus without loss of generality we omit the word "uniform".

We adapt the notions of $\mathcal{WS}^+$ and $\mathcal{WS}^-$ and define:

- $\mathcal{WS}^+(\varphi) = \{r \mid$ Eloise has an *r*-strategy for $(\mathbf{M},\mathbf{L})^{pr}$-game $(\varphi, v)\}$.
- $\mathcal{WS}^-(\varphi) = \{1 \ominus r \mid$ Abelard has an *r*-strategy for $(\mathbf{M},\mathbf{L})^{pr}$-game $(\varphi, v)\}$.

Theorem 16 tells us that in the special case of formulas without slashed quantifiers the recently defined notions of $\mathcal{WS}^+$ and $\mathcal{WS}^-$ correspond to the ones from the bargaining game. Obviously, $\mathcal{WS}^+(\varphi)$ is a lower set and $\mathcal{WS}^-(\varphi)$ is an upper set.

Lemma 23. *Let φ be a formula. Then $\|\mathcal{WS}^-(\varphi) \cap \mathcal{WS}^+(\varphi)\| \leq 1$.*

Let us consider again the formula $\varphi = \exists x \forall y P(x,y)$ from the previous example and explore how the sets $\mathcal{WS}^+(\varphi)$ and $\mathcal{WS}^-(\varphi)$ change when the second quantifier becomes independent of the first one, i.e., we shall consider the formula $\varphi' = \exists x \forall y_{/x} P(x,y)$. It is easy to see that Eloise keeps all her *r*-strategies in the sense that $\mathcal{WS}^+(\varphi') = [0, 0.4] = \mathcal{WS}^+(\varphi)$ while Abelard loses some of them as $\mathcal{WS}^-(\varphi) \subset \mathcal{WS}^-(\varphi') = [0.6, 1]$. Neither player has a *r*-strategy for $r \in (0.4, 0.6)$. This example shows that in fuzzy bargaining games with imperfect information $\mathcal{WS}^+(\varphi)$ and $\mathcal{WS}^-(\varphi)$ do not have in general to cover the whole set of truth values.

The introduction of independence leads, as in the case of the classical IF logics, to the existence of formulas without a definite truth value. We can say that in the case of IF fuzzy logic we are more specific about the indefiniteness

of the truth value. The classical "Indefinite" may be interpreted as "anywhere between Truth and Falsity" (or between 0 and 1) in the many-valued setting. In the case of fuzzy logics, the zone of indefiniteness does not extend to the whole unit interval, but corresponds in general to a proper subinterval of $[0, 1]$.

Definition 24. *If there is $a \in L$ such that $\mathcal{WS}^+(\varphi) \cap \mathcal{WS}^-(\varphi) = \{a\}$ we say that a is the truth value of φ.*

Observe that for formulas without slashed quantifiers our notion of a truth value coincides with the Tarskian one (compare with the analogous situation in the non-safe structures—see Section 6.4). In cases where there is no such element, the value of the formula cannot be characterized by a single element, but it is given by sets of values $\mathcal{WS}^+, \mathcal{WS}^-$.

6.7 Conclusion

The aim of this article was to introduce evaluation games for fuzzy logics. We concentrated on the Łukasiewicz logic and provided two kinds of game-theoretic semantics. It is possible to generalize this semantics for a general fuzzy logic (in particular for product and Gödel logic), however there is no space to do so in this article.

From the philosophical point of view our game semantics provides an alternative mode of presentation for fuzzy connectives. As Łukasiewicz logic can be seen as a substructural logic (it does not have the contraction rule), we can consider our definition as an example of a model-theoretic game semantics for a substructural logic (proof-theoretic game semantics is provided by Blass (1992)).

We have shown that, although the generalization of classical evaluation games to the realm of fuzzy logics is rather straightforward, it gives us new tools for finer analysis of various semantical issues. In particular, our interpretation of formulas as sets $\mathcal{WS}^+$ allows us to interpret formulas lacking a truth value in non-safe models under the Tarskian semantics. We showed that sets $\mathcal{WS}^+$ behave compositionally, so they can serve as basis for a new semantics over non-safe structures.

Finally, the notion of informational independence introduced in the last section leaves many open questions. We propose to use sets $\mathcal{WS}^+$ to interpret IF fuzzy formulas as well (it suggests the idea that formulas over non-safe models and IF fuzzy formulas are of a similar nature). We believe to obtain a full semantics for at least some (kind of) IF fuzzy logic this way.

Appendix: Łukasiewicz predicate logic

In the Appendix we survey the basic properties of Łukasiewicz predicate logic. Unlike in the propositional logic there is no standard completeness in the predicate case. So first we need to introduce the general algebraic semantics.

Definition 25. *An* MV-algebra *is a structure* $\mathbf{L} = (L, \oplus, \neg, 0)$ *where:*

1. $(L, \oplus, 0)$ *is a commutative monoid,*
2. $\neg\neg x = x$*,*
3. $x \oplus \neg 0 = \neg 0$*,*
4. $\neg(\neg x \oplus y) \oplus y = \neg(\neg y \oplus x) \oplus x$*.*

It can be shown that in each MV-algebra $\mathbf{L} = (L, \oplus, \neg, 0)$ the reduct $(L, \vee, \wedge, 0, 1)$ is a bounded lattice.

Definition 26. *Let* $\mathbf{L} = (L, \oplus, \neg, 0)$ *be an* MV*-algebra. We say that* $\mathbf{L}$ *is linearly ordered* MV*-algebra* (MV-chain) *if* $(L, \vee, \wedge, 0, 1)$ *is linearly ordered lattice.*

An MV*-chain* $\mathbf{L}$ *is* atomic *if it contains an element* $a \neq 0$ *(called atom of* $\mathbf{L}$*) such that for each* $b \neq 0$ *is* $a \leq b$*.*

We define the basic semantical notions ($\mathbf{L}$-evaluation, $\mathbf{L}$-model, and $\mathbf{L}$-tautology of MV-algebra $\mathbf{L}$) as usual:

Definition 27. *Let* $\mathbf{L} = (L, \oplus, \neg, 0)$ *be an* MV*-algebra. We say that a mapping e from the set of formulas to the set L is an* $\mathbf{L}$*-evaluation if:*

- $e(\varphi \oplus \psi) = e(\varphi) \oplus e(\psi)$
- $e(\neg\varphi \oplus \psi) = \neg e(\varphi)$
- $e(0) = 0$

$\mathbf{L}$*-evaluation is* $\mathbf{L}$*-model of theory* T *if* $e(\varphi) = 1$ *for each* $\varphi \in T$*. Formula* φ *is* $\mathbf{L}$*-tautology if* $e(\varphi) = 1$ *for each* $\mathbf{L}$*-evaluation e.*

We use the symbol $\models_{\mathbf{L}}$ for the semantical consequence over given MV-algebra $\mathbf{L}$ ($T \models_{\mathbf{L}} \varphi$ iff for each $\mathbf{L}$-model e of T we have $e(\varphi) = 1$). We can prove the following (strong) completeness theorem:

Theorem 28 (Strong completeness theorem)**.** *Let* T *be a theory and* φ *a formula. Then the following are equivalent:*

- $T \vdash \varphi$*.*
- $T \models_{\mathbf{B}} \varphi$ *for each* MV*-algebra* $\mathbf{B}$*.*
- $T \models_{\mathbf{B}} \varphi$ *for each* MV*-chain* $\mathbf{B}$*.*

For finite theories we could add one more equivalent condition:

- $T \models_{[0,1]_S} \varphi$*.*

We assume that the reader is familiar with the syntax and semantics of classical predicate logic. Here we refresh the basic notions of Łukasiewicz predicate logic. As mentioned above, we are forced to work with more general algebras of truth values than the standard interval $[0, 1]_S$, in particular with MV-chains. As shown in Hájek (1998) (originally in a slightly weaker form in Belluce and Chang (1963), see also a recent survey (Hájek and Cintula, 2006) this logic is strongly complete with respect to the class of all MV-chains (like in the propositional case).

For each MV-chain $\mathbf{L}$, an $\mathbf{L}$-structure for a predicate language $\mathbf{\Gamma}$ is $\mathbf{M} = (M, (P_{\mathbf{M}})_{P \in \mathbf{\Gamma}}, (f_{\mathbf{M}})_{f \in \mathbf{\Gamma}})$ where $M \neq \emptyset$ is the domain of the model, for each predicate P of arity n, $P_{\mathbf{M}}$ is an n-ary $\mathbf{L}$-fuzzy relation on M (a mapping $M^n \to \mathbf{L}$), and for each function f, $f_{\mathbf{M}}$ is a mapping $M^n \to M$. Having this, one defines for each formula φ (of the given language), the *truth value*

$\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v}$ of φ in $\mathbf{M}$ determined by the MV-chain $\mathbf{L}$ and the $\mathbf{M}$-evaluation v of free variables of φ in M in the usual (Tarskian) way.

Definition 29. *Let* $\mathbf{\Gamma}$ *be a predicate language,* $\mathbf{L}$ *an* MV*-algebra,* $\mathbf{M}$ *an* $\mathbf{L}$*-structure for* $\mathbf{\Gamma}$*,* v *an* $\mathbf{M}$*-evaluation. The* value of the term *is defined as:* $\|x\|_{\mathbf{M},v} = v(x)$ *and* $\|f(t_1,\dots,d_n)\|_{\mathbf{M},v} = f_{\mathbf{M}}(\|t_1\|_{\mathbf{M},v},\dots,\|t_n\|_{\mathbf{M},v})$*. A* truth value of the formula φ *in* $\mathbf{M}$ *for an evaluation* v *is defined*[3]*:*

$$\begin{aligned}
\|P(t_1,\ t_2,\ \dots,\ t_n)\|^{\mathbf{L}}_{\mathbf{M},v} &= P_{\mathbf{M}}(\|t_1\|_{\mathbf{M},v},\ \|t_2\|_{\mathbf{M},v},\ \dots,\ \|t_n\|_{\mathbf{M},v})\,,\\
\|\varphi \oplus \psi\|^{\mathbf{L}}_{\mathbf{M},v} &= \|\varphi\|^{\mathbf{L}}_{\mathbf{M},v} \oplus \|\psi\|^{\mathbf{L}}_{\mathbf{M},v}\,,\\
\|\neg\varphi\|^{\mathbf{L}}_{\mathbf{M},v} &= \neg\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v}\,,\\
\|0\|_{\mathbf{M},v} &= 0\,,\\
\|(\forall x)\varphi\|_{\mathbf{M},v} &= \inf\{\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v'} \mid v' \equiv_x v\}\,.\\
\|(\exists x)\varphi\|_{\mathbf{M},v} &= \sup\{\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v'} \mid v' \equiv_x v\}\,.
\end{aligned}$$

If infimum (supremum) does not exist, we take its value as undefined.

As we can see, in the general case the truth assignment is a *partial* function. To overcome this difficulty we define two classes of models:

Definition 30. *Let* $\mathbf{\Gamma}$ *be a predicate language,* $\mathbf{L}$ *an* MV*-chain,* $\mathbf{M}$ *an* $\mathbf{L}$*-structure for* $\mathbf{\Gamma}$*. We say that* $\mathbf{M}$ *is:*

- *A* safe $\mathbf{L}$*-structure, if* $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v}$ *is defined for each* φ *and* v*.*
- *A* witnessed $\mathbf{L}$*-structure, if* $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v}$ *is defined for each* φ *and* v *if we replace* sup *and* inf *in Definition 29 by* max *and* min*.*

Clearly, there are non-safe $\mathbf{L}$-structures and each witnessed $\mathbf{L}$-structure is safe, but not vice versa. We can define witnessed structures more straightforwardly: $\mathbf{M}$ is a *witnessed* $\mathbf{L}$-structure iff for each formula φ and for each evaluation v there are $a, b \in M$ such that $\|(\exists x)\varphi\|^{\mathbf{L}}_{\mathbf{M},v} = \|\varphi\|^{\mathbf{L}}_{\mathbf{M},v[x:=a]}$ and $\|(\forall x)\varphi\|^{\mathbf{L}}_{\mathbf{M},v} = \|\varphi\|^{\mathbf{L}}_{\mathbf{M},v[x:=b]}$. By $(\mathbf{M},\mathbf{L}) \models \varphi$ we denote the fact $\|\varphi\|^{\mathbf{L}}_{\mathbf{M},v} = 1$ for each $\mathbf{M}$-evaluation v. Observe that in atomic MV-chain each element has the successor and predecessor and so we get:

Lemma 31. *Let* $\mathbf{L}$ *be an atomic* MV*-chain. Then each* $\mathbf{L}$*-structure is witnessed.*

Definition 32. *The predicate Łukasiewicz logic has the axioms:*

(P) the axioms resulting from the axioms of Łukasiewicz logic by the substitution of the propositional variables by the formulas of $\mathbf{\Gamma}$*,*

(∀1) $(\forall x)\varphi(x) \to \varphi(t)$*, where* t *is substitutable*[4] *for* x *in* φ*,*

(∀2) $(\forall x)(\chi \to \varphi) \to (\chi \to (\forall x)\varphi)$*, where* x *is not free in* χ*,*

The deduction rules are modus ponens and generalization: from φ *infer* $(\forall x)\varphi$*.*

It is well known, that Łukasiewicz logic is complete with respect to the *safe* $\mathbf{L}$-structures over all MV-chains $\mathbf{L}$.

Theorem 33 (Completeness Theorem)**.** *Let* $\mathbf{\Gamma}$ *be a predicate language and* φ *a formula. Then the following are equivalent:*

- $\vdash \varphi$*.*

[3]Recall we use the same symbols for both connectives and corresponding operations. By $v \equiv_x v'$ we mean that $v(y) = v'(y)$ for each object variable y different from x.

[4]The notions of substitutability, free and bounded occurrence of a variable are defined as usual.

- $(\mathbf{M}, \mathbf{L}) \models \varphi$ *for each* MV-*chain* $\mathbf{L}$ *and each* safe $\mathbf{L}$-*structure* $\mathbf{M}$.
- $(\mathbf{M}, \mathbf{L}) \models \varphi$ *for each* MV-*chain* $\mathbf{L}$ *and each* witnessed $\mathbf{L}$-*structure* $\mathbf{M}$.

As in the propositional case we could formulate the strong completeness theorem. However, unlike in propositional case there is no standard completeness theorem. In fact, the set of predicate tautologies of the standard MV-algebra is a Π_2-complete set.

References

Belluce, L. P. and Chang, C. C. (1963). A weak completeness theorem on infinite valued predicate logic. *Journal of Symbolic Logic*, 28(1):43–50.

Blass, A. (1992). A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56(1–3):183–220.

Fermüller, C. G. (2003). Parallel dialogue games and hypersequents for intermediate logics. In Mayer, M. C. and Pirri, F. editors, *Proceedings of TABLEAUX Conference 2003*, pages 48–64, Rome.

Hájek, P. (1998). *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic*. Kluwer, Dordrecht.

Hájek, P. and Cintula, P. (2006). Triangular norm predicate fuzzy logics. To appear in Proceedings of Linz Seminar 2005.

Hintikka, J. and Sandu, G. (1989). Informational independence as a semantical phenomenon. In Fenstad, J. E., Frolov, I. T., and Hilpinen, R., editors, *Proceedings of LMPS*, volume 8, pages 571–589. North-Holland, Amsterdam.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier Science B.V., Oxford, Shannon, Tokyo, MIT Press, Cambridge, MA.

Łukasiewicz, J. (1920). O logice trojwartosciowej. *Ruch filozoficzny*, 5:170–171. (On Three-Valued Logic).

Mundici, D. (1993). Ulam games, Łukasiewicz logic and AFC*-algebras. *Fundamenta Informaticae*, 18:151–161.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press, Cambridge, MA.

Sandu, G. (1993). On the logic of informational independence and its applications. *Journal of Philosophical Logic*, 22:29–60.

Sandu, G. and Pietarinen, A.V. (2001). Partiality and games: propositional logic. *Logic Journal of the Interest Group of Pure and Applied Logic*, 9:107–127.

von Neumann, J. and Morgenstern, O. (1994). *Theory of Games and Economic Behavior*. Wiley, New York.

Chapter 7

GAMES, QUANTIFICATION AND DISCOURSE STRUCTURE

Robin Clark*
Department of Linguistics, University of Pennsylvania
rclark@babel.ling.upenn.edu

Abstract Quantifiers in natural language contribute both to the truth conditions of a sentence and to the discourse in which the sentence occurs. While a great deal of attention has been paid to truth conditions, the contributions of quantifiers to the discourse have been little studied. This paper seeks to rectify this by developing a set of game rules that account both for the truth conditional and the discourse contributions of quantified expressions.

7.1 Overview

In this paper, I would like to make some proposals on the treatment of quantifiers and their consequences for discourse in Game Theoretic Semantics.[1] For present purposes, I will mean by *quantifier* any noun phrase which contains both a determiner and a head noun along any modifiers like adjectives and relative clauses. The examples in (1) illustrate the sort of phrases I will be interested in treating[2]:

(1) a. *Aristoteleans*: every dean, all deans, some faculty member, not all students, no provost.
b. *Cardinals and bounding determiners*: at least five administrators, at most four department chairs, between three and ten trustees.

*The author wishes to acknowledge the support of the NIH, grant NS44266. Portions of this material were presented at the 2004 Prague Colloquium; thanks to all for the many helpful comments.

[1] For general treatments of GTS, see Hintikka and Sandu (1997), Hintikka and Kulas (1985), and Hintikka (1996), and the references cited there. For a treatment of some generalized quantifiers within GTS see Pietarinen (2001, 2007).

[2] I will put aside definite descriptions like *the dean* and "polyadic quantifiers" like *each. . .a different. . .* as in "each dean read a different comic book." The former case requires a more detailed discussion of the discourse model (see Clark (2005) for a game-theoretic treatment) and the latter case is too complex to treat here; but see van Benthem (1989), Keenan (1992) and Pietarinen (2007).

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

c. *Majorities*: most monkeys, more than half of the department chairs, more deans than faculty members, fewer provosts than trustees.

The expressions in (1) have long been studied by both linguists and philosophers and a sophisticated and very useful theory of them already exists.[3]

Why bother to recast the project in terms of games? One good answer is that it is always interesting and useful to rework a well-understood theory in a different way. I think, however, that there is a more interesting and telling reason to investigate the game-theoretic properties of quantifiers in natural language. To motivate things, let us consider a small text like the one in (2):

(2) Exactly three students took an exam. One passed it effortlessly. She had clearly studied for it. The others struggled with it and barely passed. They felt relief mixed with shame.

The first sentence in the text in (2) contains two quantifiers: *exactly three students* and *an exam*. Clearly, the sentence is true in some models and false in others. But the contribution of the quantifiers extends beyond the first sentence, as the second sentence illustrates. The indefinite *one* in the second sentence clearly depends on the discourse effects of the quantifier *exactly three students* from the first sentence; equally, the pronoun *it* in the second sentence depends upon the quantifier *an exam* in the first. Continuing to the third sentence, *she* must refer (in the absence of other compelling discourse information) to the student picked out by *one* in the second sentence, and so is ultimately contingent on the first sentence. Consider, next, the contribution of the phrase *the others* in the second to last sentence. This phrase is understood as referring to the remaining two students. Its interpretation is contingent upon the interaction between the quantifier *exactly three students* and the discourse anaphor *one*. Having interpreted *the others* correctly, these two students can become the antecedent of the pronoun *they* in the last sentence.

Clearly, quantifiers make contributions to our understanding of and inferences from texts, contributions that extend far beyond the level of the sentence. Generalized quantifier theory, with its emphasis on truth conditions and the functions that interpret quantifiers, has not been able to clearly address the connections between quantifiers and their effects on the discourse.[4] I will argue, though, that Game Theoretic Semantics provides a natural framework for investigating these relationships. I will be particularly concerned with how quantifier games can establish new discourse entities. I will not, however,

[3]Generalized quantifiers, first proposed in Mostowski (1957), have played a central role in the development of semantic theory. See, among many others, Barwise and Cooper (1981), van Benthem (1986), Keenan and Stavi (1986), and Keenan and Westerståhl (1997).

[4]Indeed, Dynamic Semantics (Groenendijk and Stokhof, 1991; van den Berg, 1996; Beaver, 2001) and Discourse Representation Theory (Kamp and Reyle, 1993) have attempted to address these connections.

discuss reference tracking, the problem of following discourse entities through a discourse. I take this to be a problem of the management of resources in the discourse model, a problem that is amenable to a game treatment. For games played on this level, the reader may consult Clark (2004, 2005).

Throughout this paper, I will make the standard assumptions about Game Theoretic Semantics as applied to natural language. We will suppose that there are two players, Eloïse and Abelard, the initial verifier and the initial falsifier, respectively. They are playing a zero-sum game on a sentence S relative to a model M. Eloïse wins if the sentence is verified relative to M while Abelard wins if the sentence is counter-exemplified. See Hintikka and Sandu (1996) for discussion and references.

7.2 Aristoteleans

The Aristotelean quantifiers are familiar from first-order logic. These include any quantifier synonymous with those in the familiar Aristotelean forms. These quantifiers have been well-studied in GTS and I will restrict my attention to their effects on discourse anaphora. For present purposes, I will restrict my attention to *every/all* and *some/a*, the other Aristoteleans being simple to define on this basis.

I will follow the practice in GTS of allowing players to add entities and sets of entities to a special set, the choice set I_S. This set works as a discourse model, a database of entities and sets of entities that have been invoked by the discourse to date. When a pronoun, definite description or other anaphoric element is encountered in the course of play, one player or the other must select an entity or set from I_S to serve as the target referent of the anaphoric element. For example, in a sequence of games like:

(3) A boy was following a man. The man did not notice.

The verifier will choose elements from the model to act as witnesses for the noun phrases in the first sentence. These elements will be placed in the choice set at the end of a play of the first game. The second game begins with a definite description, *the man*. The verifier is forced to make her choice of referent from I_S; hence, *the man* in the second sentence is understood to refer to an entity already invoked by the discourse. Compare (3) with:

(4) a. Every student thinks he's smart. #He has enormous self-confidence.
b. Every student thinks he's smart. They have enormous egos.

The sequence in (4)a is decidedly peculiar, while the sequence in (4)b is more acceptable.[5] To account for the differences between (3), (4)a and (4)b, we must, first, partition the choice set into two parts. One member of the partition,

[5] See Clark (2004, 2005) for a game-theoretic treatment of such texts.

I_{current}, will contain the entities invoked by the current sentence. Certain types of anaphora, an example might be reflexives like *himself* and *herself* constrain the player to choose within I_{current}. The other, $I_{\text{discourse}}$, will contain entities and sets that have been invoked by previous games in the discourse. We can think of this partition as the discourse model proper. Second, we will impose the following constraint on the passing of discourse entities from I_{current} to $I_{\text{discourse}}$:

(5) *Choice Preservation*
An entity is passed from I_{current} to $I_{\text{discourse}}$ just in case it was selected by Eloïse. Otherwise, the set X from which the entity was selected is placed in $I_{\text{discourse}}$.

To demonstrate the system, let us turn to examples. Denoting a game played on a sentence S relative to a model, M, as $G(S; M)$, then we have the following rule for *some*:

(R.some)
If the game $G(S; M)$ has reached an expression of the form:

$$Z - \text{some } X \text{ who } Y - W.$$

Then the verifier may choose an individual from the appropriate domain, say b. The game is then continued as $G(Z - b - W, b \text{ is an } X \text{ and } bY; M)$. The individual b is added to the set I_{current}.

Given the constraint in (5), if Eloïse is playing verifier when she chooses the witness b, then b will survive in the discourse model and be added to the set $I_{\text{discourse}}$. Thus, the following text is acceptable:

(6) Mary saw some student. He jumped out from behind the door.

In (6), Eloïse, playing the initial verifier, selects a witness for *some student* from the model in accord with the rule (R.some). After the first game, the players move to the second sentence in (6). In order to interpret the pronoun, *he*, Eloïse must pick an entity from $I_{\text{discourse}}$. Since her choice of student survives in $I_{\text{discourse}}$, she can select it as the target referent of *he*.

If Abelard, the initial falsifier, is playing verifier, as would be the case under negation, then his choice is forgotten, as (7) shows[6]:

(7) Mary didn't see some student. {They were/#He was} hiding in the corridor.

In (7), negation forces the verifier and the falsifier to exchange roles. Therefore, at the point where (R.some) must be played on *some student*, Abelard will be playing the verifier. His choice of student is dropped at the end of the game, although the set of students is added to the discourse model.

[6] I will prefix pragmatically odd choices with a '#.'

The rule (R.some) and Choice Preservation allow us to account for the establishment of discourse entities in this simple case. Let us contrast (R.some) with another Aristotelean, *every*:

> **(R.every)**
> If the game $G(S; M)$ has reached an expression of the form:
>
> $$Z - \text{every } X \text{ who } Y - W.$$
>
> Then the falsifier may choose an individual from the appropriate domain, say b. The game is then continued as $G(Z - b - W, b \text{ is an } X \text{ and } bY; M)$. The individual b is added to the choice set I_{current}.

Since the falsifier is playing, there is no question of the particular choice of witness selected under (R.every) surviving into the next game, although the set that contains witness does survive, according to Choice Preservation. We expect to see the following pattern:

(8) a. Every student thinks he's treated unfairly.
b. Every student passed the exam. #She studied very hard.
c. Every student passed the exam. They studied very hard.
d. Every student wrote an essay. One spelled most of the words correctly. He must have had a dictionary.

In (8)a, the falsifier selects an entity as a counterexample to the sentence. This entity is placed into I_{current} and can act as an antecedent for any anaphor that occurs within the game. Thus, the pronoun *he* in the embedded clause can denote the falsifier's choice. At the end of the game, however, the falsifier's particular choice is deleted and the set that falsifier chose from is placed in $I_{\text{discourse}}$. Thus, there is no singular referent for the pronoun *she* in (8)b and the text is peculiar, all else being equal.

We can compare the peculiarity of (8)b with the unremarkable acceptability of (8)c. Although the falsifier's choice of witness is dropped, the set from which he chose is placed in $I_{\text{discourse}}$ and can serve as the target for a plural pronoun. Consider, finally, the slightly longer text in (8)d. In the first sentence, the falsifier selects a witness and the set he chose from is evoked; by Choice Preservation, *students* is added to $I_{\text{discourse}}$. The rule for interpreting *one* is approximately:

> **(R.one)**
> When a semantical game has reached a sentence of the form:
>
> $$X - \text{one} - Y$$
>
> an individual, say b, is selected by the verifier from a set in $I_{\text{discourse}}$. The game is continued with respect to:
>
> $$X - b - Y.$$
>
> The entity b is then added to I_{current}.

In accordance with (R.one), the verifier may find a set in $I_{\text{discourse}}$ and pick an element from it to serve as the witness for *one*. Having done so, she establishes a particular discourse entity—one of the students—who survives in $I_{\text{discourse}}$ and can then serve as the antecedent for the pronoun in the third sentence.

7.3 Cardinals and bounding determiners

We turn now to the interesting cases of cardinals and bounding determiners, as exemplified in:

(9) a. At least three students passed the exam.
b. At least one (= some) dean drank eau de vie.
c. At most ten graduate students wrote papers.
d. Between three and seven trustees take viagra.
e. Exactly five deans read at the sixth grade level.

These quantifiers involve explicit numeric quantities. In (9)a and (9)b the quantifiers set a lower bound on the number of individuals with the property named in the predicate. In (9)c an upper bound is placed on the number of individuals and in (9)d and (9)e upper and lower limits are placed on the number of individuals. Thus, one might take *exactly five* to mean "more than four but less than three." Notice that we take *at least one* to be equivalent semantically to *some*, although their pragmatic effects may differ. I will, again, restrict my attention to some simple cases, the others being easy to define on their basis.

The game rules that follow differ in form from the rules for Aristoteleans presented in Section 7.2. The games for the Aristoteleans all involve the choice of a witness by one player or the other. I will propose that these quantifiers, as well as the quantifiers that follow, involve two moves. In one move, a player chooses a set of entities and, in the next move, his or her opponent selects a witness from that set.

Consider a simple cardinal determiner like *at least n*, where *n* is a positive integer. We can simulate this kind of a quantifier by allowing the verifier to select a set of entities from the model, each of which could potentially witness the relevant property. The falsifier is then allowed to select an individual from this set as a counterexample. If he is unable to do so, then the sentence must hold in the model and the verifier wins:

(R.at least *n*)
If the game $G(S;M)$ has reached an expression of the form:

$$Z - \text{at least } n\ X \text{ who } Y - W$$

then the verifier may choose a set of entities from the domain M, call it ver(M), such that $|\text{ver}(M)| \geq n$. The falsifier then selects an entity $d \in \text{ver}(M)$. Play continues on $Z - d - W, d$ is an X and $d - Y$. Both d and the contents of ver(M) are placed in I_{current}.

Notice that both the verifier's choice of the set, ver(M), and the falsifier's choice from ver(M) are placed in I_{current}, although only ver(M) will survive in the discourse model. This means that the falsifier's choice of counterexample and the verifier's choice of the set should be available as targets for anaphora within the current game. To motivate this consider the following:

(10) a. #At least three students think he's smart.
(where *he* is one of the students.)
b. At least three students think they're smart.

At first view, it would seem that a singular anaphor is ruled out in (10)a. That is, we cannot utter (10)a intending to mean that each of the set of at least three students believes of himself or herself "I am smart." I submit, however, that this is a fact about the morphosyntax of coreference; the problem is that the pronoun does not agree in number with the antecedent noun phrase, so the two cannot share reference at any level. Compare this with (10)b, which is at least three ways ambiguous. On one reading, the students have a belief about some set of individuals, namely that whose individuals are smart. We need not concern ourselves with this reading. The other two readings involve whether the students believe of the whole set of three or more students (that is, the witness set) that they all are smart or whether each member of the set believes "I am smart." In the former case, the set ver(M) is the target of the anaphor and in the latter case the falsifier's choice of individual, d, is the target of the anaphor.

When the contents of I_{current} are placed in $I_{\text{discourse}}$, the falsifier's choice of witness is, of course, deleted as required by Choice Preservation. We can immediately account for the following range of texts:

(11) a. At least five deans smoked crack. They passed out.
b. At least five deans drank Mad Dog. #He passed out.
c. At least five deans dropped acid. One jumped out the window.

Example (11)a is acceptable because the plural pronoun *they* can denote the verifier's witness set, the set of deans that smoked crack. Example (11)b is odd because, all else being equal, there is no entity in $I_{\text{discourse}}$ for the singular definite pronoun to denote. Finally, (11)c is acceptable because the verifier can pick a single element out of the witness set, now transfered to $I_{\text{discourse}}$. The explanations for each of the small texts in (11) is the same as those given for texts above involving the Aristoteleans.

Let us turn, now, to examples involving *at most n* as in:

(12) At most three politicians smoke crack.

Following Keenan and Stavi (1986), we might try to exploit the boolean structure of the algebra in which natural language determiners take their denotations and treat *at most n*(P, Q) as the boolean complement of *at least n* + 1$(P, \overline{Q})$.

That is, the falsifier and the verifier would exchange roles and play on *at least* $n + 1$ when they encounter a sentence containing *at most n*. But consider the discourse effects of a sentence containing *at most n*:

(13) At most five faculty members considered resorting to vegetarianism. They changed their minds when they realized how much work it would be.

The first sentence is true if three faculty members considered resorting to vegetarianism. Suppose that is the case. The pronoun in the next sentence refers to just those three faculty members who considered resorting to vegetarianism and to no others. Our method is to write rules that require the players to select sets that will eventually serve as potential discourse entities; the problem with exchanging roles and playing on *at least* $n+1$ is that it fails to create the needed discourse entities. We must, therefore, reject this approach.

The following rule, however, will do the trick:

(R.at most *n*)
If the game $G(S\,;M)$ has reached an expression of the form:

$$Z - \text{at most } n\ X \text{ who } Y - W$$

The verifier chooses a set of entities from the domain M, call it ver(M), such that the cardinality of ver(M) is less than or equal to n. The falsifier chooses a disjoint set of entities from M, call it fal(M), such that $|\text{ver}(M) \cup \text{fal}(M)| > n$. The game then continues on:

$Z-$every ver(M) $-W$, $Z-$no fal(M)$-W$, every ver(M) is an X who Y, every fal(M) is an X who Y.

The set ver(M) is placed in I_{current}.

The rule (R.at most *n*) is based on the idea that the verifier must choose a maximal set of cardinality bounded by *n*. If she has a winning strategy, then she must pick out every object so described and the falsifier should be unable to select an object matching that description. Since Eloïse in her role as the verifier never selects a single entity—play is carried by selecting sets and then playing on *every*, where the falsifier chooses, and *no*, where Eloïse cannot be playing as initial verifier—we do not expect definite singular discourse anaphora to be licensed by *at most n*, although ver(M) will be in the choice set and available for plural definites and indefinite anaphora. Thus, *at most n* should behave like *at least m*, which it does:

(14) a. At most five trustees know how to play *Candyland*. #He studied it at Harvard Business School.
b. At most five trustees drank Old Crow. They were trying to save money.
c. At most five trustees performed on the kazoo. One did a passable interpretation of *Die Walkyrie*.

In (14)b, the pronoun *they* refers to the five or fewer trustees who drank Old Crow; that is, the pronoun refers to the set selected by the verifier in the previous game. The game rule (R.at most *n*), when contrasted with the treatment of *at most n* as the complement of *at least n* + 1, brings out the strategic nature of interpretation.

I will put aside bounding quantifiers as below (but see Clark, 2004); for the present I will merely note some appropriate texts and leave the definition of the game rules as a puzzle for the reader:

(15) a. Between three and seven department chairs exchanged flowers. #He sneezed because he was allergic.
b. Between three and seven department chairs exchanged flowers. They decorated their hats with them.
c. Between three and seven department chairs exchanged flowers. One led a ceremonial procession down the Alps.

7.4 Majority determiners

By *majority determiners* I mean determiners like *most*, *more/less than half of the* and *n*-ary determiners like *more . . . than . . .*, as illustrated in (15):

(16) a. Most faculty eat grubs in the winter.
b. More than half of the trustees dine on Andalucian dogs.
c. More deans than faculty resort to prostitution.

These determiners, being higher-order, are of greater complexity than those we have considered up to now. *most P's are Q's* is true when the cardinality of the set of things that are both *P* and *Q* is greater than that of the set of things that are *P* but not *Q*. It may not be immediately obvious how to construct a simple game rule, based on choice, that will yield the correct result; that is, where verifier has a winning strategy just in case a majority of *P*'s have the property *Q*.

In addition, the cardinalities involved might be infinite:

(17) Most integers are not divisible by five.

Although Abelard should win on (17), it is far from obvious how to encode the meaning of (17) in a finite game, if such a thing is even possible. For the moment, I will restrict my attention to games involving majorities over finite sets.

We might try the following game rule for *most*:

(R.most)
If the game $G(S : M)$ has reached an expression of the form:

$$Z - \text{most } CN \text{ who } P_1 - W$$

where CN is a common noun and P_1 is a predicate, then the verifier picks a set of objects, call it ver(M), of cardinality at least:

$$\frac{|CN|}{2} + 1$$

The falsifier may choose an individual $d \in$ ver(M) and the game continues as:

$$G(Z - d - W, d \text{ is a } CN \text{ and } d\ P_1; M)$$

The set ver(M) is then added to the choice set I_S.

The game rule (R.most) requires that the verifier select a set whose cardinality is greater than half that of the set denoted by CN. The falsifier may then select an element of that set to test the sentence on. If the falsifier cannot select a counterexample from the set, then it must be that a majority of the elements denoted by CN have the requisite property and the verifier wins. Notice that the difference between (R.most) and the game rules for the cardinal determiners resides in the requirement that ver(M) be of a particular size.

Finally, the rule requires that the set ver(M) be placed in the choice set. The discourse effect of (R.most) should be similar to those of the cardinal determiners. That is, singular pronouns will not match, but plurals and indefinites will:

(18) a. Most deans practice fortune-telling. He is a reader of tarot cards.
b. Most deans are druids. They march about waving mistletoe.
c. Most deans hunt small game. One caught a pigeon.

Finally, let us turn now to examples of quantifiers with multiple heads like "More doctors than lawyers eat pez." Here is a candidate game rule:

(R.more-than)
If the game $G(S; M)$ has reached an expression of the form:

$$Z - \text{more } X \text{ who } Q \text{ than } Y \text{ who } R - W$$

Then the verifier picks a set of entities ver(M) of cardinality n and the falsifier likewise picks a set of entities fal(M) also of cardinality n. The falsifier picks $d \in$ ver(M) and the verifier picks $c \in$ fal(M). The game continues on: $Z - d - W$ and not $Z - c - W$ and d is an X and c is an Y and $d\ Q$ and $c\ R$.
Both ver(M) and fal(M) are added to the choice set I_S.

The effect of (R.more-than) on discourse anaphora is far less clear. As we would expect, use of a singular definite pronoun is not allowed:

(19) More deans than faculty eat three square meals a day. #He is getting fat that way.

The proper interpretation of plural definite pronouns is less clear:

(20) a. More deans than faculty eat three square meals a day. They need to keep up their blood sugar.
(*They* being the deans)

b. More deans than faculty eat three squares a day. They want to keep their weight down.
(*They* being the faculty)

It seems to me that (20)a is somewhat more comfortable than (20)b, but that the latter is still possible. I have, therefore, added both ver(M) and fal(M) to the choice set.

Equally, it seems to me that either argument of *more-than* can provide a basis for indefinite singular anaphora:

(21) a. More deans than faculty eat three squares a day. One danced a merry jig to taunt the assembled faculty.
(*One* being one of the deans.)
b. More deans than faculty eat three squares a day. One whined piteously outside the deans meeting.
(*One* being one of the faculty.)

The game rules given in this section work for finite sets. Can they be adapted for infinite sets? The treatment of generalized quantifiers given in Keenan and Stavi (1986) uses infinitary means, in the guise of arbitrary meets and joins, to derive higher-order quantifiers. For example, *most* is defined as the arbitrary meet of an infinite family of generalized quantifier denotations, built from the basic cardinal determiners.[7]

One might think that infinite would correspond to an infinite round of games over finite samples. Thus, we might let the verifier pick a sample from the model on which the game is played. These sub-games could be repeated infinitely. But, of course, this will not work; the verifier could cheat infinitely. For example, she will always be able to pick a biased set for:

(22) Most numbers are even.

and win each time. Thus, using an infinite round of finite games will not work. Instead, Eloïse and Abelard must be locked in an infinite game, whatever that means. To my mind, a better option would be for Eloïse to offer Abelard a convincing proof (or vice versa) of the truth (or falsity) of the proposition.

References

Barwise, J., and Cooper, R. (1981) Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.

Beaver, D. I. (2001) *Presupposition and Assertion in Dynamic Semantics*. CSLI, Stanford, CA.

[7]The semantic automata framework of van Benthem (1986) which simulates higher-order quantifiers with push-down automata does not even treat the problem of infinite sets, since the models are encoded as finite strings.

Clark, R. (2004) Quantifier games and reference tracking. M.S. University of Pennsylvania, presented at the 2004 Prague Colloquium.

Clark, R. (2005) Discourse and quantifier games. M.S. University of Pennsylvania.

Groenendijk, J., and Stokhof, M. (1991) Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100.

Hintikka, J. (1996) *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge.

Hintikka, J., and Kulas, J. (1985) *Anaphora and Definite Descriptions: Two Applications of Game-Theoretical Semanitcs*. D. Reidel, Dordrecht, The Netherlands.

Hintikka, J., and Sandu, G. (1997) Game-theoretical semantics. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, Eds. MIT, Cambridge, MA, pp. 361–410.

Kamp, H., and Reyle, U. (1993) *From Discourse to Logic*. Kluwer, Dordrecht, The Netherlands.

Keenan, E. (1992) Beyond the frege boundary. *Linguistics and Philosophy*, 15:199–221.

Keenan, E. L., and Stavi, J. (1986) A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9:253–326.

Keenan, E. L., and Westerståhl, D. (1997) Generalized quantifiers in linguistics and logic. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, Eds. MIT, Cambridge, MA, pp. 837–893.

Mostowski, A. (1957) On a generalization of quantifiers. *Fundamenta Mathematicæ*, 44:12–36.

Pietarinen, A.-V. (2001) Most even budged yet: Some cases for game-theoretic semantics in natural language. *Theoretical Linguistics*, 27:20–54.

Pietarinen, A.-V. (2007) Semantic games and generalised quantifiers. In *Game Theory and Linguistic Meaning*, A.-V. Pietarinen, Ed. Elsevier, Oxford. pp. 183–206.

van Benthem, J. (1986) *Essays in Logical Semantics*. D. Reidel, Dordrecht, The Netherlands.

van Benthem, J. (1989) Polyadic quantifiers. *Linguistics and Philosophy*, 12:437–465.

van den Berg, M. (1996) Dynamic generalized quantifiers. In *Quantifiers, Logic, and Language*, J. van der Does and J. van Eijck, Eds. CSLI, Stanford, CA, pp. 63–94.

Part III

DIALOGUES

Chapter 8

FROM GAMES TO DIALOGUES AND BACK

Towards a General Frame for Validity

Shahid Rahman[1] and Tero Tulenheimo[2]
[1] *U.F.R. de Philosophie, Université Lille 3*
shahid.rahman@univ-lille3.fr

[2] *Academy of Finland; Department of Philosophy, University of Helsinki*[*]
tero.tulenheimo@helsinki.fi

Abstract In this article two game-theoretically flavored approaches to logic are systematically compared: dialogical logic founded by Paul Lorenzen and Kuno Lorenz, and the game-theoretical semantics of Jaakko Hintikka. For classical propositional logic and for classical first-order logic, an exact connection between 'intuitionistic dialogues with hypotheses' and semantic games is established. Various questions of a philosophical nature are also shown to arise as a result of the comparison, among them the relation between the model-theoretic and proof-theoretic approaches to the philosophy of logic and mathematics.

8.1 Introduction

The fact that game-theoretical semantics (GTS) and dialogic are sisters has been widely acknowledged. The differences between the original approaches have been discussed too: while GTS relates to the study of *truth in a model*, dialogic has explored the possibilities of a certain type of proof-theoretic approach to *validity*. Despite the close relationship between the two approaches, no detailed, thorough analysis of their interaction has yet been undertaken. The insightful article of Saarinen (1978) is, however, a notable early attempt at a comparison of the two viewpoints. The aim of this paper is to present, by means

[*]Partially supported by a personal grant from Finnish Cultural Foundation; and partially carried out within the project 207188 of the Academy of Finland. Work done in part at UMR Savoirs et Textes, Université Lille 3.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

of analyzing the notion of validity, a systematic comparison of dialogical logic and GTS, in the hope of stimulating a fruitful dialogue between and around the two approaches.[1]

8.1.1 Characterization of semantic properties

Truth and *validity*—or *material truth* and *logical truth*, respectively—are the two most important semantic properties that logics deal with. Semantically, logics are used for making assertions about *models*, and a part of the specification of a semantics for a logic is telling under which conditions a formula of such-and-such logic is true, relative to a given model. Logical truth then means truth with respect to *all* models relative to which the semantics is defined. Truth is sometimes qualified as *material* truth, to convey the idea that this notion of truth is relative to a contingent context. For some logics the notions of truth and validity admit of generalization. Hence in first-order logic, truth (in a model) is a special case of satisfaction (in a model and under a variable assignment); and logical truth is an instance of satisfaction in every model and under all variable assignments.

A logic for which a semantics is specified in some way, typically admits of conceptually different ways of capturing the notions of truth and validity appropriate to that logic. For instance, the most common way of defining the semantics of first-order logic is by defining satisfaction conditions of its formulas relative to a model and a variable assignment, by recursion on the structure of a formula. This was Tarski's original approach in defining the semantics of first-order logic (Tarski, 1933, Tarski and Vaught, 1956). An alternative to the Tarskian way of specifying the semantics would be game-theoretical semantics (Hintikka, 1968, 1973; Hintikka and Sandu 1997), which captures the very same satisfaction conditions in terms of the existence of a winning strategy for a certain player in a semantic game, associated with a formula, a model and a variable assignment.[2]

Alternative ways of specifying a semantics are said to *characterize* the notions *defined* by the specification of the logic. In this sense, GTS serves to characterize the semantics of first-order logic that is defined by the Tarskian

[1] In linguistics, Lauri Carlson has formulated so-called 'dialogue games' in his studies of discourse analysis. (See, e.g., Carlson [1983].) While he was inspired by Hintikka, the resulting analysis goes well beyond game-theoretical semantics. It would be a possible further line of research to compare the three approaches of Hintikka's GTS, Lorenzen and Lorenz's dialogic, and Carlson's dialogue games, as applied to natural language analysis.

[2] The characterization requires applying the standard interpretation of second-order logic in the sense of Henkin (1950). Furthermore, if strategies are formulated as functions ('deterministic strategies'), the characterization is subject to assuming the *Axiom of Choice*. This assumption is not needed if strategies are formulated as 'non-deterministic' (i.e., whenever the player is to make a move, his strategy is allowed to offer him several options, instead of a single value); cf. Hodges (2006).

semantics. Similarly, dialogues associated with first-order sentences serve to characterize validity in first-order logic, i.e., the same property that under the Tarskian approach is captured by the condition 'true in every model', or, proof-theoretically, as derivability in a complete and sound proof system from the empty set of premises. (As is well known, such a proof system exists for first-order logic.) This means that the same logic would have been obtained, had any of the characterizations been used in place of the original definition of truth or validity for the logic in question.

Yet the different ways of capturing the same notions may make it possible to pose questions that would otherwise not have appeared. For example, various questions whose original motivation derives from game theory, arise in connection with logics whose semantics is defined game-theoretically. Cases in point are issues of determinacy (whether always one of the players has a winning strategy), imperfect information (whether a player is always fully informed of a past course of a play), and strategic action (what are the different ways in which a true sentence may be verified), which all have turned out to function as grounds for interesting generalizations of first-order logic—as witnessed by Hintikka's so-called IF, or 'independence-friendly', logic (see, e.g., Hintikka, 1996, 2002; Hintikka and Sandu, 1997), and the research pursued within the framework of the 'Games and Logic' paradigm of van Benthem and other Dutch logicians (see, e.g., van Benthem, 2001a, b, 2002). The usefulness of the dialogical approach in initiating novel perspectives in connection with linear logic is another example of the fruitfulness of the game-theoretical approach (Blass, 1992). Thus alternative formulations of semantics may enable asking new questions; what is more, the requisite new conceptual tools may actually make it possible to study logics that could not even be formulated in terms of the traditional tools—or whose formulation using the received tools would in any event be clumsier. Examples are logics with Henkin quantifiers (Henkin, 1961; Krynicki and Mostowski, 1995), infinitely deep languages (Hintikka and Rantala, 1976; Karttunen, 1984; Hyttinen, 1990) and Vaught sentences (Vaught, 1973; Makkai, 1977), which all extend first-order logic. They all are very naturally defined using games.

Dialogues and GTS: characterizing validity and truth. In this paper we will be concerned with game-theoretical methods, used for defining semantically important notions.

While GTS has been from the beginning clearly model-theoretically oriented, the dialogical approach has a strong connection to proof theory—historically, conceptually, and philosophically. For logics admitting a complete proof system (such as propositional logic and first-order logic), the gap between model theory and proof theory is of course bridgeable, but even so

the two backgrounds lead easily to different types of development, so much so that even the corresponding ultimate understanding of what semantics is may be affected: there are constructivistically oriented philosophers who consider proof-theoretic inference rules as meaning-constitutive and who speak of *proof-conditional* semantics for logical operators (see Ranta, 1988; Sundholm, 2002), whereas from the viewpoint of classical model theory there is no proof-theoretic component at all to the semantics of logical operators. A sense of an existing common ground between GTS and the dialogical approach is still unmistakable.

Specifically, we establish for classical propositional logic and classical first-order logic an exact connection between 'intuitionistic dialogues with hypotheses' and semantic games. Basically, we show how the existence of a winning strategy for one of the players (called *Proponent*) in a dialogue $\mathcal{D}(A; H_1, \ldots, H_n)$ corresponding to a sentence A with a finite number of hypotheses H_i of a certain type, gives rise to a family of *Eloise*'s winning strategies in semantic games $G(A, M)$, one strategy for each model M; and, conversely, how to construct a winning strategy for *Proponent* in the dialogue $\mathcal{D}(A; H_1, \ldots, H_n)$ out of *Eloise*'s winning strategies in games $G(A, M)$. The proofs are *constructive* in the sense that we explicitly show, by providing a suitable explicit recipe, how a strategy for one type of game is built using a strategy for the other type of game.[3] In fact, these explicit sets of instructions are the real content of our results—it is well known that abstractly, validity in one sense (dialogic) coincides with validity in the other sense (GTS), simply because they both characterize the notion 'true in all models'.

8.1.2 The languages considered

The languages to be considered are propositional logic and first-order logic.

Propositional logic. Given a countable set **prop** of propositional atoms (denoted $p, q, \ldots, p_0, p_1, \ldots$), we consider propositional logic (**PL**) with the connectives conjunction ($\wedge$), disjunction ($\vee$) and negation ($\neg$). (By 'countable' we mean 'finite or of size $\aleph_0$'.) Semantics of **PL** is relative to models M : **prop** $\longrightarrow$ {`true`, `false`}. Such a model M partitions the set of propositional atoms into two classes: those that are true in the model, and those that are false. We assume that the reader is familiar with the semantics of propositional logic.

[3] In connection with propositional logic such recipes can be formulated as algorithms. On the other hand, when discussing first-order logic, attention cannot be restricted to finite models. But then, infinite models are not in general representable by finite means, wherefore in connection with first-order logic we cannot even try to formulate the relevant instructions as algorithms.

In our official syntax, the implication sign ($\rightarrow$) does not appear.[4] For classical propositional logic this is no restriction, as there implication can be defined from disjunction and negation:

$$A \rightarrow B := \neg A \vee B.$$

For intuitionistic logic this *is* a genuine restriction, however, since intuitionistically implication is not definable from the other connectives. In particular, intutionistically $A \rightarrow B$ is weaker than $\neg A \vee B$: from $\neg A \vee B$ it follows intuitionistically that $A \rightarrow B$, but not *vice versa*. Mostly in this paper we consider languages without implication. Sometimes, however, we phrase definitions for the extended language involving implication, to give a fuller picture of the logical situation.

The notion of (proper) *subformula* of a formula is defined in the usual way: $Sub(p) = \varnothing$; $Sub(B \vee C) = Sub(B \wedge C) = \{B, C\} \cup Sub(B) \cup Sub(C)$; and $Sub(\neg B) = \{B\} \cup Sub(B)$. A propositional formula A is said to be in *negation normal form*, if the negation sign ($\neg$) appears in A only prefixed to atomic subformulas: if $\neg B$ is a subformula of A, then $B \in$ **prop**. It is not difficult to verify that every propositional formula has an equivalent in negation normal form:

Fact 1. *For every $A \in$ **PL** there is $B \in$ **PL** such that B is in negation normal form, and A is logically equivalent to B.*

First-order logic. Let τ be a countable vocabulary, i.e., a countable set consisting of constants $c_0, c_1, \ldots$ and relation symbols $\mathrm{R}_0, \mathrm{R}_1, \ldots$.[5] Each relation symbol is associated with a positive natural number, called its *arity*. Let a set of individual variables, $Var = \{x_0, x_1, \ldots\}$, be fixed. Constants and variables are jointly referred to as *terms*. Atomic first-order formulas are strings of the form

$$\mathrm{R}_i t_1 \ldots t_n,$$

where $\mathrm{R}_i \in \tau$ is n-ary and each t_j is a term.

The class of formulas of first-order logic of vocabulary τ, or $\mathbf{FO}[\tau]$, is obtained by closing the set of atomic formulas under conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$), as well as under universal ($\forall x_i$) and existential ($\exists x_i$) quantification, $x_i \in Var$. Sometimes we wish to consider an extension $\mathbf{FO}[\tau, =]$ of $\mathbf{FO}[\tau]$ termed *first-order logic with equality*. It is obtained from $\mathbf{FO}[\tau]$ by introducing the identity symbol '=' as an additional logical symbol (hence not

[4]The implication sign, "$\rightarrow$", is not to be confused with the sign "$\longrightarrow$" used to indicate the domain and range of a function, as when writing $f : A \longrightarrow B$.

[5]For simplicity we assume τ not to contain function symbols.

included in the vocabulary τ, all of whose symbols are non-logical), and allowing strings $t_1 = t_2$ as additional atomic formulas, for any terms t_1, t_2.

We use capital letters $A, B, C, \ldots$ from the beginning of the alphabet for arbitrary (atomic or complex) formulas. The notion of (proper) *subformula* is obtained by extending the definition of subformula of a **PL**-formula by the clauses: $Sub(\forall x_i B) = Sub(\exists x_i B) = \{B\} \cup Sub(B)$. The set $Free[B]$ of *free variables* of a formula is defined recursively as usual:

- $Free[\mathrm{E}t_1 \ldots t_n] = \{t_1, \ldots, t_n\} \cap Var$, if $\mathrm{E} \in \tau \cup \{=\}$.
- $Free[\neg B] = Free[B]$.
- $Free[(B \wedge C)] = Free[(B \vee C)] = Free[B] \cup Free[C]$.
- $Free[\forall x_i B] = Free[\exists x_i B] = Free[B] \setminus \{x_i\}$.

Formulas whose set of free variables is empty, are *sentences*. Sometimes we will write $A(x_1, \ldots, x_n)$ to indicate that the free variables of A are among $x_1, \ldots, x_n$. Semantics of $\mathbf{FO}[\tau]$ is defined relative to τ-structures, i.e., structures $\mathcal{M}$ consisting of a non-empty domain M together with interpretations of the symbols appearing in the vocabulary τ: interpretation $c_j^{\mathcal{M}}$ of a constant c_j is simply an element of the domain, while the interpretation $\mathrm{R}_i^{\mathcal{M}}$ of an n-ary relation symbol R_i is an n-ary relation on M, i.e., a subset of the product M^n. In $\mathbf{FO}[\tau, =]$, the identity symbol is rigidly interpreted by the identity relation. We assume that the reader is familiar with the semantics of first-order logic, i.e., the recursive definition of the relation "$(\mathcal{M}, \gamma) \models A$" for all $\mathbf{FO}[\tau]$-formulas, τ-structures $\mathcal{M}$ and variable assignments $\gamma : Free[A] \longrightarrow M$.

Note that the implication sign is not among the logical symbols of $\mathbf{FO}[\tau]$. However, like for classical propositional logic, also for classical first-order logic implication could be introduced as a defined connective. We will write $\mathbf{FO}[\rightarrow, \tau]$ for first-order logic with implication, i.e., the logic otherwise like $\mathbf{FO}[\tau]$ but having the definable symbol $\rightarrow$ as one of its logical constants. The negation normal form result of Fact 1 extends straightforwardly to $\mathbf{FO}[\rightarrow, \tau]$:

Fact 2. *For every $A \in \mathbf{FO}[\rightarrow, \tau]$ there is $B \in \mathbf{FO}[\rightarrow, \tau]$ such that B is in negation normal form, and A is logically equivalent to B (that is, is satisfied in exactly the same τ-structures by precisely the same variable assignments).*

8.2 Formal dialogues

Let us see what is at stake in dialogical logic by reconstructing in dialogical terms the notion of validity of first-order logic. (For a somewhat different account, see Rahman and Keiff [2005].) We first define a language $\mathcal{L}[\tau]$; this language will basically be obtained from first-order logic of vocabulary τ by adding certain metalogical symbols. For the sake of fuller exposition we consider first-order logic with implication, or $\mathbf{FO}[\rightarrow, \tau]$.

We introduce special *force symbols* ? and !. An *expression* of $\mathcal{L}[\tau]$ is either a formula of $\mathbf{FO}[\rightarrow, \tau]$, or one of the following strings:

$$L, R, \vee, \forall x_i/c_j \text{ or } \exists x_i,$$

where x_i is a variable and c_j a constant. Expressions of the latter type are referred to as *attack markers*. In addition to expressions and force symbols, for $\mathcal{L}[\tau]$ we have available *labels* **O** and **P**, standing for the players (*Proponent*, *Opponent*) of dialogues. We will refer to **P** as 'she' and to **O** as 'he'. Every expression e of $\mathcal{L}[\tau]$ can be augmented with labels **P** or **O** on the one hand, and with force symbols ? and ! on the other, so as to yield the strings

$$\mathbf{P}\text{-}!\text{-}e,\ \mathbf{O}\text{-}!\text{-}e,\ \mathbf{P}\text{-}?\text{-}e \text{ and } \mathbf{O}\text{-}?\text{-}e.$$

These strings are said to be *(dialogically) signed expressions*. Their role is to signify that in the course of a dialogue, the move corresponding to the expression e is to be made by **P** or **O**, respectively, and that the move is made as a defense (!) or as an attack (?). We will use X and Y as variables for **P** and **O**, always assuming $X \neq Y$.

8.2.1 Particle rules

Dialogues have two types of rules: particle rules and structural rules. The former are meant to provide a schematic description of the key semantic features of logical operators. The latter, again, are chosen differently for different purposes. In Section 8.2.2 a set of structural rules will be considered which allows using dialogues for strictly proof-theoretic purposes (characterizing validity).

An *argumentation form* or *particle rule* is an abstract schematic description of the way a formula, according to its outmost form, can be criticized and how the critique can be answered. It is abstract in the sense that this description can be carried out without reference to a specified context. In dialogical logic, these rules are said to state the *local semantics*: what is at stake is only the critique and the answer corresponding to a given logical constant, rather than the whole context where the logical constant is embedded—a context which varies with the choice of structural rules.[6] The particle rules fix the dialogical semantics of the logical constants of $\mathcal{L}[\tau]$ in the following way:

	$\wedge$	$\vee$	$\rightarrow$
Assertion	X-!-$A \wedge B$	X-!-$A \vee B$	X-!-$A \rightarrow B$
Attack	Y-?-L or Y-?-R	Y-?-$\vee$	Y-!-A
Defense	X-!-A *resp.* X-!-B	X-!-A or X-!-B	X-!-B

[6]There can be no particle rule corresponding to atomic formulas. On the other hand, we can consider dialogues in which *Opponent* is right at the beginning committed to a number of additional initial concessions; such initial concessions may, in particular, be atomic. This is the case with 'material dialogues' explained in Section 8.4, and in 'dialogues with hypotheses' that we will make extensive use of in the present paper.

	$\forall$	$\exists$	$\neg$
Assertion	X-!-$\forall xA$	X-!-$\exists xA$	X-!-$\neg A$
Attack	Y-?-$\forall x/c$ for any c available to Y	Y-?-$\exists x$	Y-!-A
Defense	X-!-$A[x/c]$	X-!-$A[x/c]$ for any c available to X	–

In the diagram, $A[x/c]$ stands for the result of substituting the constant c for every free occurrence of the variable x in the formula A.

Note that particle rules themselves leave it open what types of entities the objects c are which are chosen in connection with quantifier rules. Likewise the particle rules need not specify which objects c are indeed available to the relevant player at a given stage of a dialogue. In the internal division of labor between particle and structural rules, these specifications are left for the latter. For example, the structural rules corresponding to characterizing validity will specify that the entities c are individual constants from some specified set $\{c_0, c_1, \ldots\}$. The structural rules will also specify that when **P** defends an existentially quantified sentence, as well as when **P** attacks a universally quantified sentence, the constant c must be chosen among constants already introduced in the dialogue, while when **O** defends an existentially quantified sentence, as well as when **O** attacks a universally quantified sentence, the constant c must be fresh in the sense of not having been yet used in the dialogue.

A more thorough way to stress the sense in which the particle rules determine local semantics is to see these rules as defining *state* of a (structurally not yet determined) game.

Definition 3 (State of a dialogue)**.** *Let $A \in \mathbf{FO}[\rightarrow, \tau]$, and let a countable set $\{c_0, c_1, \ldots\}$ of individual constants be fixed. A* state of the dialogue $\mathcal{D}(A)$ *about the formula A is a quintuple $\langle B, X, f, e, \sigma\rangle$ such that:*

- *B is a (proper or improper) subformula of A.*
- *X-f-e is a dialogically signed expression: $X \in \{\mathbf{O}, \mathbf{P}\}$, $f \in \{?, !\}$, and $e \in \mathcal{L}[\tau]$.*
- *$\sigma : Free[B] \longrightarrow \{c_0, c_1, \ldots\}$ is a function mapping the free variables of B to individual constants.*

The component e is either a formula of $\mathbf{FO}[\rightarrow, \tau]$ or an attack marker. We stipulate that in the former case, always $e := B$.

Given a force f, let us write f' for the opposite force, i.e., let $f' \in \{?, !\} \setminus \{f\}$. Each state $\langle B, X, f, e, \sigma\rangle$ has an associated *role assignment*, indicating which player occupies the role of *Challenger* (?) and which the role of *Defender* (!). The role assignment is a function $\rho : \{\mathbf{P}, \mathbf{O}\} \longrightarrow \{?, !\}$ such that $\rho(X) = f$ and $\rho(Y) = f'$.

State $\langle B_2, X_2, f_2, e_2, \sigma_2\rangle$ is *reachable* from state $\langle B_1, X_1, f_1, e_1, \sigma_1\rangle$ if it is a result of X_1 making a move in accordance with the appropriate particle rule in the role f_1. If the role is that of *Challenger* ($f_1 = ?$), the player states an attack, whereas if the role is that of *Defender* ($f_1 = !$), the player poses a defense.

Let us take a closer look at the transitions from one state to another. Particle rules determine which state S_2 of a dialogue is reachable from a given other state S_1. Note that the player who defends need not be the same at both states. In order for S_2 to be reachable from $S_1 = \langle B, X, f, e, \sigma\rangle$, it must satisfy the following.

- *Particle rule for negation*: If $B = e$, $f = !$ and B is of the form $\neg C$, then $S_2 = \langle C, Y, !, C, \sigma\rangle$. So if **P** is *Defender* of $\neg C$ at S_1, then **O** is *Defender* of C at S_2, and **P** will challenge (counterattack) C; and dually, if **P** is *Challenger* of $\neg C$ at S_1.

Here state S_2 involves the claim that C can be defended; however, this claim has been asserted in the course of an attack, and the whole move from S_1 to S_2 counts as an attack on the initial negated formula, i.e., an attack on C. Actually, this follows from the fact that at S_2, the roles of the players are inverted as compared with S_1. Counterattack may yield from S_2 a further state, $S_3 = \langle C, X, ?, *, \sigma\rangle$, where C is the formula considered, and the attack pertains to the relevant logical constant of C, for which $*$ is a suitable attack marker.

- *Particle rule for conjunction*: If $B = e$, $f = !$ and B is of the form $C \wedge D$, then $S_2 = \langle C, X, !, C, \sigma\rangle$ or $S_2 = \langle D, X, !, D, \sigma\rangle$, according to the choice of *Challenger* between the attacks ?-L and ?-R. (Here *Challenger* is Y: Y's role is ? here.)

- *Particle rule for disjunction*: If $B = e$, $f = !$ and B is of the form $C \vee D$, then $S_2 = \langle C, X, !, C, \sigma\rangle$ or $S_2 = \langle D, X, !, D, \sigma\rangle$, according to the choice of *Defender*, reacting to the attack ?-$\vee$ of *Challenger*. (Here *Defender* is X: X's role is ! here.)

- *Particle rule for implication*: If $B = e$, $f = !$ and B is of the form $C \rightarrow D$, then $S_2 = \langle C, Y, !, C, \sigma\rangle$ and, further, state $S_3 = \langle D, X, !, D, \sigma\rangle$ is reachable from S_2. So if **P** is *Defender* of $C \rightarrow D$ at S_1, and hence **O** is *Defender* of C at S_2, it is **P** who will be *Defender* of D at S_3.

To attack an implication amounts to being prepared to defend its antecedent, and so it should be noticed that the defense of C at state S_2 counts as an attack. If **P** is *Defender* of $C \rightarrow D$ at S_1, at state S_3 reachable from S_2, either **P** may defend D, or else **P** may counterattack C, thus yielding a further state, $S_4 = \langle C, X, ?, *, \sigma\rangle$, where C is the formula considered, and the relevant logical particle of C is attacked, $*$ being an appropriate attack marker.

- *Particle rule for the universal quantifier:* If $B = e$, $f = !$ and B is of the form $\forall x D(x)$, then $S_2 = \langle D(x), X, !, D(x), \sigma[x/c_i]\rangle$, where c_i is the constant chosen by *Challenger* (who here is Y) as a response to the attack ?-$\forall x/c_i$.

As usual, the notation '$\sigma[x/c_i]$' stands for the function that is otherwise like σ, but maps the variable x to c_i. Hence if σ is already defined on x, $\sigma[x/c_i]$ is the result of reinterpreting x by c_i; otherwise it is the result of extending σ by the pair (x, c_i).

- *Particle rule for the existential quantifier:* If $B = e$, $f = !$ and B is of the form $\exists x D(x)$, then $S_2 = \langle D(x), X, !, D(x), \sigma[x/c_i]\rangle$, where c_i is the constant chosen by *Defender* (who is X here), reacting to the attack ?-$\exists x$ of *Challenger* (that is, Y).

8.2.2 Structural rules

When analyzing dialogues, we will make use of the following notions: *dialogue*, *dialogical game*, and *play* of a dialogue. It is very important to keep them conceptually distinct. Dialogical games are sequences of dialogically signed expressions, i.e., expressions of the language $\mathcal{L}[\tau]$ equipped with a pair of labels, **P**-!, **O**-!, **P**-?, or **O**-?. The labels carry information about how the dialogue proceeds. Dialogical games are a special case of plays: all dialogical games are plays, but not all plays are dialogical games. However, all plays are *sequences* of dialogical games. Finally, dialogues are simply sets of plays.

A complete dialogue is determined by game rules. They specify how dialogical games in particular, and plays of dialogues in general, are generated from the thesis of the dialogue. Particle rules are among the game rules, but in addition to them there are *structural rules*, which serve to specify the general organization of the dialogue.

Different types of dialogues have different kinds of structural rules. *When the issue is to characterize validity*—as it is for the dialogues considered in the present paper—a dialogue can be thought of as a tree, whose (maximal) branches are (finished) plays relevant for establishing the validity of the thesis. The structural rules will be chosen so that *Proponent* succeeds in defending the thesis against all allowed critique of *Opponent* if, and only if, the thesis is valid in the standard sense of the term ('true in every model'). In dialogical logic the existence of such a winning strategy for *Proponent* is typically taken as the *definition* of validity; however, this dialogical definition indeed captures the standard notion (see the discussion in connection with Definition 5 below).

Each split into two branches—into two plays—in a dialogue tree should be considered as the outcome of a propositional choice made by *Opponent*. Any choice by **O** in defending a disjunction, attacking a conjunction, and reacting to an attack against a conditional, gives rise to a new branch: a new play.

By contrast, *Proponent*'s choices do not generate new branches; and neither do *Opponent*'s choices for quantifiers (defending an existential quantifier, attacking a universal quantifier).

The participants **P** and **O** of the dialogues that we are here interested in—the dialogues used for characterizing validity—are of course idealized agents. If real-life agents took their place, it might happen that one of the players was cognitively restricted to the point of following a strategy which would make him lose against some, or even every sequence of moves by the opponent—even if a winning strategy would be available to him. The idealized agents of the dialogues are not hence restricted: their 'having a strategy' means simply that there exists, by combinatorial criteria, a certain kind of function; it does not mean that the agent possesses a strategy in any cognitive sense.

Plays of a dialogue are sequences of dialogically signed expressions. In particular, plays can always be analyzed into dialogical games: any play is of the form $\langle \Delta_1, \ldots, \Delta_n \rangle$, where the Δ_i are dialogical games. The case $n := 0$ yields the empty sequence $\langle\rangle$, referred to as the *empty play*. By stipulation the empty play is identified with the *empty dialogical game*. The members of plays other than the first member (if any) are termed *moves*, the first member being termed the *thesis*. A move is either an attack or a defense. The particle rules stipulate which moves are to be counted as attacks. Exactly those moves X-f-e whose expression component e is a first-order formula, are said to have *propositional content*. Recall that in the case of implication and negation some moves with propositional content count as attacks. (In the actual design of a dialogue there usually is a notational device to differentiate between those moves with propositional content that are attacks and those that are not.)

We move on to introduce a number of structural rules for dialogues designed for the language $\mathcal{L}[\tau]$. We will write $\mathcal{D}(A)$ for the dialogue about A, i.e., the dialogue whose thesis is A. Further, we will write $\Delta[n]$ for the member of the sequence Δ with the position n. Let A be a first-order sentence of vocabulary τ. We have the following structural rules (SR-0) to (SR-6) regulating plays $\Delta \in \mathcal{D}(A)$, i.e., members of the dialogue $\mathcal{D}(A)$.

(SR-0) (Starting rule).

(a) If A is atomic, then $\mathcal{D}(A) = \{\langle\rangle\}$, i.e., the dialogue $\mathcal{D}(A)$ contains the empty play and nothing else; cf. rule (SR-5). Otherwise the dialogically signed expression $\langle$**P**-!-$A\rangle$ belongs to the dialogue $\mathcal{D}(A)$: the *thesis* A as stated by *Proponent* constitutes a play in the dialogue about A.

(b) If Δ is any non-empty play in the dialogue $\mathcal{D}(A)$, then the thesis A has position 0 in Δ: if $\Delta \in \mathcal{D}(A)$, then $\Delta[0] = \langle$**P**-!-$A\rangle$.

(c) At even positions **P** makes a move, and at odd positions it is **O** who moves. That is, each $\Delta[2n]$ is of the form $\langle$**P**-f-$B\rangle$ for some $f \in \{?, !\}$

and $B \in Sub(A)$; and each $\Delta[2n+1]$ is similarly of the form $\langle \mathbf{O}\text{-}f\text{-}B \rangle$. Every move after $\Delta[0]$ is a reaction to an earlier move made by the other player, and is subject to the particle rules and the other structural rules.

(SR-1.I) (Intuitionistic round closing rule). Whenever player X has a turn to move, he may attack any (complex) formula asserted by his opponent, Y, or he may defend himself against the *last not already defended attack* (i.e., the attack by Y with the greatest associated natural number such that X has not yet responded to that attack).

A player may postpone defending himself as long as he can perform attacks. Only the *latest* attack that has not yet received a response may be answered: If it is X's turn to move at position n, and positions l and m both involve an unanswered attack ($l < m < n$), then player X may *not* at position n defend himself against the attack of position l.

(SR-1.C) (Classical round closing rule). Whenever player X has a turn to move, he may attack any (complex) formula asserted by his opponent, Y, or he may defend himself against *any* attack, including those which have already been defended. That is, here even redoing earlier defenses is allowed.

(SR-2) (Branching rule for plays). If in a play $\Delta \in \mathcal{D}(A)$ it is **O**'s turn to make a propositional choice, that is, to defend a disjunction, attack a conjunction, or react to an attack against a conditional, then Δ extends into two plays $\Delta_1, \Delta_2 \in \mathcal{D}(A)$,[7]

$$\Delta_1 = \Delta^\frown\alpha \quad \text{and} \quad \Delta_2 = \Delta^\frown\beta,$$

differing in the chosen disjunct, conjunct *resp.* reaction, α *vs.* β. More precisely: Let $n \leq max\{m : \Delta[m]\}$.

- If $\Delta[n] = \langle \mathbf{O}\text{-}!\text{-}B \vee C \rangle$ and $\Delta[max] = \langle \mathbf{P}\text{-}?\text{-}\vee \rangle$, then

$$\alpha := \langle \mathbf{O}\text{-}!\text{-}B \rangle \quad \text{and} \quad \beta := \langle \mathbf{O}\text{-}!\text{-}C \rangle.$$

- If $\Delta[n] = \Delta[max] = \langle \mathbf{P}\text{-}!\text{-}B \wedge C \rangle$, then

$$\alpha := \langle \mathbf{O}\text{-}?\text{-}L \rangle \quad \text{and} \quad \beta := \langle \mathbf{O}\text{-}?\text{-}R \rangle.$$

[7] If $\bar{s} = (a_0, \ldots, a_n)$ is a finite sequence and a_{n+1} is an object, $\bar{s}^\frown a_{n+1}$ is by definition the sequence $(a_0, \ldots, a_n, a_{n+1})$. Generally, if $\bar{s} = (a_0, \ldots, a_n)$ and $\bar{s}' = (a'_0, \ldots, a'_{n'})$, then $\bar{s}^\frown \bar{s}' := (a_0, \ldots, a_n, a'_0, \ldots, a'_{n'})$. If $\bar{s} = \bar{s}_1{}^\frown \bar{s}_2$, then $\bar{s}_1$ is said to be an *initial segment* of $\bar{s}$, and, if the sequence $\bar{s}_2$ is not empty, its *proper* initial segment.

- If $\Delta[n] = \langle \mathbf{O}\text{-}!\text{-}B \to C\rangle$ and $\Delta[max] = \langle \mathbf{P}\text{-}!\text{-}B\rangle$, then

$$\alpha := \langle \mathbf{O}\text{-}?\text{-}*\rangle \quad \text{and} \quad \beta := \langle \mathbf{O}\text{-}!\text{-}C\rangle,$$

  where $*$ is an attack marker corresponding to the logical form of the formula B.

No moves other than propositional moves made by **O** will trigger branching.

(SR-3) (Shifting rule). When playing a dialogue $\mathcal{D}(A)$, **O** is allowed to switch between 'alternative' plays $\Delta, \Delta' \in \mathcal{D}(A)$. More exactly, if Δ involves a propositional choice made by **O**, then **O** is allowed to continue by switching to another play—existing by the *Branching rule* (SR-2). Concretely this means that the sequence $\Delta^\frown\Delta'$ will, then, be a play, i.e., an element of $\mathcal{D}(A)$.

It is precisely the *Shifting rule* that introduces plays which are not plain dialogical games. (Dialogical games are a special case of plays: those plays that are unit sequences of dialogical games.) As an example of applying the *Shifting rule*, consider a dialogue $\mathcal{D}(A)$ proceeding from the hypotheses $B, \neg C$, with the thesis $A := B \wedge C$. If **O** decides to attack the left conjunct, the result will be the play

$$(\langle \mathbf{P}\text{-}!\text{-}B \wedge C\rangle, \langle \mathbf{O}\text{-}?\text{-}L\rangle, \langle \mathbf{P}\text{-}!\text{-}B\rangle),$$

and **O** will lose. But then, by the *Shifting rule*, **O** may decide to have another try. This time he wishes to choose the right conjunct. The result is the play

$$(\langle \mathbf{P}\text{-}!\text{-}B \wedge C\rangle, \langle \mathbf{O}\text{-}?\text{-}L\rangle, \langle \mathbf{P}\text{-}!\text{-}B\rangle, \langle \mathbf{P}\text{-}!\text{-}B \wedge C\rangle, \langle \mathbf{O}\text{-}?\text{-}R\rangle, \langle \mathbf{P}\text{-}!\text{-}C\rangle).$$

Observe that this play consists of two dialogical games, namely $(\langle \mathbf{P}\text{-}!\text{-}B \wedge C\rangle, \langle \mathbf{O}\text{-}?\text{-}L\rangle, \langle \mathbf{P}\text{-}!\text{-}B\rangle)$ and $(\langle \mathbf{P}\text{-}!\text{-}B \wedge C\rangle, \langle \mathbf{O}\text{-}?\text{-}R\rangle, \langle \mathbf{P}\text{-}!\text{-}C\rangle)$. By contrast, it is not itself a dialogical game.

(SR-4) (Winning rule for plays). A dialogical game Δ_i is *closed* if in Δ_i there appears the same positive literal in two positions, one stated by X and the other one by Y. That is, Δ_i is closed if for some $k, m < \omega$ and some positive literal $\ell \in Sub(A)$, we have: $\Delta_i[k] = \ell = \Delta_i[m]$, where $k < m$ and furthermore, k is odd if, and only if m is even. A play $\Delta = \langle \Delta_1, \ldots, \Delta_n\rangle \in \mathcal{D}(A)$ whose most recent dialogical game Δ_n is closed, is said to be *closed* as well.

A dialogical game Δ_i is *maximal* if either Δ_i is closed, or within Δ_i all rules have been applied in a maximal fashion so that the play could only continue if **O** applied the *Shifting rule*. A dialogical game is *open* if it is maximal but not closed. In particular, the empty dialogical game $\langle\rangle$ is open. A play $\Delta = \langle \Delta_1, \ldots, \Delta_n\rangle$ is *open* if it is empty or its most recent dialogical game Δ_n is open.

A play is *finished* if it is either open, or else such that no further move is allowed by the *Shifting rule*. Observe that whenever a play $\Delta \in \mathcal{D}(A)$ is finished, there is no further play $\Delta' \in \mathcal{D}(A)$ such that Δ is an initial segment of Δ'.

Winning and losing are attributes that apply to finished plays. If a finished play is open, the player who stated the thesis (that is, **P**) *loses* the play, and **O** wins it. If, again, a play is finished and closed, **P** *wins* the play and **O** loses it. In dialogues—unlike in semantic games—there is no difference between 'winning a play' and 'having a winning strategy': **O** has a winning strategy in a dialogue iff the dialogue admits of an open dialogical game iff there is a finished play of the dialogue won by **O**; and **P** has a winning strategy in a dialogue iff the dialogue admits of a play that is both finished and closed (and so won by **P**).

(SR-5) (Formal use of atomic formulas). **P** cannot introduce positive literals: any positive literal must be stated by **O** first. From this it follows that a dialogue about an atom cannot have non-empty plays. Positive literals cannot be attacked.

In what follows we will consider, when speaking of first-order logic, intuitionistic dialogues with additional hypotheses of the following form:

$$\forall x_1 \ldots \forall x_n(\mathrm{E}x_1 \ldots x_n \vee \neg \mathrm{E}x_1 \ldots x_n),$$

where E is a relation symbol of a fixed vocabulary τ, or else the identity symbol. That is, the relevant hypotheses are instances of (a universal closure of) *tertium non datur*. In the presence of such hypotheses, we could use a more general formulation of the rule (SR-5):

(SR-5*). **P** cannot introduce literals: any literal (positive or not) must be stated by **O** first. Positive literals cannot be attacked.

Before we can state the structural rule (SR-6), or the *'No delaying tactics' rule*, we need some definitions.

Definition 4 (Strict repetition of an attack/a defense)**.** (*a*) *We speak of a* strict repetition of an attack *if a move is being attacked although the same move has already been challenged with the same attack before. (Note that even though choosing the same constant is a strict repetition, the choices of* ?*-L and* ?*-R are in this context different attacks.) In the case of moves where a universal quantifier has been attacked with a new constant, moves of the following kind must be added to the list of strict repetitions:*

> A universal quantifier move is being attacked using a new constant, although the same move has already been attacked before with a constant which was new at the time of that attack.

(*b*) *We speak of a* strict repetition of a defense *if a challenging move (attack) m_1, which has already been defended with the defensive move (defense) m_2 before, is being defended against the challenge m_1 once more with the same defensive move. (Note that the left and the right disjunct give rise to two different*

defenses in this context.) In the case of moves where an existential quantifier has been defended with a new constant, moves of the following kind must be added to the list of strict repetitions:

> An attack on an existential quantifier is being defended using a new constant, although the same quantifier has already been defended before with a constant which was new at the time of that defense.

According to these definitions, neither a new defense of an existential quantifier, nor a new attack on a universal quantifier, represents a strict repetition, if *it uses a constant that is not new but is however different from the one used in the first defense (or in the first attack).*

(SR-6) ('No delaying tactics' rule). This rule has two variants, classical and intuitionistic, depending on whether the dialogue is played with the classical structural rule (SR-1.C), or with the intuitionistic structural rule (SR-1.I).

Classical: No *strict* repetitions are allowed.

Intuitionistic: If **O** has introduced a new atomic formula which can now be used by **P**, then **P** may perform a repetition of an attack. No other *strict* repetitions are allowed.

Definition 5 (Validity)**.** *A first-order sentence A is* dialogically valid *in the classical (intuitionistic) sense if all finished plays belonging to the classical (*resp. *intuitionistic) dialogue* $\mathcal{D}(A)$ *are closed.*

It is possible to prove that the dialogical definition of validity coincides with the standard definition, both in the classical and in the intuitionistic case. First formulations of the proofs were given by Lorenz in his 1961 Ph.D. thesis *Arithmetik und Logik als Spiele*. Haas (1980) and Felscher (1985) proved the equivalence for intuitionistic first-order logic (by proving the correspondence between intuitionistic dialogues and intuitionistic sequent calculi), while Stegmüller (1964) established the equivalence in the case of classical first-order logic. Rahman (1994, 88–107) proved directly the equivalence between the two types of dialogues and the corresponding semantic tableaux, from which the result extends to the corresponding sequent calculi.

Let us take two examples of dialogues, one classical and the other intuitionistic.

Example 6. Consider the classical dialogue $\mathcal{D}(p \vee \neg p)$. Its thesis is $p \vee \neg p$, where p is an atomic sentence. In Figure 8.1, a dialogical game from dialogue $\mathcal{D}(p \vee \neg p)$ is described. This dialogical game is closed:

O			**P**	
			$p \vee \neg p$	0
1	?-$\vee$	0	$\neg p$	2
3	p	2	–	
[1]	[?-$\vee$]	[0]	p	4

Figure 8.1: Classical rules, **P** wins

The outer columns indicate the position of the move inside the dialogical game, while the inner columns state the position of the earlier move which is being attacked. The defense is written on the same line with the corresponding attack: an attack together with the corresponding defense constitutes a so-called *closed round*. The sign "–" indicates that there is no possible defense against an attack on a negation.

The dialogical game of the example is closed, because after **O**'s last attack in move 3, **P** is allowed—according to the classical rule SR-1.C—to defend (once more) herself against **O**'s attack made in move 1, and so the dialogical game in question becomes closed. **P** states her new defense in move 4. (In reality **O** does not repeat his attack of move 1: what we have written between square brackets simply serves to remind of the attack against which **P** is re-acting.)

In fact the described dialogical game is a finished play of the dialogue $\mathcal{D}(p \vee \neg p)$, and actually its only finished play: **O** could not prolong the play any further by making different moves. Hence not only is the described particular dialogical game closed—in fact **P** has a winning strategy in the dialogue, i.e., she is able to win no matter what **O** does. In other words, the sentence $p \vee \neg p$ is dialogically valid in the classical sense (cf. Definition 5).

Example 7. Let us consider the intuitionistic variant of the dialogue of the above example. In Figure 8.2, a dialogical game from the intuitionistic dialogue $\mathcal{D}(p \vee \neg p)$ is described. This dialogical game is open:

O			**P**	
			$p \vee \neg p$	0
1	?-$\vee$	0	$\neg p$	2
3	p	2	–	

Figure 8.2: Intuitionistic rules, **O** wins

The dialogical game constitutes a finished play, and it is **O** who wins the play of the example: no further move by **P** is possible following the intuitionistic structural rules, and the dialogical game is open. In particular remaking an earlier move—as in the above example of a classical dialogue—is not possible.

In fact **O** has trivially a winning strategy in the intuitionistic dialogue $\mathcal{D}(p \vee \neg p)$: **P** cannot prevent, by making different moves, **O** from generating precisely the described play won by **O**. Observe, in particular, that the sentence $p \vee \neg p$ is

not dialogically valid in the intuitionistic sense. (This does not mean, of course, that thereby the sentence $\neg(p \vee \neg p)$ would be intuitionistically valid!)

8.3 Game-theoretical semantics

Semantic games (sometimes referred to as evaluation games) provide a tool for defining, or characterizing, truth of a sentence in a model. This contrasts to the role of dialogues, used for defining, or characterizing, validity of a sentence, that is, truth of a sentence in all models. The general tactics for the use of games is basically the same in both cases: define games by laying down the game rules, and then define the logical property of interest (truth, validity) by reference to the existence of a winning strategy for a certain player of the relevant two-player game.

In this section we will introduce semantic games for propositional logic and first-order logic. By way of introduction, let us look at the case of propositional logic without implication, and in negation normal form. We associate with every formula A of **PL**, and every model M : **prop** $\longrightarrow$ {`true`, `false`}, a semantic game $G(A, M)$ between two players (*Abelard* and *Eloise*). The *game rules* of the games $G(A, M)$ are defined as follows, by recursion on the structure of the formula A.

(1^{+}_{AT}) If $A = p$ with $p \in$ **prop**, a play of the game has come to an end. If $M(p) =$ `true`, *Eloise* wins $G(A, M)$, and *Abelard* loses it. Otherwise *Abelard* wins and *Eloise* loses.

(2^{-}_{AT}) Also if $A = \neg p$, a play of the game has come to an end. If $M(p) =$ `false`, *Eloise* wins $G(A, M)$, and *Abelard* loses it. Otherwise the payoffs are reversed.

(3) If $A = (B \vee C)$, then *Eloise* chooses a disjunct $D \in \{B, C\}$, and the game goes on as $G(D, M)$.

(4) If $A = (B \wedge C)$, then *Abelard* chooses a conjunct $D \in \{B, C\}$, and the game goes on as $G(D, M)$.

The above game rules do not, by themselves, suffice for defining truth and falsity of propositional formulas. For this purpose, the notion of *winning strategy* is needed. A function f is said to be a *strategy* for *Eloise* in $G(A, M)$ if it provides a choice of a disjunct for every subformula of A of the form $(B \vee C)$, depending on the choices for the conjunctions already made when playing $G(A, M)$. A strategy f is a *winning strategy* (in short, a *w.s.*) for *Eloise* if against any sequence of moves made by *Abelard* following the game rules, making the choices for disjuncts in accordance with f leads to a game $G(D, M)$ won by *Eloise*, with D a literal (i.e., a propositional atom or a negation of a propositional atom). We define:

- A is true in M, in symbols $M \models^{+}_{GTS} A$, if there exists a w.s. for *Eloise* in $G(A, M)$.
- A is false in M, symbolically $M \models^{-}_{GTS} A$, if there exists a w.s. for *Abelard* in $G(A, M)$.

Note the following about the terminology: what really is won or lost is *such* a sequence of moves made in accordance with the game rules that cannot be further extended, i.e., which has reached a (negated) atomic formula. In technical terminology of game theory, such a sequence is called a *terminal play* (or *terminal history*). Hence *games* are not things that can be won—terminal plays are. On the other hand, it is for games that players have strategies. In the above formulation of game rules, we broke against this distinction by declaring, in rules (1^{+}_{AT}) and (2^{-}_{AT}), some *games* as being won or lost by one of the players. However, the games we were speaking of were *atomic* games in the sense of games $G(\ell, M)$ with $\ell = \pm p$ for some $p \in \mathbf{prop}$.[8] For such games there are no moves for either player, so who wins and who loses is entirely dependent on the literal ℓ and the model M. For these games having a winning strategy or winning an individual play are actually one and the same thing, so the deviation from the strict terminology is justified. In our more formal definition of semantic games for first-order logic below, we will stay with the usual terminology.

8.3.1 Negation

We have now sketched the definition of semantic games for propositional logic, assuming that the formulas are in negation normal form. This assumption is by no means necessary. To avoid it, we introduce the notion of *role* of a player. There are two roles available—*Verifier* and *Falsifier*—and always exactly one player assumes the role of *Verifier* and the other player that of *Falsifier*.

In propositional logic, role distribution ρ in a game $G(A, M)$ is relative to subformula tokens of A, and can be defined as follows. Let $B \in \{A\} \cup Sub(A)$. Then:

- If B is subordinate to an even number or zero negation signs in A, then $\rho(B, Eloise) = Verifier$; and $\rho(B, Abelard) = Falsifier$.
- Otherwise $\rho(B, Eloise) = Falsifier$; and $\rho(B, Abelard) = Verifier$.

Hence in particular *Eloise* is the initial verifier, and *Abelard* the initial falsifier: $\rho(A, Eloise) = Verifier$ and $\rho(A, Abelard) = Falsifier$. Having the role

[8] If A is a formula, we write '$B = \pm A$' as a shorthand notation for '$B \in \{A, \neg A\}$'.

distribution available, we may define semantic games $G(A, M)$ for arbitrary propositional formulas in the language without implication by replacing the condition (2^-_{AT}) for atomic negation by the following condition (2):

(2) If $A = \neg B$, then the game goes on as $G(B, M)$. (Observe that by definition $\rho(A, \mathit{Eloise}) = \rho(B, \mathit{Abelard})$.)

Furthermore, the payoffs of the players now depend on the role distribution; these payoffs were specified for formulas in negation normal form in conditions (1^+_{AT}) and (2^-_{AT}). Replace now the above condition (1^+_{AT}) by (1):

(1) If $A = p$ with $p \in \mathbf{prop}$, a play of the game has come to an end. If $M(p) = \texttt{true}$, then the player whose current role is *Verifier* wins $G(A, M)$, and the player whose current role is *Falsifier* loses it. Otherwise the player occupying the role of *Falsifier* wins and the one whose role is *Verifier* loses. (The player carrying a given role is determined by the function ρ: for each player $\mathcal{P}$, see whether the value $\rho(A, \mathcal{P})$ is *Verifier* or *Falsifier*.)

The rules (1), (2), (3) and (4) hence obtained define game rules for the full language of propositional logic without implication.

In classical propositional logic implication $B \to C$ is definable as $\neg B \vee C$. In accordance with this, we might introduce a game rule for (classical) implication in GTS as follows. First the definition of role distribution must be defined for the language with implication:

Definition 8 (Role distribution)**.** *If $B \in \{A\} \cup \mathit{Sub}(A)$, let n_B be the number of negation signs to which a subformula token B is subordinate in A; and let a_B be the number of* those *implication signs to which B is subordinate in A* and *which will yield B if their antecedents are chosen.*

- *If the number $n_B + a_B$ is even or equal to zero, then* $\rho(B, \text{Eloise}) = \text{Verifier}$*; and* $\rho(B, \text{Abelard}) = \text{Falsifier}$.
- *Otherwise* $\rho(B, \text{Eloise}) = \text{Falsifier}$*; and* $\rho(B, \text{Abelard}) = \text{Verifier}$.

(Mnemonics: 'n_B' for 'negation', and 'a_B' for 'antecedent'.)

For example, if $A := (B \to C)$, then $n_B = 0$ and $a_B = 1$; and if $A := \neg(B \to C) \to D$, then $n_B = 1$ and $a_B = 2$; whereas $n_C = 1$ and $a_C = 1$. Further, $n_D = a_D = 0$. Now the game rule for (classical) implication is:

(5) If $A = (B \to C)$, then *Eloise* chooses either the antecedent B or the consequent C. (Observe that by definition she assumes in the former case the current role of *Abelard*, while in the latter case she keeps her own current role: $\rho(B, \mathit{Eloise}) = \rho(A, \mathit{Abelard})$, but $\rho(C, \mathit{Eloise}) = \rho(A, \mathit{Eloise})$.)

8.3.2 Games in extensive form

We move on to describe semantic games for first-order logic. Attention will be restricted to sentences in negation normal form. (Recall that by Fact 2 all first-order sentences can be equivalently written in such a form.) We could define the games in the semi-formal fashion in which the semantic games for propositional logic were introduced above. To make things more explicit, however, we prefer to introduce them in what is termed in game theory the *extensive form* of a game.[9] That is, these games will be tuples $\langle N, H, Z, P, \langle u_i \rangle_{i \in N} \rangle$, where:

(i) N is the set of players of the game.

(ii) H is a set of sequences of elements from some given set $\mathcal{A}$ of *actions*. The members of H are called *histories*, or *plays* of the game.

(iii) Z is the set of *terminal histories* of the game, which, in the case that all histories are of finite length, are simply histories that cannot be extended by any action so as to yield a further history.

(iv) $P : H \setminus Z \longrightarrow N$ is the *player function* which assigns to every non-terminal history the player whose turn it is to move.

(v) For each $i \in N$, u_i is the *payoff function* for player i, that is, a function that specifies the payoffs (win or loss) of player i at terminal histories.

Satisfaction of first-order formulas of vocabulary τ is defined relative to τ-structures and variable assignments. Variable assignments can be taken to be functions from the set of free variables of a formula to elements of the domain of the relevant structure. Accordingly, systematically the most natural practice in GTS is to associate a semantic game with every formula $A \in \mathbf{FO}[\tau]$, τ-structure $\mathcal{M}$ and a variable assignment $\gamma : \mathit{Free}[A] \longrightarrow \mathit{dom}(\mathcal{M})$. In what follows, attention will be restricted to models with a countable domain. By the *Downward Löwenheim-Skolem theorem* a first-order formula is satisfied in a countable model if it is satisfiable at all. Hence theoretically there is no need for allowing models of larger than countable cardinality—given that our interest is in the expressive power of first-order logic.

We will associate with every first-order formula A (in negation normal form and written in vocabulary τ), every τ-structure $\mathcal{M}$ and every variable assignment $\gamma : \mathit{Free}[A] \longrightarrow \mathit{dom}(\mathcal{M})$, a game

$$G(A, \mathcal{M}, \gamma)$$

[9] Explicitly introducing semantic games in extensive form was suggested in Sandu and Pietarinen (2001, 2003). The underlying idea is of course as old as game-theoretical semantics itself, cf. e.g. Saarinen (1978). The distinction between the extensive and strategic (or 'normalized') form of a game was introduced in von Neumann and Morgenstern, 1944, see esp. Sections 11, 12), where these two forms were also shown to be strictly equivalent. What is at stake, then, is a familiar thing in a different, but obvious guise. This is not to deny that heuristically the two presentations may serve different purposes.

in extensive form. The set N of players of these games will be $\{Abelard, Eloise\}$. Due to the way in which semantic games for first-order logic will be specified, these games will be two-player zero-sum games of perfect information.

The related set of actions $\mathcal{A}$ consists of pairs (A_i, γ_i), where A_i is a subformula of A and γ_i is an assignment whose domain is the set $Free[A_i]$ of free variables of A_i. Histories of $G(A, \mathcal{M}, \gamma)$ are defined recursively, simultaneously with the player function:

1. $(A, \gamma) \in H$.
2. If $h = \langle (A_0, \gamma_0), \ldots, (A_n, \gamma_n) \rangle \in H$, then:
   - If $A_n = (\psi \wedge \phi)$, then $P(h) = Abelard$, and $h^\frown(\psi, \gamma_n) \in H$ and $h^\frown(\phi, \gamma_n) \in H$. If again $A_n = (\psi \vee \phi)$, then $P(h) = Eloise$, and $h^\frown(\psi, \gamma_n) \in H$ and $h^\frown(\phi, \gamma_n) \in H$.
   - If $A_n = \forall x\psi$, then $P(h) = Abelard$, and for all $a \in dom(\mathcal{M})$: $h^\frown(\psi, \gamma_n[x/a]) \in H$. If, on the other hand, $A_n = \exists x\psi$, then $P(h) = Eloise$, and for all $a \in dom(\mathcal{M})$: $h^\frown(\psi, \gamma_n[x/a]) \in H$.

So elements of the set H will be precisely all combinatorially possible sequences that can be built given that a quantifier is 'interpreted' by an element of the relevant domain, and a binary propositional connective by choosing its left or right term. Because the depth of a first-order formula is always finite,[10] all histories will likewise be of finite length.

Terminal histories are members h of H that cannot be further extended so as to yield a history. Hence the last member of a terminal history is of the form $(\mathrm{R}x_1 \ldots x_n, \gamma_i)$ or $(\neg \mathrm{R}x_1 \ldots x_n, \gamma_i)$. The players' payoffs (1 for win and -1 for loss) on terminal histories $h \in Z$ are determined as follows:

- $u_\exists(h) = 1$, if $(\mathcal{M}, \gamma_n) \models \ell$, where ℓ is the (negated) atomic formula in the last member of h; otherwise $u_\exists(h) = -1$.
- $u_\forall(h) = -u_\exists(h)$.

Note that since formulas were assumed to be in negation normal form, the definition of payoff need not be made dependent on a role distribution. On the other hand, the semantics is readily generalized to arbitrary first-order formulas (in the language without implication) simply by associating all plays with a role distribution and stipulating:

- $u_\exists(h, \rho) = 1$, if $[\rho(Eloise) = Verifier$ and $(\mathcal{M}, \gamma_n) \models \ell]$ or $[\rho(Eloise) = Falsifier$ *and* $(\mathcal{M}, \gamma_n) \not\models \ell]$, where ℓ is the (negated) atomic formula in the last member of h; otherwise $u_\exists(h, \rho) = -1$.

[10] Depth of a first-order formula A is the maximum number of syntactically subordinate quantifiers and propositional connectives appearing in A.

- $u_{\forall}(h, \rho) = -u_{\exists}(h, \rho)$.

The definition of games $G(A, \mathcal{M}, \gamma)$ in extensive form is now complete. Observe that the specification of an extensive form of a game corresponds to laying down game rules as done above for propositional logic. We still must tell what a (winning) strategy is, in order to be able to define (or characterize) the satisfaction relation of first-order logic. This time we are in a position to do it more explicitly than above.

A strategy for *Eloise* is any function $f : P^{-1}(\{Eloise\}) \longrightarrow \mathcal{A}$ such that if $P(h) = Eloise$, then $h^\frown f(h) \in H$. In other words, a strategy for *Eloise* yields exactly one choice—in compliance with the game rules—for any history at which *Eloise* is supposed to move. It is clear that there always exist strategies for *Eloise*.[11] A strategy f for *Eloise* is a *winning strategy* (w.s.), if there exists a set $W \subseteq H$ such that all of the following conditions hold: (a) $(A, \gamma) \in W$; (b) whenever $h \in W$ and $P(h) = Abelard$, then any $h^\frown(A_i, \gamma_i)$ which belongs to H also belongs to W; (c) W is closed under applications of f, i.e., if $h \in W$ and $P(h) = Eloise$, then $h^\frown f(h) \in W$; and (d) all terminal histories in W are wins for *Eloise*. The idea behind defining the winning condition of a strategy f in relation to a set W is that W is the set containing exactly those histories that are of relevance for f being a winning strategy. In particular W contains no such histories that cannot be realized given that *Eloise* follows the strategy f. The set W may be referred to as a 'plan of action'.[12]

The above definition of a w.s. incorporates one possible exact formulation of the requirement that a w.s. of a player must yield a win against every sequence of moves made by the opponent. The notion of (winning) strategy for *Abelard* can be defined analogously. The satisfaction relation for first-order logic is now definable as follows:

- A is satisfied in $\mathcal{M}$ under γ, in symbols $(\mathcal{M}, \gamma) \models^{+}_{GTS} A$, if there exists a w.s. for *Eloise* in $G(A, \mathcal{M}, \gamma)$.

- A is dissatisfied in $\mathcal{M}$ under γ, in symbols $(\mathcal{M}, \gamma) \models^{-}_{GTS} A$, if there exists a w.s. for *Abelard* in $G(A, \mathcal{M}, \gamma)$.

For sentences A—i.e., in the case that γ is empty—we speak of truth (falsity) instead of satisfaction (dissatisfaction). It follows from the results of von Neumann and Morgenstern (1944) that every two-player zero-sum game of

[11] Supposing the domain is a set, as opposed to a proper class—such as the universe V of set theory. Evaluated relative to V, there are for instance no strategies for *Eloise* in the game for $\forall x \exists y(x = y)$: no function $V \longrightarrow V$ is a set.

[12] The idea of defining winning strategies by reference to such sets W was proposed by G. Sandu (personal communication), in his capacity of a Ph.D. supervisor of Tero Tulenheimo.

perfect information with payoffs in the set $\{1, -1\}$ is *determined*, i.e., such that one of the players has a w.s. in the game.[13] That is, we have:

$$(\mathcal{M}, \gamma) \models^{-}_{GTS} A \quad \text{if and only if} \quad (\mathcal{M}, \gamma) \not\models^{+}_{GTS} A.$$

8.3.3 Skolem functions

It is useful to observe the connection between *Eloise's* winning strategies in semantic games $G(A, \mathcal{M}, \gamma)$ on the one hand, and *Skolem functions* of a first-order formula on the other.[14] If A is a first-order formula of vocabulary τ in negation normal form, and such that no two occurrences of an existential quantifier in it are associated with the same variable, its *Skolem form* is a second-order $\Sigma^1_1(\tau)$-formula[15] obtained as follows:

1. For each existential quantifier $\exists x_i$ of A, introduce a second-order function variable, f_i. Replace all occurrences of the first-order variable x_i in A by the term $f_i(t_1, \ldots, t_{n_i})$, where the t_j are first-order variables which are obtained from those universal quantifiers $\forall t_j$ to which $\exists x_i$ is syntactically subordinate in A. From the result of carrying out these replacements for all i, erase all occurrences of existential quantifiers. Write A_{Sk} for the resulting formula.

2. If the existential quantifiers of A are $\exists x_1, \ldots, \exists x_n$, define the Skolem form of A to be the $\Sigma^1_1(\tau)$-formula $SK(A) := \exists f_1 \ldots \exists f_n A_{Sk}$.

If $f_1, \ldots, f_k$ are the function symbols appearing in A_{Sk}, and $F_1, \ldots, F_k$ are functions that interpret these function symbols so as to satisfy the formula A_{Sk} in a τ-structure $\mathcal{M}$, then these functions are termed *Skolem functions*, and the sequence $(F_1, \ldots, F_k)$ is called a *full array of Skolem functions* for the formula A relative to the structure $\mathcal{M}$.

The syntactic requirement we imposed on A above, to the effect that in A no two occurrences of an existential quantifier may be associated with the same variable, is by no means necessary; it was only assumed to simplify the presentation. On the other hand the requirement is no restriction from the viewpoint of expressivity, since every first-order formula is logically equivalent to a formula of that form. Not even the requirement that A be in negation normal form is necessary—it could be avoided by making a distinction between positive

[13] This statement is usually taken to be a consequence of the so-called 'Zermelo's theorem'. However, the theorem according to which all zero-sum games of perfect information are determined is not due to Zermelo. For this piece of history, see Schwalbe and Walker (2001).

[14] What came to be known as Skolem functions were introduced in Skolem (1920).

[15] '$\Sigma^1_1(\tau)$' denotes the fragment of second-order logic of vocabulary τ, whose formulas are (logically equivalent to formulas) of the form $\exists X_1 \ldots \exists X_n A$ or $\exists f_1 \ldots \exists f_n B$, where A is a first-order formula of the vocabulary $\tau \cup \{X_1, \ldots, X_n\}$ and B is a first-order formula of the vocabulary $\tau \cup \{f_1, \ldots, f_n\}$, the X_i being relation symbols of any arity, and the f_i function symbols of any arity.

and negative occurrences of quantifiers, based on whether a quantifier appears subordinate to an even or an odd number of negation signs.

Assuming the *Axiom of Choice*, it is possible to prove that if A is such a first-order formula of vocabulary τ for which $SK(A)$ is defined, then for all τ-structures $\mathcal{M}$ and all assignments γ:

$$(\mathcal{M}, \gamma) \models A \quad \text{if and only if} \quad (\mathcal{M}, \gamma) \models SK(A),$$

where for second-order logic, its standard interpretation in the sense of Henkin (1950) is used.

8.4 Characterizing truth

The main issue to be discussed in this paper is how to build a bridge between GTS and dialogical logic from the viewpoint of characterizing validity (in propositional logic, as well as in first-order logic). However, let us first discuss from the dialogical point of view the question of characterizing truth of a sentence (or, more generally, satisfiability of a formula) relative to a model.

There are two rather straightforward approaches one can assume; they give rise to what are known as 'alethic' and 'material' dialogues (see, e.g., Rahman and Keiff, 2005). As dialogues are designed for dealing with validity, some additional ingredient must be introduced into dialogues in order to make them capable of dealing with material truth. *Alethic dialogues* are simply obtained by relativizing a dialogue to a model. Hence a part of the specification of an alethic dialogue in the case of propositional logic will be a valuation function, and in the case of first-order logic a τ-structure for an appropriate vocabulary τ.

By contrast, the idea behind *material dialogues* is to avoid having an extra component to dialogues (such as a specification of a model); they are meant to do with the resources of dialogues designed for dealing with validity, and the idea is to 'approximate' a characterization of truth by adding a sufficient amount of *additional hypotheses*—taken to be *initial concessions* of *Opponent*—which will serve to specify a model by using the resources of the object language only.

What is a sufficient amount, then? In the case of propositional logic, when discussing the truth of a formula A, any relevant model can indeed be specified in terms of **PL**-formulas, namely literals: atomic formulas or their negations. What is more, it suffices to specify a *finite* number of such literals. The relevant models are identified by going through all propositional atoms p_i *appearing in* A (there are only finitely many of these atoms) and choosing, for all i, either p_i itself or its negation $\neg p_i$. In this way any relevant model—any truth-value distribution on the relevant atoms—can be specified.

For first-order logic, this approach has the obvious downside that in general there is no way of capturing a τ-structure in terms of a finite number of first-order sentences. Take for example a {P}-structure $\mathcal{M}$ with an infinite domain

M = $\{d_i : i < \omega\}$, where P is unary. For one thing, to exhaustively describe the interpretation of P in $\mathcal{M}$ in terms of first-order sentences, an infinite list $\langle \ell_i : i < \omega \rangle$ of literals is needed, where $\ell_i := \mathrm{P}c_i$ if $d_i \in \mathrm{P}^{\mathcal{M}}$, and $\ell_i := \neg \mathrm{P}c_i$ if $d_i \notin \mathrm{P}^{\mathcal{M}}$. (By stipulation, the constant c_i stands for the element d_i.) Mathematically there is of course nothing problematic with such infinite lists of hypotheses. But one *desideratum* in designing dialogues typically is that it should be possible to think of them as humanly manageable, ideally temporal processes. Such a process cannot really begin by going through an infinite number of hypotheses. This is why material dialogues with an infinity of hypotheses should be considered as something deeply unsatisfactory.

For another thing—even granting such an 'unrealistic' way of fixing the interpretations of relation symbols—one should find a syntactic way of specifying the domain of the model considered. In particular, the fact that the members of the list $\langle \ell_i : i < \omega \rangle$ jointly mention the infinite set $\{c_0, c_1, \ldots\}$ of individual constants only indicates that in the corresponding model there are individuals (denoted by) $c_0, c_1, \ldots$. In this framework nothing precludes that there are further individuals, not named by any of the constants c_i. As a matter of fact, the straightforward idea of material first-order dialogues—formulated by adding all (negated) atomic sentences true in the relevant model as additional hypotheses of *Opponent*—is doomed to failure. The ultimate reason for this has been pointed out by Hintikka (1987, 251–252). For, the idea could work only if it was possible to obtain all sentences true in a given model $\mathcal{M}$ as logical consequences of the so-called diagram of that model: the set of atomic sentences and negated atomic sentences true in $\mathcal{M}$. However, it is a model-theoretic fact that this is not generally possible—there are models $\mathcal{M}$ and (complex) sentences A true in $\mathcal{M}$ such that A is not derivable from the diagram of $\mathcal{M}$ using any sound and complete proof system for first-order logic, say dialogues.

Below in Section 8.6.1 we will develop a novel approach for defining reasonable material dialogues for first-order logic. For reasons just mentioned, the additional hypotheses of *Opponent* in such dialogues cannot simply be sentences indicating the interpretations of the relevant relation symbols relative to a fixed set of individual constants. A part of the results obtained in Section 8.5 about propositional logic, and in Section 8.6 about first-order logic, is that the appropriate formulations of material dialogues actually capture the notion of truth, i.e., coincide with the usual definition of truth of sentences of these logics.

8.5 Characterizing validity in propositional logic

We will establish an explicit correspondence result between dialogical logic and GTS in characterizing validity in connection with classical propositional logic. The dialogues considered will be intuitionistic dialogues $\mathcal{D}(A)$

with hypotheses (initial concessions). These are simply intuitionistic dialogues about the thesis A, where *Opponent* concedes right from the beginning some finite number of propositions $B_1, \ldots, B_n$. We denote these dialogues as $\mathcal{D}(A; B_1, \ldots, B_n)$. (When the hypotheses are clear from the context, we allow simply writing $\mathcal{D}(A)$ for the dialogue.)

The additional hypotheses of *Opponent* can be thought of as specific material assumptions that *Proponent* is free to make use of and to which *Opponent* remains committed. From the model-theoretic perspective they are naturally viewed as giving rise to a partial specification of a model. In this paper we will consider specified instances of the *law of excluded middle* as hypotheses to which **O** is committed. Specifically, if the thesis is A, we assume that *Opponent* concedes $p \vee \neg p$ for all propositional atoms p appearing in A.

It will be shown how to turn a winning strategy of *Proponent* in an intuitionistic dialogue $\mathcal{D}(A; \mathcal{H})$ with hypotheses,

$$\mathcal{H} = \{p \vee \neg p : p \text{ is an atom appearing in } A\},$$

into a family of winning strategies of *Eloise* in games $G(A, M)$, where M is an arbitrary truth-value distribution; and conversely, we will show how to obtain a winning strategy of *Proponent* in dialogue $\mathcal{D}(A; \mathcal{H})$ from a family of *Eloise*'s winning strategies in games $G(A, M)$. The fact that the correspondence result is of this form has two notable features. Indirectly, it shows that the only classical assumption relevant for GTS in connection with propositional logic is that the models are total in the sense that every propositional atom is either true or false (no propositional indeterminacy). Further, the result provides an explicit method of constructing a winning strategy in one type of game from a family of winning strategies in games of the other type, and *vice versa*.

Before moving to the general proof, let us take an example.

Example 9. Let $A := (\neg p \wedge q) \vee (p \vee \neg q)$. We show how to transform *Proponent*'s winning strategy in the intuitionistic dialogue

$$\mathcal{D}(A; p \vee \neg p, q \vee \neg q)$$

into a family of *Eloise*'s winning strategies in games $G(A, M)$ with $M : \{p, q\} \longrightarrow \{\texttt{true}, \texttt{false}\}$; and *vice versa*.

($\Longrightarrow$) Suppose f is a w.s. for **P** in $\mathcal{D}(A)$. Let us assume that f first makes **P** to pose questions about **O**'s initial concessions:

$$?\text{-}p \vee \neg p \quad \text{and} \quad ?\text{-}q \vee \neg q.$$

We may suppose that f is of such form without loss of generality, because **P** might not finish any play of the dialogue without receiving an answer to at least one such question, and by assumption **P** is able to finish all relevant plays. Combinatorially, there are four possible pairs of answers by **O**:

$$(p, q) \text{ or } (\neg p, q) \text{ or } (p, \neg q) \text{ or } (\neg p, \neg q).$$

Since f is a winning strategy, the continuation of the dialogue must be as follows. In the first, third and fourth case, the strategy f tells to pick out the right disjunct of A (i.e., $p \vee \neg q$), whereas in the second case it makes **P** to choose the left disjunct (that is, $\neg p \wedge q$).

In the first-mentioned cases f tells **P** to go on by picking a disjunct whose truth **O** already had conceded: p in the first case, either p or $\neg q$ in the third case, and $\neg q$ in the fourth case. By contrast, supposing **O** conceded $(\neg p, q)$, **P** can reply by the relevant conjunct to either of the possible questions by **O** concerning $\neg p \wedge q$.

Let us now see how such a w.s. of **P** transforms into a family of winning strategies of *Eloise*, one strategy for each game $G(A, M)$ with $M : \{p, q\} \longrightarrow \{\texttt{true}, \texttt{false}\}$. Let M be an arbitrary truth-value distribution $\{p, q\} \longrightarrow \{\texttt{true}, \texttt{false}\}$. M determines a particular set of answers by **O** to **P**'s questions about **O**'s initial concessions in $\mathcal{D}(A)$:

> **O** responds to ?-$p \vee \neg p$ by `left`, if $M(p) = \texttt{true}$, and
> by `right`, if $M(p) = \texttt{false}$;
> **O** responds to ?-$q \vee \neg q$ by `left`, if $M(q) = \texttt{true}$, and
> by `right`, if $M(q) = \texttt{false}$.

We define a strategy g for *Eloise* making use of **P**'s winning strategy f in dialogue $\mathcal{D}(A)$. Because f is a w.s., it tells what to do in particular in the case that **O** has 'specified the model M' with his answers.

For the outmost disjunction in $(\neg p \wedge q) \vee (p \vee \neg q)$, let g yield the same choice (`left` or `right`) that f yields as a response to the corresponding question ?-$\vee$ asked by **O**. Hence the choice provided by g is `left` precisely when the model M makes p false and q true (since this is when f yields the choice `left`). In that case there are no more moves in the play of $G(A, M)$ and *Eloise* wins the play. If *Eloise* chooses `right` for the outmost disjunction, let g make her choose as follows for the inner disjunction: `left` if f tells to choose `left` when asked about that disjunction, and `right` otherwise. Clearly this too leads to a play won by *Eloise*. We may conclude, then, that g is a w.s. for *Eloise* in $G(A, M)$.

($\Longleftarrow$) Suppose that for every model $M : \{p, q\} \longrightarrow \{\texttt{true}, \texttt{false}\}$, there is a w.s. (say g_M) for *Eloise* in the corresponding game $G(A, M)$. This means that the strategy g_M tells to choose `left` for the outmost disjunction of $(\neg p \wedge q) \vee (p \vee \neg q)$, if $M(p) = \texttt{false}$ and $M(q) = \texttt{true}$, and otherwise makes *Eloise* to choose `right`. In the right disjunct, g_M then goes on to yield the choice `left`, if $M(p) = \texttt{true}$, and the choice `right`, if $M(q) = \texttt{false}$. (If both $M(p) = \texttt{true}$ and $M(q) = \texttt{false}$, then g_M yields one of the two choices.)

Let us define a strategy f for **P** in dialogue $\mathcal{D}(A)$ as follows. First f makes **O** to answer both questions, ?-$p \vee \neg p$ and ?-$q \vee \neg q$. These answers determine a truth-value distribution $M : \{p, q\} \longrightarrow \{\texttt{true}, \texttt{false}\}$, and in particular for

this model M, *Eloise* has—by assumption—a winning strategy (g_M) in the semantic game $G(A, M)$. We use g_M in defining how to continue in dialogue $\mathcal{D}(A)$. Let f yield the same choice `left` for the outmost disjunct of A that g_M yields for that disjunction. In the case of choosing first `right`, let f further provide the same choice for the inner disjunction that is provided by g_M to the inner disjunction. Obviously f then always leads to a choice of a (positive or negative) literal already conceded by **O**. That is, f is a winning strategy for **P** in dialogue $\mathcal{D}(A)$.

8.5.1 The correspondence result

Let us observe the following fact about tokens of subformulas B of a propositional formula A, and choices for disjunctions and conjunctions to which B is syntactically subordinate in A.

Observation 10. *In the case of propositional logic, tokens of subformulas B of a given formula A unambiguously reveal the choices for conjunctions and disjunctions made in order to arrive at B in a semantic game. For instance, the innermost token of 'p' in the formula* $(p \vee (p \wedge q))$ *is identified by the choices* `right` *and* `left` *for* $\vee$ *and* $\wedge$*, respectively.*

On the one hand, in a dialogue one and the same token of a connective may have been visited by Challenger *several times. On the other hand, any subformula token B reveals* only the most recent *left/right choices for the conjunctions and disjunctions that syntactically precede B, made by* **P** *and* **O** *in a dialogue about the thesis A. Obviously for the continuation of a propositional dialogue only the subformula reached matters: only those most recent choices matter.*[16]

We now move on to state and prove a theorem linking GTS and dialogical logic in view of characterization of validity in propositional logic.

Theorem 11. *Let A be any formula of propositional logic. The following conditions are equivalent:*

(*i*) *There is a w.s. for* Proponent *in* $\mathcal{D}(A; p_1 \vee \neg p_1, \ldots, p_n \vee \neg p_n)$*;*

(*ii*) *For every M, there is a w.s. for* Eloise *in* $G(A, M)$*;*

where $p_1, \ldots, p_n$ *are the propositional atoms appearing in A. Furthermore, there is an algorithm turning* Proponent*'s winning strategy into a family of* Eloise*'s winning strategies, and* vice versa*.*

We prove Theorem 11 in two steps, by first establishing Lemma 12 and then Lemma 14.

[16]The situation is more complicated in the case of first-order logic; Section 8.6.2 is devoted to discussing this issue.

Lemma 12. *Condition* (*ii*) *of Theorem 11 implies its condition* (*i*).

Proof. Let $p_1, \ldots, p_n$ be the propositional atoms that actually appear in A. Consider the 2^n different truth-value distributions (models)

$$M_j : \{p_1, \ldots, p_n\} \longrightarrow \{\mathtt{true}, \mathtt{false}\} \qquad (j := 1, \ldots, 2^n).$$

Suppose that for every M_j, there is a w.s. for *Eloise* in $G(A, M_j)$. We must show that there is a w.s. for **P** in the intuitionistic dialogue $\mathcal{D}(A; p_1 \vee \neg p_1, \ldots, p_n \vee \neg p_n)$. **O**'s initial concessions are

$$p_i \vee \neg p_i \qquad (i := 1, \ldots, n).$$

We describe a strategy f of **P** in $\mathcal{D}(A)$.

(i) To begin with, f tells **P** to pose, for all i, the question

$$?\text{-}p_i \vee \neg p_i.$$

O's answers to these questions will, then, determine a truth-value distribution $M : \{p_1, \ldots, p_n\} \longrightarrow \{\mathtt{true}, \mathtt{false}\}$. By assumption there is a w.s. for *Eloise* in $G(A, M)$, call it g. We go on to construct a continuation of **P**'s strategy f in $\mathcal{D}(A)$ using the strategy g. Recall that the dialogue $\mathcal{D}(A)$ will proceed intuitionistically.

(ii) We will first correlate recursively every subformula token B of A with a play of the semantic game $G(A, M)$. The thesis A is associated with the empty sequence $\langle\rangle$. Suppose, then, that a subformula token B is already associated with a play h. We will write $\sharp[B]$ for the number of negation signs to which B is subordinate in A.

- If $B = C \wedge D$ and $\sharp[B]$ is even or equal to zero, or $B = C \vee D$ and $\sharp[B]$ is odd, then C is associated with $h^\frown C$ and D with $h^\frown D$.
- If $B = C \vee D$ and $\sharp[B]$ is even or equal to zero, or $B = C \wedge D$ and $\sharp[B]$ is odd, and $g(h) = C$, then C is associated with $h^\frown g(h)$; whereas if $g(h) = D$, then it is D that is associated with $h^\frown g(h)$.
- If $B = \neg C$, then C is associated with h.

Define now f so that if **O** asks **P** to choose a disjunct of $C \vee D$ (the disjunction sign hence appearing under an even number of negation signs), and the history associated with $C \vee D$ is h, then **P** chooses the uniquely determined disjunct $E \in \{C, D\}$ such that the disjunct in question is associated with a history—and hence by construction with a history of the form $h^\frown g(h)$, where $g(h) = E$. (Observe that in order to have arrived at the very subformula $C \vee D$, **P** must have made exactly those choices for

the connectives for which she has moved that are encoded in the history h.) Similarly, if **O** asks to choose a conjunct of $C \wedge D$ (the conjunction sign appearing under an odd number of negation signs), let f make **P** choose the uniquely determined conjunct with an associated history; such a history is then of the form $h^\frown g(h)$. In short, then, we define f so that it 'respects associated histories'.

Claim 13. *The strategy f is a winning strategy for* **P** *in* $\mathcal{D}(A; p_1 \vee \neg p_1, \ldots, p_n \vee \neg p_n)$.

Proof. Due to its definition, f leads to a literal ℓ (or a conjunction of literals $\ell_1, \ldots, \ell_l$) true in M. This means that **O** has conceded p (if $\ell = p$), or has conceded $\neg p$ if $\ell = \neg p$. In the former case **P** is indeed in a position to reply p (**O** has already conceded it); in the latter case **O** may only challenge $\neg p$ by contradicting himself. □

Let $\mathcal{G} = \{g_j : j < 2^n\}$ be a family of *Eloise*'s winning strategies, one for each model $M_j : \{p_1, \ldots, p_n\} \longrightarrow \{\texttt{true}, \texttt{false}\}$. The algorithm for generating a winning strategy for **P** in $\mathcal{D}(A; p_1 \vee \neg p_1, \ldots, p_n \vee \neg p_n)$ consists, then, in first making **O** to determine a model M_j by his answers to questions ?-$p_i \vee \neg p_i$. The w.s. g_j of *Eloise* in $G(A, M_j)$ then determines a labeling of subformulas by plays of $G(A, M_j)$ as explained above. This, in turn, directly defines a winning strategy for **P**, proceeding from **O**'s specific answers to **P**'s questions about **O**'s initial concessions, as was shown above. □

It is important to observe the following fact about the definition of **P**'s strategy f in the above proof. Actually, if the formula A considered is intuitionistically valid, then from the point of view of dialogic there is no need to use concessions of the form $p \vee \neg p$. However, for the sake of our aims, it is convenient to assume that **O** explicitly makes concessions of that form, and that **P** always starts by asking questions about these concessions. It is interesting to note that the resulting games are of a kind that Kuno Lorenz has called *strenge Dialogspiele*: games where not only defenses, but also attacks can be performed *only once* (cf. Lorenzen and Lorenz, 1978, 120–126). Dialogues of this type are related to the representation of certain connectives of linear logic and could even be considered as having anticipated them. Furthermore, part of the critique against the intuitionistic dialogic that Andreas Blass puts forward on the first pages of his beautiful 1992 paper is pointing out the asymmetry between conjunction and disjunction. Now if one implements an algorithm which stipulates to attack first the appropriate instances of the law of excluded middle, the alleged asymmetry will disappear: no question whether *Proponent* is entitled to concede an atom or not will arise later in the same dialogical game.

Lemma 14. *Condition* (*i*) *of Theorem 11 implies its condition* (*ii*).

Proof. Suppose **P** has a w.s. in the dialogue

$$\mathcal{D}(A; p_1 \vee \neg p_1, \ldots, p_n \vee \neg p_n),$$

and call it f. Such a strategy need not consist of first asking **O** about all hypotheses, but we may without loss of generality assume that f is such a strategy. Let $M : \{p_1, \ldots, p_n\} \longrightarrow \{\texttt{true}, \texttt{false}\}$ be an arbitrary model. We must show that *Eloise* has a w.s. in $G(A, M)$. First consider a dialogical game belonging to $\mathcal{D}(A)$, where **P** has received such answers to her questions about **O**'s initial concessions that these constitute precisely the model M: **O** has replied p_i if $M(p_i) = \texttt{true}$, and otherwise has replied $\neg p_i$.

Let us associate with every subformula B of A for which it is **P** who makes a move in $\mathcal{D}(A)$, a subformula E of B, as follows.

> If $B = C \vee D$ and $\sharp[B]$ is even or equal to zero, or $B = C \wedge D$ and $\sharp[B]$ is odd, and f gives $E \in \{C, D\}$ as a response to the question ?-$\vee$ (*resp.* as the choice of a conjunct), let B be associated with E.

Define a strategy g for *Eloise* in the semantic game $G(A, M)$ by putting

$$g(h) := f(B),$$

if B is the subformula of the form $C \vee D$ (under an even number of negation signs, or none) or of the form $C \wedge D$ (under an odd number of negation signs) corresponding to which *Eloise* must make a move at the history h. By Observation 10 we may assume that the value of the strategy f only depends on the subformula token B.

> **Claim 15.** *The strategy g is a w.s. for* Eloise *in* $G(A, M)$.
>
> *Proof.* The last move by *Eloise* made in accordance with g leads to a literal ℓ (or a conjunction of literals). By the definition of g, if $\ell = p$, then p is true in M (since in that case **O** must have conceded it before). If, on the other hand, $\ell = \neg p$, then p is false in M, for then **O** must contradict himself by conceding p. □

Let f be **P**'s w.s. in $\mathcal{D}(A)$. The algorithm for generating a family $\mathcal{F}$ of *Eloise*'s winning strategies in games $G(A, M)$, one for each model

$M : \{p_1, \ldots, p_n\} \longrightarrow \{\texttt{true}, \texttt{false}\}$, consists of transforming f into such a winning strategy f_0 which first asks **O** to reply to each of the n questions of the form ?-$p_i \vee \neg p_i$. **O**'s answers then determine a model M, relative to which f_0 further determines, as explained above, a labeling which defines a winning strategy for *Eloise* in the semantic game $G(A, M)$. □

8.6 Characterizing validity in first-order logic

Before moving on to formulate and prove the correspondence result between dialogical logic and GTS in characterizing validity in connection with classical first-order logic, we will discuss conceptual issues that are forced upon us by the dialogical framework. In particular, we think of dialogues as intrinsically 'finitist' processes; no dialogical game for instance may involve going through an actual infinity of attacks and defenses. On the other hand, for our correspondence result we need a way of representing τ-structures on the level of dialogues. How are we, then, to deal with the fact that such structures can have an infinite domain?

8.6.1 The framework

It was observed that in the case of propositional logic, a finite list of initial concessions $p_i \vee \neg p_i$ can be produced simply out of the propositional atoms $p_1, \ldots, p_n$ appearing in the propositional formula considered. It was noted that hence it is possible to determine any model $M : \{p_1, \ldots, p_n\} \longrightarrow \{\texttt{true}, \texttt{false}\}$ of propositional logic by a set of answers by means of which *Opponent* can defend himself against *Proponent*'s attacks on these initial concessions; and conversely, any such set of answers determines a model $M : \{p_1, \ldots, p_n\} \longrightarrow \{\texttt{true}, \texttt{false}\}$. On the other hand, it was seen in the beginning of Section 8.4 that in the case of first-order logic, there is no straightforward way in which to represent first-order structures by finitary means in dialogues. Is it, however, possible to find a new type of concession, so that a finite set of *Opponent*'s concessions of such a type could, after all, be used for specifying the relevant first-order structures?

First note that in order to identify a τ-structure, there are two things to determine: (1) its domain, and (2) the interpretations of the relation symbols of the vocabulary τ. To formulate a connection between dialogues and semantic games in characterizing validity in first-order logic, a prerequisite is to be able to reconstruct τ-structures within the dialogical framework. Any given semantic game is played on a fixed τ-structure, and we must find a way of saying to what having fixed such a τ-structure corresponds in a dialogue. To this end, then, we must be able to reconstruct the notion of domain, as well as to determine interpretations of relation symbols on such domains, using dialogues—despite the problems observed.

Next recall that τ-structures can be taken to be of countable size (cf. Section 8.3.2). This is well suited from the proof-theoretic perspective, as in proof theory we deal with at most countably many values of variables. Indeed, the individuals of a proof theorist are defined in terms of linguistic symbols (individual constants), and in the dialogical framework it surely would be out of purpose to have more than countably many such symbols. As vehicles of proof theory, dialogues must not make reference to models at all. Proof theory is *formal*—it must be possible to carry out proofs on the *syntactic* level. (This is so notwithstanding any philosophical arguments to the effect that proofs or inference rules are constitutive of meaning and in this sense serve to link language to what the language is about.)

We will proceed as follows. Let us fix a countably infinite stock $\{c_0, c_1, \ldots\}$ of individual constants.[17] We will be interested in arbitrary countable models, wherefore a finite number of constants would not suffice as syntactic substitutes for individuals. On the other hand, since we are, *inter alia*, interested in finite models—in addition to countably infinite ones—we cannot simply take the individual constants $c_0, c_1, \ldots$ themselves as representing individuals; otherwise all models that we would manage to syntactically represent would have $\{c_0, c_1, \ldots\}$ as their infinite domain. Instead, we will include in the dialogues a mechanism that will mimic the semantic phenomenon of several individual constants standing for the same object.

Representing domains. We include among *Opponent*'s initial concessions in all dialogues that we will consider the following sentence:

$$\forall x_1 \forall x_2 (x_1 = x_2 \vee x_1 \neq x_2).$$

Here the symbol '=' is syntactically subject to equivalence relation axioms (reflexivity, symmetry, transitivity), together with a substitution rule saying that if **O** has conceded both $c_i = c_j$ and $A[x/c_i]$, then **P** may ask **O** to concede $A[x/c_j]$; the only available defense for **O** being to concede indeed $A[x/c_j]$.[18]

Given that the constants to be substituted for quantified variables will be taken from the set $\{c_0, c_1, \ldots\}$, what type of question concerning the sentence $\forall x_1 \forall x_2 (x_1 = x_2 \vee x_1 \neq x_2)$ should we pose in order for the answer to identify a domain—which can be either finite or countably infinite? Simply putting forward a question ?-$\forall x_1/c_i$, followed by a question ?-$\forall x_2/c_j$, followed by a question ?-$(c_1 = c_2 \vee c_1 \neq c_2)$ would not do; this maneuver would lead to identifying a domain in terms of constants c_i (intuitively standing for its elements) only if repeated *infinitely* many times. But starting a dialogue with

[17] Technically, fixing the set of available individual constants in this way may be construed as a specific structural rule.

[18] The equivalence relation axioms can be construed either as metalogical rules or else, simply, as further initial concessions by **O**. The substitution rule is a structural rule of its own.

going through an infinity of hypotheses to get the model right would hardly be in keeping with the idea of a dialogue as a (humanly manageable, ideally temporal) process, as already noted.

Clearly, we should ask, once and for all, *Opponent* to choose a *Skolem function* for the disjunction symbol of the sentence

$$\forall x_1 \forall x_2 (x_1 = x_2 \vee x_1 \neq x_2).$$

A Skolem function

$$f : \{c_0, c_1, \ldots\} \times \{c_0, c_1, \ldots\} \longrightarrow \{\texttt{left}, \texttt{right}\}$$

expressly states, for each possible pair of values (c_i, c_j) chosen for the pair of variables (x_1, x_2), whether *Opponent* considers c_i and c_j to be proof-theoretically interchangable or not. Intuitively, then, if *Opponent* chooses the left disjunct, c_i and c_j 'stand for the same object', whereas if he chooses the right disjunct, c_i and c_j are taken to 'stand for distinct objects'.

Technically, what *Opponent*'s choice of a Skolem function f accomplishes is to induce an equivalence relation $\sim_f$ among pairs of constants from the set $\{c_0, c_1, \ldots\}$: we have $c_i \sim_f c_j$ if and only if $f(c_i, c_j) = \texttt{left}$. Now the index of the equivalence relation $\sim_f$ will be precisely the cardinality of the domain represented by the Skolem function f.[19] Clearly any countable cardinality can be represented by a suitable choice of f. The domain corresponding to the choice of f is simply the quotient set $\{c_0, c_1, \ldots\}/\sim_f$. The equivalence classes of the set $\{c_0, c_1, \ldots\}$ are then representatives of model-theoretic individuals.

We must introduce a rule allowing to pose such 'second-order' questions, with a Skolem function as the response. Before doing so, let us consider how to determine interpretations of the relation symbols of a given vocabulary.

Representing interpretations. For a finite vocabulary τ, among *Opponent*'s initial concessions will be included the sentences

$$\forall x_1 \ldots \forall x_n (\mathrm{R} x_1 \ldots x_n \vee \neg \mathrm{R} x_1 \ldots x_n),$$

one sentence for each $\mathrm{R} \in \tau$ of arity n. Again, we must allow *Proponent* to ask, once and for all, *Opponent* to choose a Skolem function for the disjunction symbol of such a sentence. For instance, a Skolem function

$$f : \{c_0, c_1, \ldots\} \longrightarrow \{\texttt{left}, \texttt{right}\}$$

for the disjunction symbol in the sentence $\forall x (\mathrm{P} x \vee \neg \mathrm{P} x)$ expressly indicates, for each possible constant c_i that can be substituted for x, which of the disjuncts

[19] The *index* of an equivalence relation $\sim$ on a set X is by definition the cardinality of the quotient set $X/\sim$, i.e., the cardinality of the set of all equivalence classes of X under the relation $\sim$.

Opponent considers being satisfied by c_i. This is exactly what it means to specify the extension of the predicate P relative to a domain whose individuals are all named by at least one constant in the set $\{c_0, c_1, \ldots\}$.

Observe that if *Opponent* chooses his responses at random, the choice of a Skolem function f for, say, a unary relation symbol P can contradict his choice of a Skolem function i corresponding to the identity sign. That is, it can happen that $i(c_i, c_j) = \texttt{left}$, while $f(c_i) \neq f(c_j)$. Hence c_i and c_j intuitively stand for the same object, but yet only one of them serves to satisfy the predicate P. However, evidently *Opponent* can always make his choices of Skolem functions for the additional hypotheses in a coherent way—so that the choice of the function corresponding to the identity sign does not lead to a contradiction in view of his choices of functions corresponding to relation symbols. We will subsequently always assume that *Opponent* indeed makes such coherent choices.

Let us then move on to introduce a new mode of question, which enables to ask about a Skolem function for an operator. In what follows, we sometimes use a barred x, i.e., $\bar{x}$, to stand for a finite sequence of variables, $x_1 \ldots x_n$, and $\forall\bar{x}$ to stand for the block $\forall x_1 \ldots \forall x_n$. When asked about a sentence of the form $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$ with $\mathrm{E} \in \tau \cup \{=\}$, the question

??-$\vee$

must be answered by providing a second-order object, namely a Skolem function $f : \{c_0, c_1, \ldots\}^n \longrightarrow \{\texttt{left}, \texttt{right}\}$ for the unique token of the disjunction symbol $\vee$ appearing in $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$. (We write the '?' two times to indicate that the answer should be a second-order object.) Suppose **O** asserts that f is such a function. Then if E is the identity symbol, a domain is thereby determined (precisely the constants c_i, c_j with $c_i \sim_f c_j$ will be thought of as standing for the same object), while if E is a relation symbol from τ, the function f specifies which atomic sentences involving E are taken to be true (in the sense of being conceded by **O**) and which false (in the sense that their negations are conceded by **O**). The concession concerning the identity sign might be contradictory with the concessions concerning the relation symbols, but as just noted we may always assume that the concessions are in effect mutually coherent. Then indeed, a model can be extracted from **O**'s replies to **P**'s questions ??-$\vee$ about his initial concessions $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$. If **O**'s reply to the question about the identity sign was i, then his replies to the questions about relation symbols will serve to specify the interpretations of these relation symbols on the domain determined by the Skolem function i. For instance, if f is **O**'s reply to a question concerning a unary relation symbol P, then P is taken to be satisfied by precisely those equivalence classes ξ determined by the Skolem function i for which $f(c) = \texttt{left}$ for some (and hence all) $c \in \xi$.

Once **O** has laid down his choices of Skolem functions, **P** can draw all kinds of inferences from them. For instance she can check whether a relation symbol

R is satisfied by at least one tuple, by seeing whether `left` lies in the image of the Skolem function corresponding to R.[20]

The questions ??-$\vee$ give rise to the following new structural rule:

(SR-7) (Skolem function rule). If **O** has conceded that f is a Skolem function for $\vee$ in $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$, then **O** must also, if asked, concede all instances of this second-order concession. That is, for any tuple $\bar{c}$ interpreting the variables $\bar{x}$, he must concede $\mathrm{E}\bar{c}$, if $f(\bar{c})$ = `left`, and $\neg\mathrm{E}\bar{c}$, if $f(\bar{c})$ = `right`. Accordingly, once **O** has replied by some f to a question ??-$\vee$, **P** is always entitled to pose the question ?-$f/\bar{c}$, for any tuple $\bar{c}$, asking **O** to confirm that indeed he concedes that the tuple $\bar{c}$ satisfies the disjunct $f(\bar{c})$. **O** has no real choice for his answer: the reply is fully predetermined by his choice of f and the requirement that **O** must be coherent in his replies.[21]

Remark 16. *What we phrased above in terms of Skolem functions may look like an ascent from the essentially syntactic approach of proof theory to a more model-theoretic approach, where we speak of objects or sets of objects serving as denotations for such linguistic items as individual constants or relation symbols. For, in the above approach it is not sufficient that* **O** *choose a function symbol,* f. *In addition, he must commit himself to an infinity of identities of the form*

$$\mathrm{f}(c_{i_1}, \ldots, c_{i_n}) = b$$

where $b \in \{$`left`, `right`$\}$ *and the* c_{i_j} *are arbitrary individual constants. When saying that* **O** *chooses a Skolem function we hence mean that he makes at one stroke a potentially infinite number of concessions of syntactic nature. The concession is a 'second-order concession', but still an essentially syntactic or formal one.*

Furthermore, the language we are dealing with remains a first-order language, where all quantification is over individuals only. In particular, the language $\mathcal{L}[\tau]$ *itself does not involve quantification over functions. Having available the possibility of asking for Skolem functions is simply an addition to the* repertoire of questions that can be posed about concessions *concerning formulas of the relevant first-order language. It does not add to what the formulas themselves state.*

Are such second-order questions as ??-$\vee$ not quite far-fetched from the dialogical viewpoint, anyway? There are two things to observe by way of

[20] If $f : A \longrightarrow B$ is a function, we write $Im(f)$ for the *image* of f, that is, for the set $\{f(x) : x \in A\}$. Hence $Im(f)$ is a (proper or improper) subset of B.

[21] Skolem functions could of course appear in dialogues more generally than as Skolem functions of disjunctions appearing in contexts like $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$. The name "Skolem function rule" is hence undeservedly general. However, in the present paper no further use of Skolem functions is made, and we stay with the terminology.

answering this type of criticism. First, the reason why we are introducing this device into dialogues in the first place is not motivated by the dialogues themselves, but our desire to prove a first-order analogue to Theorem 11. That is, we wish to establish an explicit bidirectional link between dialogues on the one hand, and semantic games on the other, in characterizing the validity of first-order sentences. Given this background of generalizing Theorem 11, the most important thing to take care of is that the only non-intuitionistic inputs to dialogues come from the additional initial concessions of **O**. We *must* be able to reconstruct τ-structures on the level of dialogues, and we must be able to do it by asking only finitely many questions. Asking about Skolem functions for finitely many disjunctions does the job, while asking about an infinity of individual constants whether they serve to satisfy a given predicate does not.

Secondly, ironically perhaps, the question form ??-$\vee$ is, arguably, more in the spirit of intuitionism than first meets the eye—and therefore particularly suitable in the dialogical approach, which was historically motivated precisely by intuitionism. Namely, for an intuitionist, committing oneself to the truth of a sentence $\forall x(\mathrm{P}x \vee \neg \mathrm{P}x)$, say, necessarily involves committing oneself to a method of verifying the sentence. An intuitionist takes $\forall x(\mathrm{P}x \vee \neg \mathrm{P}x)$ as true only if he not only knows it to be true (there is the epistemic dimension to the notion of truth for an intuitionist), but also knows *why* it is true. In less pictorial language, he must be in a possession of a function which for every value of x chooses either the left or the right disjunct, according to whether the value does or does not satisfy the predicate P. From this viewpoint, one could even expect that intuitionistically, the questions about the concessions of a player would *need* to be responded, first and foremost, by listing Skolem functions of the operators appearing in the conceded sentence. We will not require this in general in what follows, but take it that there are good enough reasons not to consider our framework, in its intended context, as *ad hoc* (cf. here Hintikka, 1996, Chapter 11).

Example 17 (With implication in the language)**.** *Consider the intuitionistic dialogue for the first-order sentence* $\exists x(\mathrm{Q}x \to \forall y \mathrm{Q}y)$, *with the additional initial concessions* $\forall x_1 \forall x_2(x_1 = x_2 \vee x_1 \neq x_2)$ *and* $\forall x(\mathrm{Q}x \vee \neg \mathrm{Q}x)$ *of* **O**. *That is,* **O** *concedes that the interpretations of the symbols* '=' *and* 'Q' *respect the* law of excluded middle. *The following is a description of* **P***'s winning strategy in such a dialogue. First,* **P** *asks* **O** *to choose Skolem functions for the disjunction symbols in the extra hypotheses. Suppose* **O** *chooses, respectively, functions*

$$f : \{c_0, c_1, \ldots\} \times \{c_0, c_1, \ldots\} \longrightarrow \{\texttt{left}, \texttt{right}\},$$

and

$$g : \{c_0, c_1, \ldots\} \longrightarrow \{\texttt{left}, \texttt{right}\}.$$

If **O***'s choices involve a contradiction, i.e., if there are* c_i *and* c_j *such that* $f(c_i, c_j) = \texttt{left}$ *but* $g(c_i) \neq g(c_j)$, **P** *may proceed as follows. By asking the*

question ?-f/c_1c_j *about the concession concerning identity, and by asking the questions* ?-g/c_i *and* ?-g/c_j *about the concession involving the relation symbol* Q, **P** *will force* **O** *to concede* $c_i = c_j$ *together with* $\mathrm{Q}c_i$ *and* $\neg\mathrm{Q}c_j$ *(or with* $\mathrm{Q}c_j$ *and* $\neg\mathrm{Q}c_i$*). By the relevant substitution rule* **P** *will then further be able to make* **O** *to run into a plain contradiction, conceding both* $\mathrm{Q}c_j$ *and* $\neg\mathrm{Q}c_j$.

Suppose, then, that **O***'s choices of* f *and* g *are free from contradiction, i.e., meet the following condition:*

$$\textit{if } f(c_i, c_j) = \texttt{left}\textit{, then } g(c_i) = g(c_j).$$

By having chosen such f *and* g, **O** *has in effect determined a countable* {Q}*-structure* $\mathcal{M}$ *with the domain* $\{c_0, c_1, \ldots\}/\sim_f$ *and the following interpretation of* Q*:*

$$\xi \in \mathrm{Q}^{\mathcal{M}} \quad \Leftrightarrow_{\textbf{def}} \quad f(c_i) = \texttt{left},$$

where $\xi := \{c_j : c_i \sim_f c_j\}$ *and* c_i *is an arbitrarily chosen representative of* ξ, $c_i \in \xi$*. (This definition of* $\mathrm{Q}^{\mathcal{M}}$ *is independent of the choice of the representative* c_i *of the equivalence class* ξ*, precisely because* f *and* g *are not mutually contradictory.) Then let* **P** *proceed as follows:*

(1) *If* $Im(g) = \{\texttt{left}\}$*, i.e., if* **O** *concedes* $\mathrm{Q}c_i$ *for all constants* c_i*, then let* **P** *choose* c_0 *as a response to* **O***'s question* ?-$\exists x$*.* **O** *then attacks the implication by conceding* $\mathrm{Q}c_0$ *and* **P** *must defend* $\forall y\mathrm{Q}y$*. Supposing* **O** *asks* ?-$\forall x/c_i$*,* **P** *first asks* **O** *to acknowledge that* $\mathrm{Q}c_i$ *(i.e., she asks* ?-f/c_i*), which the latter must do as he already conceded that* g *is a Skolem function for* $\vee$ *in the hypothesis about* Q*. But then* **P** *may reply to* **O***'s question* ?-$\forall x/c_i$ *by* $\mathrm{Q}c_i$*, which yields a finished play won by* **P**.

(2) *If* $\texttt{right} \in Im(g)$*, let i be the smallest i such that* $g(c_i) = \texttt{right}$*, and let* **P** *ask* **O** *to concede* $\neg\mathrm{Q}c_i$*, which the latter must do. Let then* **P** *answer to* **O***'s question* ?-$\exists x$ *by* c_i*.* **O** *may only attack by conceding* $\mathrm{Q}c_i$*, hence contradicting himself. This yields a finished play won by* **P**.

It is not difficult to see how the run of the above dialogue would change if we had, in place of the sentence $\exists x(\mathrm{Q}x \rightarrow \forall y\mathrm{Q}y)$ involving an implication sign, its classical equivalent $\exists x(\neg\mathrm{Q}x \vee \forall y\mathrm{Q}y)$. The construction of **P**'s winning strategy would be essentially the same, the only difference being that in the latter case **P** herself could choose the disjunct in $\neg\mathrm{P}c_i \vee \forall y\mathrm{P}y$, after having chosen a constant c_i as a value for x.

But what about the *validity* of the sentence $\exists x(\mathrm{Q}x \rightarrow \forall y\mathrm{Q}y)$? Why should we think that **P**'s having a winning strategy in the above-described dialogue would amount to this sentence being *valid*, i.e., *true in all structures* of vocabulary {Q}? Let us first see why we should think it will be true in all *countable* {Q}-structures.

In our dialogical reconstructions of models, we obtain models whose domains are *partitions* of the set of individual constants. Given a vocabulary τ, let us write C_τ for the class of all τ-structures whose domain is a partition of the set $\{c_0, c_1, \ldots\}$. In these models, an n-ary relation symbol is interpreted by an n-tuple $\langle \xi_1, \ldots, \xi_n \rangle$, where the ξ_i are cells of the partition that constitutes the domain of the model. Now any countable {Q}-structure is isomorphic to a {Q}-structure from the class $C_{\{Q\}}$.

Because truth of a first-order sentence is trivially preserved under isomorphism,[22] in order to determine whether such a sentence is valid relative to the class of all countable models it suffices to consider models from the class C_τ. But this means that the dialogical analysis of *validity* of the first-order sentence $\exists x(Qx \to \forall y Qy)$ with respect to countable models will work if, and only if, its analysis of *truth* of this sentence relative to models from the class C_τ will work. Here its truth in a model $\mathcal{M}$ is analyzed as the existence of a w.s. for **P** in the intuitionistic dialogue with the thesis $\exists x(Qx \to \forall y Qy)$ and the extra hypotheses $\forall x_1 \forall x_2(x_1 = x_2 \vee x_1 \neq x_2)$ and $\forall x(Qx \vee \neg Qx)$, in particular such a dialogue in which **O** has responded to **P**'s questions about the extra hypotheses by Skolem functions f and g satisfying the following conditions: $g(c_i) = \texttt{left}$ iff $[c_i] \in Q^{\mathcal{M}}$; and if $f(c_i, c_j) = \texttt{left}$, then $g(c_i) = g(c_j)$. Finally, by Downward Löwenheim–Skolem theorem, the generality of the analysis is not threatened by the restriction to countable models.

The above remarks imply that the dialogical framework sketched here is in principle well suited for our purpose. To see that it really is so, it essentially remains to show that the dialogical analysis of truth of sentences relative to models from the class C_τ works in the desired way. This will be shown in Theorem 22. In anticipation, observe the following.

Remark 18. *Consider a dialogue $\mathcal{D}(A;\mathcal{H})$ with additional hypotheses, involving a hypothesis $\forall x_1 \forall x_2(x_1 = x_2 \vee x_1 \neq x_2)$ about identity. Let i be* **O***'s reply corresponding to this hypothesis, and write $\sim_i$ for the induced equivalence relation on the set $\{c_0, c_1, \ldots\}$. Let us say that* **P***'s strategy f in $\mathcal{D}(A;\mathcal{H})$* respects *the Skolem function i, if the following holds: for any n-tuples $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$ on which f is defined, if $x_1 \sim_i x'_1, \ldots, x_n \sim_i x'_n$, then $f(x_1, \ldots, x_n) \sim_i f(x'_1, \ldots, x'_n)$. It is straightforward to prove that if* **P** *has a winning strategy in $\mathcal{D}(A;\mathcal{H})$, she has a winning strategy respecting i therein. Intuitively, this can be thought of as meaning that the values of the strategy only depend on the objects used as its arguments, not on the names used to refer to these objects.*

[22] Isomorphism between structures is an extreme case of back-and-forth equivalence between them, and it is well known that back-and-forth equivalence entails elementary equivalence, i.e., the fact that the structures satisfy precisely the same first-order sentences. For details see e.g. Hodges (1997, Sections 3.2, 3.3).

Observe that requiring **O** to choose a Skolem function for the disjunction symbol of every sentence of the form $\forall\bar{x}(\mathrm{E}\bar{x} \vee \neg\mathrm{E}\bar{x})$ with $\mathrm{E} \in \tau \cup \{=\}$ can be seen as a straightforward generalization of what happens in the propositional case. For, in the case of propositional logic, **O** had additional concessions of the form $p_i \vee \neg p_i$, one for each propositional atom appearing in the formula under consideration, and **P** could make **O** to choose a disjunct, one for each such sentence. Hence what **P** could ask **O** to choose was nothing else than a Skolem function δ_i for the relevant disjunction, $\delta_i \in \{\mathtt{left}, \mathtt{right}\}$. In the propositional case the Skolem function happens to be a zero-place function (i.e., a constant), but what is at stake is the same in both cases: determining a disjunct for every instance of *tertium non datur* involved. For propositional logic, we *need not* phrase **O**'s commitment in terms of Skolem functions, since we can give a 'finitist' description of the relevant information even without resorting to functions. This option being not in general available in first-order case (due to the possibly infinite number of values of variables), we must there ascend to the level of Skolem functions to get the analogy right.

Note that n initial concessions of **O** yield in the propositional case 2^n ways for **O** to take sides with respect to the disjuncts of the additional hypotheses, while in the case of first-order logic n hypotheses yield in general a *continuum* of possible reactions, as the are $2^{(\aleph_0)^{m_i}}$ different functions $f : \{c_0, c_1, \ldots\}^{m_i} \longrightarrow \{\mathtt{left}, \mathtt{right}\}$ corresponding to a hypothesis with an m_i-ary relation symbol, and

$$\prod_{i:=1}^{n} 2^{(\aleph_0)^{m_i}} = (2^{\aleph_0})^n = 2^{\aleph_0}.$$

Furthermore, it may be noted that **O** also has a continuum of distinct ways of reacting to the hypothesis involving the identity sign, and hence there is a continuum of pairwise distinct domains built in terms of equivalence classes of the set $\{c_0, c_1, \ldots\}$ corresponding to **O**'s different choices of Skolem function for this hypothesis.

Before moving on to state and prove the correspondence theorem, we will establish a normal form of winning strategies of **P** in intuitionistic dialogues with additional hypotheses, for the language of first-order logic without implication.

8.6.2 History-insensitive strategies

In dialogues, players make moves in turns. The resulting sequences of moves $\langle s_0, \ldots, s_n\rangle$ differ from histories (plays) of semantic games most essentially in that **O** is allowed to remake his earlier moves: (*a*) If an earlier propositional move has turned out to lead to a loss for him, **O** may remake that move. This possibility is granted by the *Shifting rule*, i.e., the structural

rule (SR-3). And (*b*) **O** is always entitled to remake his attack on a universal quantifier, and his defense of an existential quantifier.

Due to (*a*), sequences of moves in dialogues—plays of dialogues—are *not* in general dialogical games. For instance, if first a closed dialogical game $\langle s_0, \ldots, s_n \rangle$ is produced (a dialogical game which is hence lost by **O**), **O** is allowed to reconsider any earlier propositional move s_i that he has made (with $i \leq n$) and remake it, hence in effect shifting to the dialogical game $\langle s_0, \ldots, s_{i-1} \rangle$, which may then continue, say by the moves $\langle s'_i, \ldots, s'_k \rangle$. In the end the play

$$\langle s_0, \ldots, s_n, s_0, \ldots, s_{i-1}, s'_i, \ldots, s'_k \rangle$$

will have been constructed. This play is not a dialogical game. The earlier moves are not literally undone by remaking an earlier move. This fact makes it necessary to be careful when speaking of sequences of moves in dialogues, and especially when speaking of strategies of the players in dialogues.

By the rules of dialogues, if **O** remakes his earlier move s_i—hence returning to a subformula B of A superordinate to the subformula C appearing in the last member s_n of the play $\langle s_0, \ldots, s_n \rangle$—the continuation of the play is determined by that subformula B, and any moves s_j with $i < j \leq n$ will be of no relevance to the future course of the play. This is because making a different move instead of s_i, all subformulas of B must now be gone through, regardless of whether some of them already had received a corresponding move s_j ($i < j \leq n$).

At any given position in a play, the sequence of earlier moves is finite, and hence for any subformula token concerning which a move is made in that sequence, there is in the sequence the *last* move made for that subformula token. If $\bar{s}$ is a play, we write $\mathsf{up}(\bar{s})$ for the reduct of $\bar{s}$ consisting of the most recent moves for given tokens of conjunctions, disjunctions, and universal and existential quantifiers: the latest 'updates' of values given to the operators involved. Obviously the length of $\mathsf{up}(\bar{s})$ is at most that of $\bar{s}$. It is also clear that $\mathsf{up}(s_0, \ldots, s_m)$ is determined by the last remade move s_{n+1} in $\langle s_0, \ldots, s_m \rangle$: if s_{n+1} remakes the move s_i, then the sequence $\mathsf{up}(s_0, \ldots, s_m)$ will consist of the suffix $\langle s_{n+1}, \ldots, s_m \rangle$ and the prefix $\mathsf{up}(s_0, \ldots, s_{i-1})$, where $\mathsf{up}(s_0, \ldots, s_{i-1})$ again is determined by its latest revised move, and so on. Hence the sequence $\mathsf{up}(\bar{s})$ can be computed in finitely many steps.

Note that for distinct plays $\bar{s}$, $\bar{s}'$ with possibly divergent lengths corresponding to one and the same subformula token, we may, then, well have: $\mathsf{up}(\bar{s}) = \mathsf{up}(\bar{s}')$. On the other hand, if $\bar{s} \neq \bar{s}'$ and f is an arbitrary strategy of **P**, we may of course still have: $f(\bar{s}) \neq f(\bar{s}')$. There is combinatorially no reason why a strategy could not in principle give distinct values for distinct argument sequences, no matter what those sequences have in common. This observation motivates the following definition.

Definition 19. *A strategy f of* **P** *in a dialogue $\mathcal{D}(A)$ is* history-insensitive*, if any plays $\bar{s}$ and $\bar{s}'$ belonging to $\mathcal{D}(A)$, which correspond to a move for one and the same subformula token B of A, satisfy:*

$$\mathrm{up}(\bar{s}) = \mathrm{up}(\bar{s}') \implies f(\bar{s}) = f(\bar{s}').$$

That is, the value of a history-insensitive strategy only depends on the most recent choices made, during a play, for the operators interpreted in the course of that play.

The idea behind the definition is this: In a dialogue, **O** may in one and the same play revise his choices as he pleases—in cases (*a*) and (*b*) as listed in the beginning of the present subsection. This gives rise to the possibility of a multitude of distinct plays with the same most recent choices corresponding to tokens of subformulas of the thesis A of the dialogue. The value of a history-sensitive strategy only takes into account the most recent choice for any given occurrence of an operator interpreted by **O**: it gives its value for a play $\bar{s}$ as a function of the reduct $\mathrm{up}(\bar{s})$ of $\bar{s}$.

Observe that whenever plays $\bar{s}, \bar{s}'$ correspond to one and the same subformula token B of A, i.e., when their last member involves this token, they necessarily have their *most recent conjunctive/disjunctive choices* in common (cf. Observation 10). Still they may in general differ in their choices for quantifiers.

The following lemma will be of use later. It says that we may always assume **P**'s winning strategy in an intuitionistic dialogue with hypotheses to be history-insensitive in the sense of Definition 19.

Lemma 20. *Let A be a sentence of* $\mathbf{FO}[\tau, =]$*, and let $\tau[A]$ be the set of relation symbols from τ appearing in A.*[23] *If there is a winning strategy for* **P** *in the intuitionistic dialogue $\mathcal{D}(A; \mathcal{H})$ with hypotheses,*

$$\mathcal{H} := \{\forall \bar{x}(\mathrm{E}\bar{x} \vee \neg \mathrm{E}\bar{x}) : \mathrm{E} \in \tau[A] \cup \{=\}\},$$

then there is a history-insensitive winning strategy for **P** *therein.*

Proof. Suppose **P** has a w.s. in $\mathcal{D}(A; \mathcal{H})$, call it f. We may assume it first makes **O** to choose a Skolem function for the disjunction in each hypothesis—relative to the fixed set $\{c_0, c_1, \ldots\}$ of constants—hence in effect determining a model $\mathcal{M}$ whose domain is the set $\{c_0, c_1, \ldots\}/\sim_i$, where $\sim_i$ is the equivalence relation induced by the reply i that **O** gives to the question concerning the hypothesis about the identity sign. If the tokens of subformulas of A corresponding to which it is **P**'s turn to make a move are $B_1, \ldots, B_m$, the strategy f induces strategy functions $f_1, \ldots, f_m$, each f_i providing a move for B_i as a function of **O**'s earlier moves.

[23] Recall that the syntax of $\mathbf{FO}[\tau, =]$ does not involve the implication sign.

Observe first that if $\bar{s}$ is a play of $\mathcal{D}(A;\mathcal{H})$ won by **P**, it corresponds to a literal $\ell = \pm \mathrm{R}c_{i_1} \ldots c_{i_n}$ that is true in $\mathcal{M}$, in the sense that $\langle [c_{i_1}] \ldots [c_{i_n}] \rangle \in \mathrm{R}^{\mathcal{M}}$ *resp.* $\langle [c_{i_1}] \ldots [c_{i_n}] \rangle \notin \mathrm{R}^{\mathcal{M}}$. Clearly, the truth of the literal only depends on whether the tuple $\langle [c_{i_1}] \ldots [c_{i_n}] \rangle$ indeed is contained in the interpretation of the relation symbol R in $\mathcal{M}$. But where do the relevant constants $c_{i_1}, \ldots, c_{i_n}$ come from? They come from *the most recent* choices for quantifiers in the relevant sequence of moves $\bar{s}$. In other words, they come from $\mathsf{up}(\bar{s})$. So the winning conditions in the dialogue are in terms of reducts $\mathsf{up}(\bar{s})$, not in terms of full plays $\bar{s}$.

Consider, then, different possible plays $\bar{s}, \bar{s}'$ at which a move corresponding to B_i is to be made, and which satisfy: $\mathsf{up}(\bar{s}) = \mathsf{up}(\bar{s}')$. Now we may well have that $f_i(\bar{s}) \neq f_i(\bar{s}')$. We must show that there is, however, a history-insensitive strategy function g_i for B_i, i.e., one that satisfies $g_i(\bar{s}) = g_i(\bar{s}')$ for all sequences $\bar{s}, \bar{s}'$ at which a move corresponding to B_i is to be made and which satisfy $\mathsf{up}(\bar{s}) = \mathsf{up}(\bar{s}')$.

Let us define strategy functions g_i for **P** as follows. We first group plays $\bar{s}$ corresponding to B_i in equivalences classes, by putting in a class $[\bar{s}]$ all sequences $\bar{s}'$ such that $\mathsf{up}(\bar{s}) = \mathsf{up}(\bar{s}')$. (These equivalence classes $[\bar{s}]$ must not be confounded with the $\sim_i$-equivalence classes such as the $[c_{i_k}]$ above.) For each equivalence class we choose a representative $\bar{s}$, and then for each $\bar{s}' \in [\bar{s}]$, we set

$$g_i(\bar{s}') := f_i(\mathsf{up}(\bar{s})).$$

(Note that $\mathsf{up}(\bar{s})$ itself is a combinatorially possible play of $\mathcal{D}(A)$, so f_i is defined on it.)

We claim that $g := (g_1, \ldots, g_m)$ constitutes a winning strategy for **P** in $\mathcal{D}(A;\mathcal{H})$. Suppose for contradiction that there is a play $\bar{s}$ where **P** has moved according to g but which is won by **O**. This means that $\bar{s}$ corresponds to a literal ℓ which is false in $\mathcal{M}$: Either $\ell = \mathrm{R}c_{i_1} \ldots c_{i_n}$ and **P** cannot concede it since **O** has not conceded it earlier, or $\ell = \neg \mathrm{R}c_{i_1} \ldots c_{i_n}$ and **O** has already chosen $c_{i_1} \ldots c_{i_n}$.

Consider, then, those c_{i_k} that are chosen by **P**. Each such c_{i_k} is chosen by a strategy function g_i of **P**, as a function of a fixed sequence of **O**'s earlier moves, namely the sequence $\mathsf{up}(\bar{s})$ of the most recently made choices for the operators of B_i. But then there is a play of $\mathcal{D}(A;\mathcal{H})$ where **O** makes exactly those moves on the basis of which the moves c_{i_k} of **P** are chosen: a play where **O** never revises his moves. Then these moves of **P** are in fact made following the winning strategy f, and yet **P** loses. This is a contradiction. □

Before stating and proving the theorem describing the connection between GTS and dialogical logic in the case of first-order logic, let us consider a specific example.

Example 21. Think of the sentence

$$A := \forall x \exists y Rxy \vee \exists x \forall y \neg Rxy.$$

We show how to transform *Proponent*'s winning strategy in the dialogue $\mathcal{D}(A;\mathcal{H})$ with $\mathcal{H} := \{\forall x \forall y(x = y \vee x \neq y), \forall x \forall y(Rxy \vee \neg Rxy)\}$ into a family of *Eloise*'s winning strategies in games $G(A, \mathcal{M})$ with $\mathcal{M} \in \mathcal{C}_{\{R\}}$; and *vice versa*.

($\Longrightarrow$) Suppose f is a w.s. for **P** in $\mathcal{D}(A;\mathcal{H})$. We may assume, without loss of generality, that following f, **P** first poses questions ??-$\vee$ about **O**'s initial concessions, that is, asks **O** to specify Skolem functions for the disjunction signs of the sentences $\forall x \forall y(x = y \vee x \neq y)$ and $\forall x \forall y(Rxy \vee \neg Rxy)$. **O**'s reply to each question is one out of $2^{\aleph_0}$ distinct functions of type $\{c_0, c_1, \ldots\} \times \{c_0, c_1, \ldots\} \longrightarrow \{\texttt{left}, \texttt{right}\}$, call these functions d_1 and d_2, respectively. As noted in Remark 18, we may assume, without loss of generality, that f respects d_1.

The following reasoning applies to the Skolem function d_2 corresponding to the hypothesis about R. If for every c_i there is c_j such that $d_2(c_i, c_j) = \texttt{left}$, then the strategy f, being a winning strategy, makes **P** to choose the left disjunct, whereas if for some c_i it holds that every c_j satisfies $d_2(c_i, c_j) = \texttt{right}$, then f makes **P** to choose the right disjunct.

In the first-mentioned case, if **O** asks ?-$\forall x/c_i$ and then further asks ?-$\exists y$ concerning **P**'s concession $\exists y \mathrm{R} c_i x$, **P**'s strategy f gives a constant $f(c_i) = c_k$ such that $d_2(c_i, c_k) = \texttt{left}$ (by assumption such c_k exists), and then poses to **O** the question ?-$d_2/c_i c_k$, i.e., asks him explicitly to admit that indeed according to the Skolem function d_2, we have $Rc_i c_k$. Then **P** replies to the question ?-$\exists y$ by conceding $Rc_i c_k$, which she can do, since **O** already conceded this literal. Thereby **P** wins the play.

In the second case, i.e., if f guides **P** to the right disjunct, f further provides a constant c_0 so that **P** concedes $\forall y \neg Rc_0 x$, and the constant satisfies $d_2(c_0, c_j) = \texttt{right}$ for all c_j. Then **O** picks out some c_i and asks **P** to concede $\neg Rc_0 c_i$, which she can do. If, then, **O** attacks this concession by himself conceding $Rc_0 c_i$, **P** may put forward to **O** the question ?-$d_2/c_0 c_i$, i.e., ask him to admit that indeed $\neg Rc_0 c_i$. The Skolem function d_2 commits **O** to this concession. Hence **O** contradicts himself and **P** ends up winning the play.

Let us now see how such a w.s. f of **P** transforms into a family $\mathcal{G}$ of winning strategies of *Eloise*, one strategy for each semantic game $G(A, \mathcal{M})$, where $\mathcal{M} \in \mathcal{C}_{\{R\}}$. Let $\mathcal{M}$ be any such model. Hence there are uniquely determined Skolem functions d_1' and d_2' for the disjunction signs appearing in the sentences $\forall x \forall y(x = y \vee x \neq y)$ and $\forall x \forall y(Rxy \vee \neg Rxy)$, respectively, which jointly identify precisely the model $\mathcal{M}$—namely, the functions d_1' and d_2' such that c_i and c_j are members of the same element ξ of $dom(\mathcal{M})$ iff $d_1(c_i, c_j) = \texttt{left}$, and

$$\langle [c_i], [c_j] \rangle \in R^{\mathcal{M}} \text{ iff } d_2(c_i, c_j) = \texttt{left}.$$

We define a strategy g for *Eloise* making use of **P**'s w.s. f in $\mathcal{D}(A)$. Because f is a winning strategy, it tells what to do in particular in the case that **O** has 'specified the model $\mathcal{M}$' with his answers to the questions ??-$\vee$ about the initial concessions.

For the disjunction in A, let g tell *Eloise* to choose the same disjunct that f tells **P** to choose as a reply to the corresponding question ?-$\vee$ by **O**. Hence g chooses, by definition, `left` precisely when the model $\mathcal{M}$ satisfies $\forall x \exists y Rxy$ (since this is when f has **P** to make this choice of a disjunct). Then if *Abelard* picks out an equivalence class ξ, fix $c_i \in \xi$ as its representative, and let *Eloise* reply $g(\xi) := [f(c_i)]$, i.e., to choose as witness the equivalence class generated by the witness $f(c_i)$ that **P** chooses if **O** asks him about the existential quantifier $\exists y$, having just previously instantiated x by c_i. (The definition of g does not depend on the choice of the representative, because f respects d_1.) Clearly this leads to a play won by *Eloise*. If, again, g makes *Eloise* to choose `right`, let g make *Eloise* to go on by picking out $[c]$ if the constant that f picks out whenever the play has led to the right disjunct is c. Again *Eloise* clearly wins the resulting play. We may conclude, then, that g is a w.s. for *Eloise* in $G(A, \mathcal{M})$.

($\Longleftarrow$) Suppose that for every $\mathcal{M} \in C_{\{R\}}$, *Eloise* has a w.s. ($g_{\mathcal{M}}$) in the semantic game $G(A, \mathcal{M})$. Since $g_{\mathcal{M}}$ is a winning strategy, it makes *Eloise* to choose the left disjunct in A, if $\mathcal{M}$ makes true $\forall x \exists y Rxy$, and otherwise makes her to choose the right disjunct. In the left disjunct $g_{\mathcal{M}}$ then goes on to dictate a choice $\zeta = g_{\mathcal{M}}(\xi)$, as a function of *Abelard*'s choice of ξ; the values will satisfy $\langle \xi, \zeta \rangle \in R^{\mathcal{M}}$, since $g_{\mathcal{M}}$ is a w.s. for *Eloise*. And in the right disjunct *Eloise* will choose a constant ξ_0 such that for all ζ, $\langle \xi_0, \zeta \rangle \notin R^{\mathcal{M}}$.

Define a strategy f for **P** in the dialogue $\mathcal{D}(A; \mathcal{H})$ as follows. First f makes **O** to answer to the questions ??-$\vee$ about the additional hypotheses. **O**'s replies yield Skolem functions d_1 and d_2, which jointly determine a structure $\mathcal{M} \in C_{\{R\}}$. For this structure $\mathcal{M}$, in particular, *Eloise* has—by assumption—a w.s. ($g_{\mathcal{M}}$) in the semantic game $G(A, \mathcal{M})$. We use $g_{\mathcal{M}}$ in defining how to continue playing in $\mathcal{D}(A)$.

Let f provide the reply `left` to the question ?-$\vee$, if $g_{\mathcal{M}}$ chooses `left` for the disjunction; otherwise let f reply `right`. In the former case, if c_i is the constant **O** chooses for the variable x, and $g_{\mathcal{M}}(\xi) = [\zeta]$ with $c_i \in \xi$, let c_j be a representative of ζ and put $f(c_i) := c_j$. Then first make **O** to admit that $Rc_i f(c_i)$, which he must do since $f(c_i)$ is chosen on the basis of *Eloise*'s winning strategy relative to the model $\mathcal{M}$ that **O** himself fixed by his choices of the Skolem functions d_1 and d_2; and finally let **P** concede the literal $Rc_i f(c_i)$. In the latter case, let f interpret x on the basis of g. Namely, for each equivalence class ξ_0 choose a representative c_0, and if g yields ξ_0, let f yield $c_0 \in \xi_0$. Then if **O** asks **P** to concede $\neg Rc_0 c_i$, let **P** make **O**, in turn, concede $\neg Rc_0 c_i$. The latter must concede this negative literal, because the Skolem functions d_1 and d_2 determine the model $\mathcal{M}$ and $\langle [c_0], [c_i] \rangle \notin R^{\mathcal{M}}$. Then **P** may freely concede

$\neg Rc_0c_i$, whereafter **O** can only attack **P**'s concession by contradicting himself. Obviously f always leads to a play won by **O**. That is, f will be a w.s. for **P** in dialogue $\mathcal{D}(A)$.

We are in a position to formulate and prove the correspondence result between GTS and dialogical logic in characterizing validity of first-order sentences.

8.6.3 The correspondence result

Theorem 22. *Let A be any sentence of* $\mathbf{FO}[\tau]$*, and let* $\tau[A]$ *be the set of relation symbols from* τ *appearing in A. The following conditions are equivalent:*

(*i*) *There is a w.s. for* Proponent *in the intuitionistic dialogue* $\mathcal{D}(A;\mathcal{H})$*;*

(*ii*) *For every* τ*-structure* $\mathcal{M} \in C_{\tau[A]}$*, there is a w.s. for* Eloise *in the semantic game* $G(A, \mathcal{M})$*;*

where the set $\mathcal{H}$ *of additional hypotheses consists of the sentence* $\forall x_1 \forall x_2(x_1 = x_2 \vee x_1 \neq x_2)$*, together with the sentences of the form* $\forall x_1 \ldots \forall x_n(\mathrm{R}x_1 \ldots x_n \vee \neg \mathrm{R}x_1 \ldots x_n)$*, one sentence for each relation symbol* $\mathrm{R} \in \tau[A]$*, n being the arity of* R*. Furthermore, there is an explicit set of instructions turning a family of* Eloise*'s winning strategies into* Proponent*'s winning strategy; and an explicit set of instructions turning* Proponent*'s winning strategy into a family of* Eloise*'s winning strategies in games played on models from the class* $C_{\tau[A]}$*.*

As in the case of propositional logic, we prove the correspondence theorem in two steps: first we establish Lemma 23 and then Lemma 25.

Lemma 23. *Condition* (*ii*) *of Theorem 22 implies its condition* (*i*)*.*

Proof. Suppose that for every model $\mathcal{M} \in C_{\tau[A]}$, there is a w.s. for *Eloise* in $G(A, \mathcal{M})$. It must be shown that there is a w.s. for **P** in the intuitionistic dialogue $\mathcal{D}(A;\mathcal{H})$. Let us move on to describe a strategy f of **P** in $\mathcal{D}(A;\mathcal{H})$.

To begin with, the strategy f tells **P** to ask, for every $\mathrm{E} \in \tau[A] \cup \{=\}$, **O** to reply to the question ??-$\vee$ with regard to every hypothesis $\forall \bar{x}(\mathrm{E}\bar{x} \vee \neg \mathrm{E}\bar{x})$. **O**'s answers will, then, determine a $\tau[A]$-structure $\mathcal{M}$ whose domain is a partition of the set $\{c_0, c_1, \ldots\}$. (The specific partition is determined by the Skolem function that **O** chooses in responding to the question about the hypothesis $\forall x_1 \forall x_2(x_1 = x_2 \vee x_1 \neq x_2)$, as indicated in Section 8.6.1.) By hypothesis, there is in particular a w.s. (g) for *Eloise* in the game $G(A, \mathcal{M})$ corresponding to this model $\mathcal{M}$. We go on to construct a continuation of **P**'s strategy f in $\mathcal{D}(A;\mathcal{H})$ using *Eloise*'s strategy g.

Let us first generate $\mathcal{W}$ as the class of those histories of the semantic game $G(A, \mathcal{M})$ that are realizable when *Eloise* follows the strategy g: First, let

$(A, \langle\rangle) \in \mathcal{W}$. Second, suppose we have $h \in \mathcal{W}$, with (B, γ) as the last member of h. Then:

- If $B = C \wedge D$ and $\rho(B, Eloise) = Verifier$, or $B = C \vee D$ and $\rho(B, Eloise) = Falsifier$, then for all $E \in \{C, D\}$: $h^\frown(E, \gamma) \in \mathcal{W}$.
- If $B = C \vee D$ and $\rho(B, Eloise) = Verifier$, or $B = C \wedge D$ and $\rho(B, Eloise) = Falsifier$, then $h^\frown(f(h), \gamma) \in \mathcal{W}$.
- If $B = \forall x D(x)$ and $\rho(B, Eloise) = Verifier$, or $B = \exists x D(x)$ and $\rho(B, Eloise) = Falsifier$, then for all ξ: $h^\frown(D(x), \gamma[x/\xi]) \in \mathcal{W}$.
- If $B = \exists x D(x)$ and $\rho(B, Eloise) = Verifier$, or $B = \forall x D(x)$ and $\rho(B, Eloise) = Verifier$, then $h^\frown(D(x), \gamma[x/f(h)]) \in \mathcal{W}$.
- If $B = \neg C$, then $h^\frown(C, \gamma) \in \mathcal{W}$. Observe that by definition of $G(A, \mathcal{M})$, $\rho(C, Eloise) = \rho(B, Abelard)$.

Given that the initial hypotheses have already been processed, and that a model $\mathcal{M} \in C_{\tau[A]}$ has thereby been determined, plays of $\mathcal{D}(A; \mathcal{H})$ correspond to plays of $G(A, \mathcal{M})$ in a straightforward way. The only difference is that whereas in the dialogue $\mathcal{D}(A; \mathcal{H})$ moves corresponding to quantifiers are choices of individual constants, the quantifier moves in the semantic game $G(A, \mathcal{M})$ involve choosing an equivalence class consisting of individual constants. Now if a play $\bar{s}$ of $\mathcal{D}(A; \mathcal{H})$ has been reached, and **O** asks **P** to choose a disjunct (conjunct) of a formula B, the strategy f is defined by letting it pick out the unique disjunct (conjunct) E such that $h^\frown(E, \gamma) \in \mathcal{W}$, where h is the play of $G(A, \mathcal{M})$ determined by $\mathsf{up}(\bar{s})$, and the last member of h is (B, γ). (The history h is obtained from $\mathsf{up}(\bar{s})$ by replacing each member of $\mathsf{up}(\bar{s})$ which is an individual constant by the equivalence class generated by that individual constant.)

Further, if, having arrived at a play $\bar{s}$, **O** asks **P** to choose a value to the existential quantifier $\exists x$ appearing in a given subformula $B := \exists x D(x)$ (the quantifier hence appearing under an even number of negation signs), choose for every $\xi \in dom(\mathcal{M})$ a representative $c_i \in \xi$, and let f tell **P** to choose $c_i \in \xi$ with $h^\frown(D(x), \gamma[x/\xi]) \in \mathcal{W}$, where h is the play of $G(A, \mathcal{M})$ determined by $\mathsf{up}(\bar{s})$, and the last member of h is (B, γ). Define f similarly in the case that the subformula $\forall x D(x)$ appears under an odd number of negation signs. Hence defined, the strategy f indeed respects the Skolem function i chosen by **O** as a response to the question about the additional hypothesis concerning the identity sign. This is because f is defined using fixed representatives of the

equivalence classes determined by i (or, in other words, fixed representatives of the elements of the domain of the model $\mathcal{M}$ determined by having chosen this Skolem function).

Claim 24. *The strategy f is a w.s. for* **P** *in $\mathcal{D}(A;\mathcal{H})$.*

Proof. Due to its definition, f leads to a literal $\ell = \pm \mathrm{R}c_{i_1} \dots c_{i_n}$ (or a conjunction of literals $\ell_1, \dots, \ell_l$) true in $\mathcal{M}$. This means that **O** has conceded $\mathrm{R}c_{i_1} \dots c_{i_n}$ (positive literal), or has conceded $\neg \mathrm{R}c_{i_1} \dots c_{i_n}$ (negative literal). In the former case **P** is indeed in a position to reply ℓ (**O** has already conceded it); in the latter case the play continues so that **O** concedes ℓ, hence contradicting himself. □

Let $\mathcal{G} = \{g_i : i < \lambda\}$ be a family of *Eloise*'s winning strategies in games $G(A, \mathcal{M})$ with $\mathcal{M} \in C_{\tau[A]}$. The explicit set of instructions for generating a winning strategy for **P** in $\mathcal{D}(A;\mathcal{H})$ consists, then, of first making **O** to determine a model $\mathcal{M}$ by his answers to questions ??-$\vee$ corresponding to the symbols in $\tau[A] \cup \{=\}$. The w.s. g_i of *Eloise* in $G(A, \mathcal{M})$ then determines the construction of the set $\mathcal{W}$ as explained above, which directly defines a winning strategy for **P** proceeding from **O**'s answers to **P**'s questions concerning **O**'s initial concessions. □

We move on to prove the other direction of the correspondence theorem.

Lemma 25. *Condition* (*i*) *of Theorem 22 implies its condition* (*ii*).

Proof. Suppose **P** has a w.s., f, in the intuitionistic dialogue $\mathcal{D}(A;\mathcal{H})$. As in the propositional case, such a strategy need not consist of first asking **O** about all hypotheses, but we may without loss of generality assume it does. By Lemma 20, we may assume that f is history-insensitive. Further, we may assume that if i is **O**'s reply to the question concerning the identity sign, the strategy f respects i.

We must show that for all structures $\mathcal{M} \in C_{\tau[A]}$, there is a w.s. for *Eloise* in $G(A, \mathcal{M})$. Let $\mathcal{M} \in C_{\tau[A]}$ be arbitrary. First consider a play of $\mathcal{D}(A)$, where **P** has received such answers to her questions about the initial concessions of **O** that these constitute precisely the model $\mathcal{M}$: The Skolem function i corresponding to the hypothesis about identity sign satisfies: $i(c_i, c_j) = \texttt{left}$ iff c_i and c_j belong to the same cell of the partition of the set $\{c_0, c_1, \dots\}$ (the partition being given by the domain of $\mathcal{M}$); and for every $\mathrm{R} \in \tau[A]$, the corresponding Skolem function d_{R} satisfies $d_{\mathrm{R}}(c_{i_1}, \dots, c_{i_n}) = \texttt{left}$ iff $\langle [c_{i_1}], \dots, [c_{i_n}] \rangle \in \mathrm{R}^{\mathcal{M}}$.

Let us construct a set $\mathcal{V}$ of pairs consisting of subformulas of B of A and sequences $\mathtt{up}(\bar{s})$ of updated moves made by **P** and **O** in the course of the dialogue $\mathcal{D}(A;\mathcal{H})$. Note that it makes sense to apply a strategy of **P** to an updated

sequence $\mathtt{up}(\bar{s})$, because such sequences are themselves possible plays of dialogues, namely plays where **O** has not revised any move once made. (Hence they are, in particular, dialogical games.) As in Section 8.5.1, we denote here by $\sharp[B]$ the number of negation signs to which the subformula token B is syntactically subordinate in a given formula A. First, we put $(A, \langle\rangle) \in \mathcal{V}$. Suppose, then, that $(B, \bar{s}) \in \mathcal{V}$.

- If $B = C \wedge D$ and $\sharp[B]$ is even or equal to zero, or $B = C \vee D$ and $\sharp[B]$ is odd, then $(C, \mathtt{up}(\bar{s})^\frown \mathtt{left}), (D, \mathtt{up}(\bar{s})^\frown \mathtt{right}) \in \mathcal{V}$.
- If $B = C \vee D$ and $\sharp[B]$ is even or equal to zero, or $B = C \wedge D$ and $\sharp[B]$ is odd, then $(E, \bar{s}^\frown f(\bar{s})) \in \mathcal{V}$, where $E \in \{C, D\}$ is the disjunct/conjunct identified by the value $f(\bar{s}) \in \{\mathtt{left}, \mathtt{right}\}$.
- If $B = \forall x D(x)$ and $\sharp[B]$ is even or equal to zero, or $B = \exists x D(x)$ and $\sharp[B]$ is odd, then for every c_i, $(D(x), \mathtt{up}(\bar{s})^\frown (x, c_i)) \in \mathcal{V}$.
- If $B = \exists x D(x)$ and $\sharp[B]$ is even or equal to zero, or $B = \forall x D(x)$ and $\sharp[B]$ is odd, then $(D(x), \bar{s}^\frown f(\bar{s})) \in \mathcal{V}$, where $f(\bar{s}) = (x, c_i)$ is the assignment of a constant to the variable x, as determined by the strategy f.
- If $B = \neg C$, then $(C, \bar{s}) \in \mathcal{V}$. By the definition of the dialogue $\mathcal{D}(A; \mathcal{H})$, we have that $\sharp[C]$ is even or equal to zero iff $\sharp[B]$ is odd.

By construction, the sequences $\bar{s}$ appearing in pairs $(B, \bar{s}) \in \mathcal{V}$ are not full sequences of moves made by the players of the dialogue—but reducts of such full sequences consisting of the most recent updates of the moves of the players. (It is of course only **O**'s moves that can have been updated.) However, by the fact that we may assume f to be history-insensitive (Lemma 20), it makes sense to apply f to these sequences.

We define a strategy g of *Eloise* in the semantic game $G(A, \mathcal{M})$ as follows. If h is a history of $G(A, \mathcal{M})$ at which it is *Eloise*'s turn to move—hence corresponding to a subformula $B \in \{C \vee D, \exists x D(x)\}$ with $\sharp[B]$ even or equal to zero, or to a subformula $B \in \{C \wedge D, \forall x D(x)\}$ with $\sharp[B]$ odd—consider the pair $(B, \bar{s})$ determined by the history h; the choice is unique up to the choice of representatives of the relevant equivalence classes. (The sequence $\bar{s}$ is obtained from h by replacing each member of h which is an equivalence class of individual constants by an arbitrarily chosen representative of the equivalence class.) Then find the uniquely determined pair $(B, \bar{s}^\frown a)$ in $\mathcal{V}$, with a of the form (x, c_i), or $\mathtt{left}$, or $\mathtt{right}$, and let g make *Eloise* respond to the appropriate question by choosing $(x, [c_i])$, $\mathtt{left}$, or $\mathtt{right}$, respectively.

Claim 26. *The strategy g is a w.s. for* Eloise *in* $G(A, \mathcal{M})$.

Proof. The last move of *Eloise* made in accordance with g leads to a literal ℓ (or a conjunction of literals). By the definition of g, if $\ell =$

$\mathrm{R}c_{i_1} \ldots c_{i_n}$, then ℓ is true in $\mathcal{M}$ (since then **O** must have chosen such a Skolem function corresponding to R that it commits him to conceding ℓ). If, on the other hand, $\ell = \neg\mathrm{R}c_{i_1} \ldots c_{i_n}$, then ℓ is false in $\mathcal{M}$, for then **O** must contradict himself by conceding $\mathrm{R}c_{i_1} \ldots c_{i_n}$. □

Let f be a winning strategy for **P** in $\mathcal{D}(A;\mathcal{H})$. The explicit set of instructions for generating a family $\mathcal{G}$ of *Eloise*'s winning strategies in the semantic games $G(A,\mathcal{M})$ with $\mathcal{M} \in \mathcal{C}_{\tau[A]}$ consists, then, in first transforming f into such a winning strategy g_0 which first asks **O** to reply to the questions ??-$\vee$ about the hypotheses in $\mathcal{H}$. **O**'s answers to those questions then determine a model $\mathcal{M} \in \mathcal{C}_{\tau[A]}$, relative to which g_0 further induces, as explained above, a set $\mathcal{V}$ which defines a w.s. for *Eloise* in $G(A,\mathcal{M})$. □

8.7 Concluding remarks

By way of conclusion, let us touch upon certain methodological issues, and point to questions that remain for future research to tackle.

8.7.1 What to read into the two analyses?

Hintikka's game-theoretical semantics was originally motivated by the wish to see how to make formal sense of the later Wittgenstein's idea of language game. The connection is not merely verbal; Hintikka has on numerous occasions made use of game-theoretical semantics in discussing the issue of meaning constitution—and the role of Wittgensteinian language games as rule-governed activities mediating the relation between language and reality (see, e.g., Hintikka and Hintikka, 1986; Hintikka, 1993). Among other things he has made it clear that, as he conceives them, his semantic games are not games played by the language users; the above *Abelard* and *Eloise* are not Jack and Jill having a conversation; they are idealized parties or poles of an equally idealized activity which is constitutive of language–world relations.

The dialogical approach, on the other hand, was originally motivated by considerations related to intuitionistic logic. This framework is typically sympathetic to philosophical positions such as anti-realism about meaning, and is related to constructivism and the conception that inference rules are meaning-constitutive. Now, the interesting philosophical point here is that also dialogical logic was conceived against the background of Wittgenstein's language games. Moreover, the so-called *Erlangen–Konstruktivismus* aimed explicitly at implementing language games in logic. Indeed, already in 1967 Kuno Lorenz, in his *Habilitationsschrift*,[24] delved into the relation between the first and the second Wittgenstein, challenging the standard intepretations

[24] Published in 1970 under the title *Elemente der Sprachkritik*.

which understood Wittgenstein's later work as a refutation of his earlier work. This approach yielded several other publications, including an introductory book *Logische Propädeutik* on the philosophy of language and logic, written by Kamlah and Lorenzen (1973), and a more comprehensive book *Konstruktive Logik, Ethik und Wissenschaftstheorie* by Paul Lorenzen and Oswald Schwemmer on logic, ethics and the philosophy of science (1975).

A research question that still remains open is that of comparing the dialogical interpretation of Wittgenstein with Hintikka's interpretation. (Cf. the suggestive tripartite paper by Dascal et al. [1995].) However, one important difference must be pointed out. From the very beginnings language games were understood in the dialogical school as delivering the semantic basis to the proof-theoretic and anti-realist meaning theory of intuitionistic logic. This opened a path to philosophers like Michael Dummett, who some time later developed these ideas further in the framework offered by Gentzen systems. It is not difficult to read statements of adherents of the two approaches as implying that in the dialogical framework, the lion's share of importance is put on proof theory, while the developments based on game-theoretical semantics virtually ignore proof theory and put almost all weight on model theory.

Be that as it may, it is our conviction that it is perfectly possible and furthermore potentially very fruitful for both parties—GTS and dialogical logic—to abstract away from any motivating factors, and to begin by concentrating on the mathematically formulated theories themselves. This enables laying down various explicit correspondence results between the two approaches, as witnessed by the present paper. Such results—once achieved—will then actually facilitate the philosophical assessment of the relative vices and virtues of the two approaches. The present results, for instance, lend themselves to the reading that in the formulations of propositional logic and first-order logic discussed here (logics without implication), the classical assumptions of GTS really are limited to the assumption that the atomic predicates obey *the law of excluded middle*—classical, that is, as opposed to intuitionistic. Otherwise semantic games and intuitionistic dialogues proceed perfectly on a par.

8.7.2 Open questions

Various questions suggest themselves for future research. Among them is the question whether the results of the present paper can be generalized to languages of propositional logic and first-order logic *with* implication. The problem is a genuine one, since implication is not reducible to the other connectives in intuitionistic logic—which is the logical basis of the dialogues considered here.

As pointed out above, the main results of this paper can be understood as saying that the classical import of GTS is in the assumption that the models are 'determinate' in the sense that atomic formulas satisfy *tertium non datur*.

In propositional logic this means that the relevant notion of negation ($\neg$) satisfies, for every propositional atom p: $M \models \neg p$ iff $M \not\models p$; and in first-order logic that $\mathcal{M} \models \neg(a_1 = a_2)$ iff $a_1^{\mathcal{M}} \neq a_2^{\mathcal{M}}$, and, if $\mathrm{R} \in \tau$,

$$\mathcal{M} \models \neg \mathrm{R}a_1 \ldots a_n \quad \text{iff} \quad (a_1^{\mathcal{M}}, \ldots, a_n^{\mathcal{M}}) \notin \mathrm{R}^{\mathcal{M}}.$$

There are several ways to modify our framework. For one thing, we can consider 'partial' models in connection with GTS, i.e., models where there are atoms p such that neither p nor $\neg p$ is true, or interpretations of relation symbols R such that *not all* tuples $(a_1^{\mathcal{M}}, \ldots, a_n^{\mathcal{M}})$ in the complement of $\mathrm{R}^{\mathcal{M}}$ serve to make $\neg \mathrm{R}a_1 \ldots a_n$ true. (For a game-theoretical study of partiality in connection with propositional logic, see Sandu and Pietarinen, 2001.) For another thing, on the side of dialogues we may drop the initial concessions of *Opponent*. Then, considering logics with implication, we may study the interrelations of GTS and dialogic. It should be noted that such a comparison cannot be begun before further conceptual decisions are made. Perhaps the single biggest question that we leave for future research moving along the lines of this paper, is how to represent intuitionistic implication in GTS. It is by no means evident how to do it. Implication should be represented game-theoretically in such a way that its relationships with other logical connectives (notably disjunction and negation) would end up being correct from the intuitionistic viewpoint. For instance, negation should be definable in terms of implication as $\neg A := A \to \bot$; the disjunction $\neg A \vee B$ should have as its logical consequence intuitionistic implication $A \to B$, but not *vice versa*; and A should have as its logical consequence the double negation $\neg\neg A$ (defined via implication), but not *vice versa*. A brute-force solution would of course be to introduce an embedding t from intuitionistic first-order logic into quantified modal logic **S4** with expanding first-order domains,[25] and to formulate GTS relative to this modal logic, systematically considering modal formulas $t(A)$ in place of first-order formulas A. However, might there not be a game-theoretical formulation of intuitionistic logic with less air of arbitrariness?

While it certainly serves the purpose of understanding the relations of the two approaches of GTS and dialogical logic to abstract away from the philosophical convictions that have historically motivated them, it is of conceptual interest to see whether such increased understanding throws light on the underlying philosophical views. In particular, attempting to formulate dialogues characterizing validity in connection with incomplete logics—say second-order logic, IF ('Independence-Friendly') first-order logic, and various incomplete modal logics[26]—would throw light on whether the dialogical approach is ultimately proof-theoretic rather than model-theoretic.

[25] For intuitionistic first-order logic and **S4**, see, e.g., Kontchakov et al. (2005).
[26] For incomplete modal (and tense) logics, see Blackburn et al. (2002, 212–216).

Moreover, such research would perhaps help to clarify the notion of *formal dialogue* (or *formal model*) implicit in the dialogical approach—the notion which was an answer of Paul Lorenzen and Kuno Lorenz to Alfred Tarski—in fact, Tarski challenged the dialogicians to provide an adequate proof-theoretic notion which could compete with the concept of truth in a model.[27] Indeed, in which way does incompleteness get manifested in dialogues? One possible argumentation path is the following: If the existence of a winning strategy for *Proponent* means ultimately provability from the empty set of premises rather than truth in every model, then for incomplete logics some valid sentences should fail to be associated with *Proponent*'s winning strategy. On the other hand, if it is truth in every model that is game-theoretically characterized by dialogues, then they are not first and foremost vehicles of proof theory. A second possible kind of response would start with the notions of informal and canonical proof, and proceed to distinguish between the proof-conditional semantics of the logical connectives and the notion of validity, in a way analogous to what is defended nowadays by some substructuralists.

In brief, incomplete logics are a good test case for the generality of the kinds of correspondence results that we have studied in the present paper. Here, once more, two types of argument follow. The first one will claim that if such results can be proven also for incomplete logics, dialogical logic is not ultimately proof-theoretic by nature—while if they cannot, then the correspondence between GTS and dialogic breaks down precisely because the latter really deals with proof theory and the former with semantics and model theory. The second type of argument would insist that this only shows (i) that despite the different approaches to meaning, GTS and dialogic have both an adequate notion to handle conditional reasoning, i.e., reasoning proceeding from hypotheses; and (ii) that the fact mentioned in (i) does not mean that one approach must reduce to the other.

Certainly, it will be difficult to work out how exactly the wider philosophical viewpoints associated with dialogic on the one hand (pragmatism, anti-realism about meaning, logical pluralism), and with GTS on the other (meaning constitution in terms of rule-governed human activities), are connected with the technical details of the two approaches. Yet findings about the interconnection between the two approaches, or about the lack thereof, will serve to clarify even these philosophical positions, insofar as such results show to which extent expectations of the logical repercussions of these views are perhaps underdetermined by the views themselves.

We believe and wish to have indicated in the present paper, in particular, that studying the interrelation of the two game-theoretically formulated approaches

[27] This happened in 1957–1958, when Lorenzen was visiting Princeton, cf. Lorenz (2001, 257).

to logic—game-theoretical semantics and dialogical logic—promises to be highly useful for the purpose of better understanding the logical nature of both views, and, in general, to have sketched the import of such a comparative study to the better understanding of the larger philosophical surroundings of these views.

Acknowledgments

We wish to thank Laurent Keiff and Manuel Rebuschi for useful comments on earlier phases of this paper, and Manuel Rebuschi in particular for pointing out the paper of Saarinen (1978), containing an insightful comparison of dialogic and game-theoretical semantics.

References

Blackburn, P., de Rijke, M., and Venema, Y. (2002). *Modal Logic*. Cambridge University Press, Cambridge.

Blass, A. (1992). A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56: 183–220.

Carlson, L. (1983). *Dialogue Games. An Approach to Discourse Analysis*. Reidel, Dordrecht.

Dascal, M., Hintikka, J., and Lorenz, K. (1995). Jeux dans le langage/Games in Language/Spiel in der Sprache. In Dascal, M., Gerhardus, D., Lorenz, K., and Meggle, G., editors, *Sprachphilosophie/Philosophy of Language/La Philosophie du langage*, pages 1371–1390. De Gruyter, Berlin.

Felscher, W. (1985). Dialogues, strategies and intuitionistic provability. *Annals of Pure and Applied Logic*, 28:217–254.

Haas, G. (1980). Hypothesendialoge, konstruktiver Sequenzenkalkül und die Rechtfertigung von Dialograhmenregeln. In Gethmann, C. F., editor, *Theorie des wissenschaftlichen Argumentierens*, pages 136–161. Suhrkamp, Frankfurt.

Henkin, L. (1950). Completeness in the theory of types. *Journal of Symbolic Logic*, 15(2): 81–91.

Henkin, L. (1961). Some remarks on infinitely long formulas. In *Infinitistic Methods*, pages 167–183. Pergamon, Oxford.

Hintikka, J. (1968). Language-games for quantifiers. Americal Philosophical Quarterly Monograph Series 2: Studies in Logical Theory. Basil Blackwell, Oxford.

Hintikka, J. (1973). *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*. Clarendon, Oxford.

Hintikka, J. (1987). Game-theoretical semantics as a synthesis of verificationist and truth-conditional meaning theories. In LePore, E., editor, *New Directions in Semantics*. Academic, London.

Hintikka, J. (1993). The original *Sinn* of Wittgenstein's philosophy of mathematics. In Puhl, K., editor, *Wittgenstein's Philosophy of Mathematics*, pages 24–51. Hölder–Pichler–Tempsky, Vienna.

Hintikka, J. (1996). *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge.

Hintikka, J. (2002). Hyperclassical logic (a.k.a. IF logic) and its implications for logical theory. *Bulletin of Symbolic Logic*, 8(3):404–423.

Hintikka, J. and Rantala, V. (1976). A new approach to infinitary languages. *Annals of Mathematical Logic*, 10:95–115.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Amsterdam.

Hintikka, M. B. and Hintikka, J. (1986). *Investigating Wittgenstein*. Basil Blackwell, Oxford.

Hodges, W. (1997). Model theory. In Rota, G.-C., editor, *Encyclopedia of Mathematics and Its Applications*, volume 42. Cambridge University Press, Cambridge. First published 1993.

Hodges, W. (2006). Logic and games. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy (Summer 2006 Edition)*. `http://plato.stanford.edu/archives/sum2006/entries/logic-games/`.

Hyttinen, T. (1990). Model theory for infinite quantifier logics. *Fundamenta Mathematicæ*, 134:125–142.

Kamlah, W. and Lorenzen, P. (1973). *Logische Propädeutik*. Bibilographisches Institut, Mannheim.

Karttunen, M. (1984). Model theory for infinitely deep languages. *Annales Academiæ Scientiarum Fennicæ*, 50.

Kontchakov, R., Kurucz, A., and Zakharyaschev, M. (2005). Undecidability of first-order intuitionistic and modal logics with two variables. *Bulletin of Symbolic Logic*, 11(3):428–438.

Krynicki, M. and Mostowski, M. (1995). Henkin quantifiers. In Krynicki, M., Mostowski, M., and Sczerba, L. W., editors, *Quantifiers: Logics, Models and Computation*, volume 1, pages 193–262. Kluwer, Dordrecht.

Lorenz, K. (1961). *Arithmetik und Logik als Spiele*. Ph.D. thesis, Christian-Albrechts-Universität Zu Kiel.

Lorenz, K. (1970). *Elemente der Sprachkritik*. Suhrkamp, Frankfurt.

Lorenz, K. (2001). Basic objectives of dialogue logic in historical perspective. *Synthese*, 127:255–263.

Lorenzen, P. and Lorenz, K. (1978). *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt.

Lorenzen, P. and Schwemmer, O. (1975). *Konstruktive Logik, Ethik und Wissenschaftstheorie*. Bibilographisches Institut, Mannheim.

Makkai, M. (1977). Admissible sets and infinitary logic. In Barwise, J., editor, *Handbook of Mathematical Logic*, pages 233–281. North-Holland, Amsterdam.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT, Cambridge, MA.

Rahman, S. (1994). *Über Dialoge, Protologische Kategorien und andere Seltenheiten*. Peter Lang, Frankfurt.

Rahman, S. and Keiff, L. (2005). On how to be a dialogician. In Vanderveken, D., editor, *Logic, Thought and Action*, volume 2: Logic, Epistemology and Unity of Science, pages 359–408. Springer, Dordrecht.

Ranta, A. (1988). Propositions as games as types. *Synthese*, 76:377–395.

Saarinen, E. (1978). Dialogue semantics versus game-theoretical semantics. In *Proceedings of the Biennial Meeting of the Philosophy of Science Association (PSA)*, volume 2: Symposia and Invited Papers, pages 41–59. The University of Chicago Press, Chicago, IL.

Sandu, G. and Pietarinen, A.-V. (2001). Partiality and games: Propositional logic. *Logic Journal of the IGPL*, 9(1):107–127.

Sandu, G. and Pietarinen, A.-V. (2003). Informationally independent connectives. In Mints, G. and Muskens, R., editors, *Games, Logic, and Constructive Sets*, pages 23–41. CSLI, Stanford.

Schwalbe, U. and Walker, P. (2001). Zermelo and the early history of game theory. *Games and Economic Behaviour*, 34:123–137.

Skolem, Th. (1920). Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Skrifter utgit av Videnskabsselskapet i Kristiania, I. Matematisk-naturvidenskabelig klasse no. 4.*

Stegmüller, W. (1964). 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.

Sundholm, G. (2002). Proof theory and meaning. In Gabbay, D. M. and Guenthner, F., editors, *Handbook of Philosophical Logic*, volume 9, pages 165–198. Kluwer, Dordrecht, second edition.

Tarski, A. (1983). The concept of truth in the languages of the deductive sciences. In Corcoran, J., editor, *A. Tarski: Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, pages 152–278. Hackett, Indianapolis, IN. Polish; original in Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno–Fizycznych 34, Warsawm, 1933.

Tarski, A. and Vaught, R. L. (1956). Arithmetical extensions of relational systems. *Compositio Mathematica*, 13:81–102.

van Benthem, J. (2001a). Games in dynamic epistemic logic. *Bulletin of Economic Research*, 53(4):219–248. Proceedings LOFT-4, Torino, Bonanno, G. and van der Hoek, W., editors.

van Benthem, J. (2001b). *Logic and Games. Lecture Notes (Draft Version)*, ILLC, Amsterdam.

van Benthem, J. (2002). Extensive games as process models. *Journal of Logic, Language and Information*, 11:289–313.

Vaught, R. L. (1973). Descriptive set theory in $l_{\omega_1\omega}$. In Mathias, A. and Rogers, H., editors, *Cambridge Summer School in Mathematical Logic*, volume 337 of *Lecture Notes in Mathematics*, pages 574–598. Springer, Berlin.

von Neumann, J. and Morgenstern, O. (2004). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, sixtieth-anniversary edition. (First appeared in 1944.)

Chapter 9

REVISITING GILES'S GAME

*Reconciling Fuzzy Logic and Supervaluation**

Christian G. Fermüller
Technische Universität Wien
chrisf@logic.at

Abstract We explain Giles's characterization of Łukasiewicz logic via a dialogue game combined with bets on results of experiments that may show dispersion. The game is generalized to other fuzzy logics and linked to recent results in proof theory. We argue that these results allow one to place *t*-norm based fuzzy logics in a common framework with supervaluation as a theory of vagueness.

9.1 Introduction

Robin Giles (1974, 1977) has presented a strategic two-person game as a formal model of reasoning in physical theories, in particular quantum theory. Giles strictly separates the treatment of logical connectives from the problem of assigning meaning to atomic propositions in the presence of uncertainty. For the systematic stepwise reduction of arguments about compound statements to arguments about their atomic subformulas he refers to Paul Lorenzen's *dialogue game rules* (see, e.g., Lorenzen, 1960). Atomic formulas are interpreted as assertions about (yes/no-)results of elementary experiments with dispersion. (That is, the experiments may yield different results when repeated; only the *probability* of a particular answer is known.) Finally, it is stipulated that each player has to pay a fixed amount of money to the other player for every false atomic assertion. Giles discovered that the propositions that a player can assert initially in the sketched game without having to expect a loss of money on average coincide with those that are valid in Łukasiewicz logic Ł, a logic introduced for different purposes in the 1920s (Łukasiewicz, 1920).

*Partly supported by the FWF grants P16539–N04 and P18563–N12.
We wish to thank Helge Rückert, George Metcalfe and Ondrej Majer for valuable comments on a previous version of this paper.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

Giles's remarkable result dates back to 1974; in more recent years Ł has emerged as one of several fundamental *fuzzy logics*. (See, e.g., Hájek, 1998; Cignoli et al., 1999.) With hindsight, Giles has addressed an important philosophical challenge concerning vagueness: how to derive a 'fuzzy logic' from first principles of approximate reasoning? (For alternative approaches to this foundational problem, see, e.g., Cignoli et al., 1999; Paris, 1997; Ruspini, 1991.)

We aim at two different tasks. First, we want to place Giles's theorem in the context of recent results in the proof theory of fuzzy logics. In particular, we indicate how Giles's game can be generalized to other important fuzzy logics and point out that strategies in the corresponding games are related to analytic proofs in so-called *r*-hypersequent calculi (Ciabattoni et al., 2005). A second task arises from a seeming paradox: the game-theoretic characterization of fuzzy logics eliminates all reference to fuzziness. More exactly, instead of talking about degrees of truth one talks about probabilities of success of elementary experiments. So how does the game-based analysis of fuzzy logics relate to their degree-theoretic semantics? This question is of particular significance, since experts insist on the fundamental difference between probabilities (degrees of belief) on the one hand, and degrees of truth (reflecting vagueness) on the other. (See, e.g., Dubois and Prade, 1980; Hájek, 1998, 2002 for a clear and concise explication of this difference.)

More generally, one may ask whether the game-based analysis can shed light on the relation between truth-functional fuzzy logics and competing models of approximate reasoning. Considering the highly contentious discourse on vagueness in analytic philosophy, our aim, although limited, is rather ambitious. We claim that the relevant games provide a way to reconcile the intuitions behind two prominent, but seemingly contradicting theories of vagueness: namely the degree-theoretic approach and *supervaluation* with respect to admissible precisifications. We will interpret both approaches to vagueness as combining a *classical* analysis of logical connectives with a *non-classical* interpretation of the semantic status of atomic propositions. Towards this aim, we show that not only supervaluation, but also degree-based fuzzy logics can be analyzed in terms of admissible precisifications of vague propositions. The dramatic difference in the respective judgements on logical validity does not disappear, but will be seen to result from the different syntactic levels at which supervaluation and fuzzy logics, respectively, refer to precisifications.

This paper is organized as follows. We begin with a short review of *t*-norm based fuzzy logics, in particular Ł, P, and G (in Section 9.2). This is followed by a presentation of Giles's game for Ł (in Section 9.3). We then connect the game with recent results in the proof theory of fuzzy logics (Section 9.4) and

generalize these results to include the logics P, CHL, and G (Section 9.5).[1] This will leave us with the challenge of interpreting the game based characterization of fuzzy logics in terms of conceptions of vagueness (as explained in Section 9.6). To address this challenge, we connect (in Section 9.7) the semantic machinery of 'supervaluation' with that of *t*-norm-based fuzzy logics. In the conclusion (Section 9.8), we hint at further topics for research.

We point out that, mainly due to lack of space, we confine our investigations to propositional logic.

9.2 *t*-norm based fuzzy logics

Fuzzy logics arise by stipulating that, in the presence of vague notions and propositions, truth comes in degrees. This view is very controversial among philosophers of vagueness. (See, e.g., Keefe, 2000; Williamson, 1994; Keefe and Smith, 1987 for an overview of the vagueness discourse in analytic philosophy.) Although we think that serious reflections on the philosophical foundation of logical formalisms are unavoidable in judging their adequateness, one may profit from recognizing at the outset that the 'degrees of truth' approach leads to a mathematically sound, robust and non-trivial formalism. It is not our intention to enter the debate on the significance of mathematical models in philosophical logic here, but we subscribe explicitly to the view that *as broad as possible* a collection of mathematical structures and tools should be available to every expert—whether philosopher, logician, computer scientist, or technician—in the search for an adequate model of reasoning in a given context.

The degree-theoretic approach to approximate reasoning has motivated dozens of different formalisms. Following Petr Hájek (1998, 2002), we cite some 'design decisions' that lead to the definition of a family of logics worth exploring in this context:

1. The set of truth degrees (truth values) is represented by the real unit interval $[0, 1]$. The usual order relation $\leq$ serves as comparison of truth degrees; 0 represents absolute falsity, and 1 absolute truth.

2. The truth value of a compound statement shall only depend on the truth values of its subformulas. In other words: the logics are truth functional.

3. The truth function for conjunction ($\&$) should be a continuous, commutative, associative, and (in both arguments) monotonically increasing function $* : [0, 1]^2 \to [0, 1]$, where $0 * x = 0$ and $1 * x = x$. In other words: $*$ is a continuous *t*-norm.

[1] Sections 9.3–9.5 extend the brief remarks in the final section of Ciabattoni et al. (2005).

4. The residuum $\Rightarrow_*$ of the t-norm $*$—i.e., the unique function $\Rightarrow_*$: $[0,1]^2 \to [0,1]$ satisfying $x \Rightarrow_* y = \max\{z \mid x * z \leq y\}$—serves as the truth function for implication. The truth function for negation is defined as $\lambda x[x \Rightarrow_* 0]$. (Observe that this is analogous to the relation between conjunction, implication and negation in classical logic.)

The three most fundamental continuous t-norms and their residua are:

	t-norm	associated residuum
Łukasiewicz	$x *_{\text{Ł}} y = \max(0, x + y - 1)$	$x \Rightarrow_{\text{Ł}} y = \min(1, 1 - x + y)$
Gödel	$x *_{\text{G}} y = \min(x, y)$	$x \Rightarrow_{\text{G}} y = \begin{cases} 1 & \text{if } x \leq y \\ y & \text{otherwise} \end{cases}$
Product	$x *_{\text{P}} y = x \cdot y$	$x \Rightarrow_{\text{P}} y = \begin{cases} 1 & \text{if } x \leq y \\ y/x & \text{otherwise} \end{cases}$

Any continuous t-norm is an ordinal sum construction of these three (see, e.g., Hájek, 1998). Note that the minimum and maximum of two values that serve as alternative truth functions for conjunction ($\wedge$) and disjunction ($\vee$), respectively, can be expressed in terms of $*$ and $\Rightarrow_*$: $\min(x, y) = x*(x \Rightarrow_* y)$ and $\max(x, y) = \min((x \Rightarrow_* y) \Rightarrow_* y, (y \Rightarrow_* x) \Rightarrow_* x)$.

We arrive at the following definition of propositional logics associated with a continuous t-norm:

Definition 1. *For a continuous t-norm $*$ with residuum $\Rightarrow_*$, we define a logic $\mathbf{L}_*$ based on a language with binary connectives $\to$, $\&$, constant $\bot$, and defined connectives $\neg A =_{def} A \to \bot$, $A \wedge B =_{def} A \,\&\, (A \to B)$, $A \vee B =_{def} ((A \to B) \to B) \wedge ((B \to A) \to A)$. A* valuation *for $\mathbf{L}_*$ is a function v assigning to each propositional variable a truth value from the real unit interval* $[0,1]$*, uniquely extended to v^* for formulas by:*

$$v^*(A \,\&\, B) = v^*(A) * v^*(B), \quad v^*(A \to B) = v^*(A) \Rightarrow_* v^*(B), \quad v^*(\bot) = 0.$$

A formula A is valid in $\mathbf{L}_*$ iff $v^*(A) = 1$ for all valuations v^* pertaining to the t-norm $*$.

The logics $\mathbf{L}_{*_{\text{Ł}}}$, $\mathbf{L}_{*_{\text{G}}}$, and $\mathbf{L}_{*_{\text{P}}}$, are called Łukasiewicz logic Ł, Gödel logic G, and Product logic P, respectively. Computational properties as well as semantic aspects of these logics, including their relation to other important logics, are well studied. (Again, Hájek, 1998 is the standard reference.) Various corresponding proof systems have been presented. Below, we will refer to the very recent systems **HŁ** of Metcalfe, Olivetti, and Gabbay (Metcalfe et al., 2005) for Ł, and **rH** of Ciabattoni, Fermüller, and Metcalfe (Ciabattoni et al., 2005) that provides a uniform treatment of Ł, G, and P.

9.3 Giles's game for Ł

As already mentioned, Giles (1974, 1977) arrived at his analysis of Łukasiewicz logic irrespective of any reflections on vagueness or *t*-norms. His corresponding game consists of two largely independent building blocks:

1. Betting for positive results of experiments. Two players—let us say me and you—agree to pay 1€ to the opponent player for every false statement they assert. By $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ we denote an *elementary state* in the game, where I assert each of the q_i in the multiset $\{q_1, \ldots, q_n\}$ of atomic statements (i.e., propositional variables), and you, likewise, assert each atomic statement $p_i \in \{p_1, \ldots, p_m\}$.

Each propositional variable q refers to an experiment E_q with binary (yes/no) result. The statement q can be read as 'E_q yields a positive result'. Things get interesting as the experiments may show dispersion; i.e., the same experiment may yield different results when repeated. However, the results are not completely arbitrary: for every run of the game, a fixed *risk value* $\langle q \rangle^r \in [0, 1]$ is associated with q, denoting the probability that E_q yields a negative result.[2]

For the special atomic formula $\bot$ (*falsum*) we define $\langle \bot \rangle^r = 1$. The risk associated with a multiset $\{p_1, \ldots, p_m\}$ of atomic formulas is defined as $\langle p_1, \ldots, p_m \rangle^r = \sum_{i=1}^{m} \langle p_i \rangle^r$. The risk $\langle \rangle^r$ associated with the empty multiset is defined as 0. The risk associated with an elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ is calculated from my point of view. Therefore the condition $\langle p_1, \ldots, p_m \rangle^r \geq \langle q_1, \ldots, q_n \rangle^r$ expresses that I do not expect any loss (but possibly some gain) when betting on the truth of atomic statements, as explained above.

2. A dialogue game for the reduction of compound formulas. Giles follows Paul Lorenzen (see, e.g., Lorenzen, 1960) in constraining the meaning of connectives by reference to rules of a dialogue game that proceeds by systematically reducing arguments about compound formulas to arguments about their subformulas.

For brevity, we will assume that formulas are built up from propositional variables, the falsity constant φ, and the connective $\rightarrow$ only.[3] The central dialogue rule can be stated as follows:

[2] Giles (1977) attempts to provide a tangible meaning to the basic notions that arise in formalizing physical theories. Our use of Giles's game is independent from this original motivation.

[3] Note that in Ł all other connectives can be defined from $\rightarrow$ and φ alone, since we may define $A \,\&\, B$ as $(A \rightarrow (B \rightarrow \varphi)) \rightarrow \varphi$. The other connectives are defined as indicated in Definition 1.

(R) If I assert $A \to B$ then, whenever you choose to attack this statement by asserting A, I have to assert also B. (And vice versa, i.e., for the roles of me and you switched.)

This rule reflects the idea that the meaning of implication is specified by the principle that an assertion of 'if A, then B' ($A \to B$) obliges one to assert B, if A is granted.

In contrast to dialogue games for intuitionistic logic (Lorenzen, 1960; Felscher, 1985; Krabbe, 1988; Fermüller, 2003a), no special regulations on the succession of moves in a dialogue are required here. However, we assume that each assertion is attacked at most once: this is reflected by the removal of $A \to B$ from the multiset of all formulas asserted by a player during a run of the game, as soon as the other player has either attacked by asserting A, or has indicated that she will not attack $A \to B$ at all. Note that every run of the dialogue game ends in an elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$. Given an assignment $\langle\cdot\rangle^r$ of risk values to all p_i and q_i we say that I *win* the corresponding run of the game if I do not expect any loss, i.e., if $\langle p_1, \ldots, p_m\rangle^r \geq \langle q_1, \ldots, q_n\rangle^r$.

As an almost trivial example consider the game where I initially assert $p \to q$ for some atomic formulas p and q; i.e., the initial state is $[\|p \to q]$. In response, you can either assert p in order to force me to assert q, or explicitly refuse to attack $p \to q$. In the first case, the game ends in the elementary state $[p\|q]$; in the second case it ends in state $[\|]$. If an assignment $\langle\cdot\rangle^r$ of risk values gives $\langle p\rangle^r \geq \langle q\rangle^r$, then I win, whatever move you choose to make. In other words: I have a winning strategy for $p \to q$ in all assignments of risk values where $\langle p\rangle^r \geq \langle q\rangle^r$.

Theorem 2 (R. Giles (1974, 1977))**.** *Every assignment $\langle\cdot\rangle^r$ of risk values to atomic formulas occurring in a formula F induces a valuation $v_{\langle\cdot\rangle^r}$ for Łukasiewicz logic Ł such that $v_{\langle\cdot\rangle^r}(F) = 1$ iff I have a winning strategy for F in the game presented above.*

Corollary 3. *F is valid in Ł iff for all assignments of risk values to atomic formulas occurring in F, I have a winning strategy for F.*

9.4 Connecting strategies and proofs

There is a well-known correspondence between winning strategies in dialogue games and cut-free proofs in adequate versions of Gentzen's sequent calculus. For the case of Lorenzen's original dialogue game and (a variant of) Gentzen's **LJ** for intuitionistic logic this has been demonstrated, for example, in Felscher (1985). A similar, even more straightforward relation holds between Gentzen's **LK** and Lorenzen-style dialogue games for classical logic. Game-based characterizations have been presented for many other logics, including modal logics, paraconsistent logics and substructural logics. To name

just one result of relevance to our context, a correspondence between *parallel* versions of Lorenzen's game and so-called hypersequent calculi for intermediary logics, including the fuzzy logic G, has been established in Fermüller (2003a) and Ciabattoni and Fermüller (2003).

Returning to the game presented in Section 9.3, we note that Giles proved Theorem 2 without formalizing the concept of strategies. However, to reveal the close relation to analytic proof systems we need to define structures that allow us to formally register possible choices for both players. These structures, called *disjunctive strategies* or, for short, *d-strategies* appear at a different level of abstraction to strategies. The latter are only defined with respect to given assignments of risk values (and may be different for different assignments), whereas *d*-strategies abstract away from particular assignments.

Definition 4. *A d-*strategy *(for me) is a tree whose nodes are disjunctions of states:*

$$[A_1^1, \ldots, A_{m_1}^1 \| B_1^1, \ldots, B_{n_1}^1] \bigvee \ \ldots \ \bigvee [A_1^k, \ldots, A_{m_k}^k \| B_1^k, \ldots, B_{n_k}^k]$$

which fulfill the following conditions:

1. *All leaf nodes denote disjunctions of elementary states.*
2. *Internal nodes are partitioned into I-nodes and you-nodes.*
3. *Any I-node is of the form* $[A \to B, \Gamma \| \Delta] \bigvee \mathcal{G}$ *and has exactly one successor node of the form* $[B, \Gamma \| \Delta, A] \bigvee [\Gamma \| \Delta] \bigvee \mathcal{G}$*, where* $\mathcal{G}$ *denotes a (possibly empty) disjunction of states, and* Γ, Δ *denote (possibly empty) multisets of formulas.*
4. *For every state* $[\Gamma \| \Delta]$ *of a you-node and every occurrence of* $A \to B$ *in* Δ*, the you-node has a successor of the form* $[A, \Gamma \| B, \Delta'] \bigvee \mathcal{G}$ *as well as a successor of the form* $[\Gamma \| \Delta'] \bigvee \mathcal{G}$*, where* Δ' *is* Δ *after removal of one occurrence of* $A \to B$*. (The multiset of occurrences of implications at the right hand sides is non-empty in you-nodes.)*[4]

We call a d-strategy winning *(for me) if, for all leaf nodes* ν *and for all possible assignments of risk values to atomic formulas, there is a disjunct* $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ *in* ν*, such that* $\langle p_1, \ldots, p_m \rangle^r \geq \langle q_1, \ldots, q_n \rangle^r$.

In game theory, a winning strategy (for me) is usually defined as a function from all possible states, where I have a choice, into the set of my possible moves. Note that winning strategies in the latter sense exist for all assignments of risk values if and only if a winning *d*-strategy exists.

[4]For a total of n occurrences of compound formulas on the right-hand sides of states in a you-node, there are $2n$ successor nodes, corresponding to $2n$ possible moves for you.

Strictly speaking, we have only defined d-strategies (and therefore, implicitly, also strategies) with respect to some given regulation that, for each possible state, determines who is to move next. Each consistent partition of internal nodes into I-nodes and you-nodes corresponds to such a regulation. However, it has been demonstrated by Giles (1974, 1977) that the order of moves is irrelevant for determining my expected gain. Therefore no loss of generality is involved here.

The defining conditions for I-nodes and you-nodes clearly correspond to possible moves for me and you, respectively, in the dialogue game. Thus Giles's theorem can be reformulated in terms of d-strategies. More interestingly, conditions 3 and 4 also correspond to the introduction rules for implication in the hypersequent calculus **HŁ** for Ł, defined in Metcalfe et al. (2005).

Hypersequents, due to Pottinger and Avron (Avron, 1991), are a natural and useful generalization of Gentzen's sequents. A hypersequent is just a multiset of sequents written as

$$\Gamma_1 \vdash \Delta_1 \mid \cdots \mid \Gamma_n \vdash \Delta_n$$

The interpretation of component sequents $\Gamma_i \vdash \Delta_i$ varies from logic to logic. But the |-sign separating the individual components is always interpreted as a classical disjunction (at the meta-level). The logical rules for introducing connectives refer to single components of a hypersequent. The only difference to sequent rules is that the relevant sequents live in a (possibly empty) context $\mathcal{H}$ of other sequents, called side-hypersequent. The rules of **HŁ** for introducing implication are:

$$\frac{B, \Gamma \vdash \Delta, A \mid \mathcal{H}}{A \to B, \Gamma \vdash \Delta \mid \mathcal{H}}\ (\to, l) \qquad \frac{A, \Gamma \vdash \Delta, B \mid \mathcal{H} \qquad \Gamma \vdash \Delta \mid \mathcal{H}}{\Gamma \vdash \Delta, A \to B \mid \mathcal{H}}\ (\to, r)$$

Observe that rules $(\to, l)$ and $(\to, r)$ are just syntactic variants of the defining conditions 3 and 4 for d-strategies. To sum up: the logical rules of **HŁ** can be read as rules for constructing generic winning strategies in Giles's game.

9.5 Other fuzzy logics: variants of the game

We have shown that a formalization of generic strategies for Giles's game (d-strategies) reveals a direct correspondence with the hypersequent system **HŁ** for Ł. What about other fuzzy logics? Can one generalize the discovered correspondence to include P, G, and related logics?

Giles's characterization of Ł combines Lorenzen-style dialogue rules for the analysis of connectives with bets on positive results of elementary experiments. But note that the phrase 'betting for positive results of (a multiset of) experiments' is ambiguous. As we have seen, Giles identifies the combined risk for such a bet with the *sum* of risks associated with the single experiments. However, other ways of interpreting the combined risk are worth exploring. In particular, we are interested in a second version of the game, where an elementary

state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ corresponds to my single bet that *all* experiments associated with the q_i, where $1 \leq i \leq n$, show a positive result, against your single bet that *all* experiments associated with the p_i $(1 \leq i \leq m)$ show a positive result. A third form of the game arises if one decides to perform only *one* experiment for each of the two players, where the relevant experiment is chosen by the opponent.

To achieve a direct correspondence between the three mentioned betting schemes and the t-norm based semantics of the connectives in Ł, P, and G, respectively, we invert risk values into probabilities of *positive* results (yes-answers) of the associated experiments. More formally, the *value* of an atomic formula q is defined as $\langle q \rangle = 1 - \langle q \rangle^r$; in particular, $\langle \bot \rangle = 0$.

My expected gain in the elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ in Giles's game for Ł is the sum of money that I expect you to have to pay me minus the sum that I expect to have to pay you. This amounts to $\sum_{i=1}^{m}(1 - \langle p_i \rangle) - \sum_{i=1}^{n}(1 - \langle q_i \rangle)$ €. Therefore, my expected gain is greater or equal to zero iff $1 + \sum_{i=1}^{m}(\langle p_i \rangle - 1) \leq 1 + \sum_{i=1}^{n}(\langle q_i \rangle - 1)$ holds. The latter condition is called winning condition $W_{\sum}$.[5]

In the second version of the game, you have to pay me 1 € unless all experiments associated with the p_i test positively, and I have to pay you 1 € unless all experiments associated with the q_i test positively. My expected gain is therefore $1 - \prod_{i=1}^{m}\langle p_i \rangle - (1 - \prod_{i=1}^{n}\langle q_i \rangle)$ €; the corresponding winning condition $W_{\prod}$ is $\prod_{i=1}^{m}\langle p_i \rangle \leq \prod_{i=1}^{n}\langle q_i \rangle$.

To maximize the expected gain in the third version of the game I will choose a $p_i \in \{p_1, \ldots, p_m\}$ where the probability of a positive result of the associated experiment is least; and you will do the same for the q_i's that I have asserted. Therefore, my expected gain is $(1 - \min_{1 \leq i \leq m}\langle p_i \rangle) - (1 - \min_{1 \leq i \leq n}\langle q_i \rangle)$ €. Hence the corresponding winning condition $W_{\min}$ is $\min_{1 \leq i \leq m}\langle p_i \rangle \leq \min_{1 \leq i \leq n}\langle q_i \rangle$.

We thus arrive at the following definitions for the value of a multiset $\{p_1, \ldots, p_n\}$ of atomic formulas, according to the three versions of the game:

$$\langle p_1, \ldots, p_n \rangle_{Ł} = 1 + \sum_{i=1}^{n}(\langle p_i \rangle - 1) = (\sum_{i=1}^{n}\langle p_i \rangle) - (n - 1)$$
$$\langle p_1, \ldots, p_n \rangle_{P} = \prod_{i=1}^{n}\langle p_i \rangle$$
$$\langle p_1, \ldots, p_n \rangle_{G} = \min_{1 \leq i \leq n} \langle p_i \rangle .$$

For the empty multiset we define $\langle \rangle_{Ł} = \langle \rangle_{P} = \langle \rangle_{G} = 1$.

In contrast to Ł, the dialogue game rule (R) does not suffice to characterize P and G. To see this, consider the state $[p \to \bot \| q]$. According to rule (R) I may assert p in order to force you to assert $\bot$. Since $\langle \bot \rangle = 0$, the resulting elementary state $[\bot \| p, q]$ fulfills the winning conditions $\langle \bot \rangle \leq \langle p \rangle \cdot \langle q \rangle$ and

[5] The term 'winning condition' is slightly misleading here, since I can lose money in a particular run of the game even if this condition holds; only a non-negative *expected* gain is guaranteed by what we choose to call 'winning condition'.

$\langle\bot\rangle \leq \min\{\langle p\rangle, \langle q\rangle\}$, that correspond to P and G, respectively. However, this is at variance with the fact that for assignments where $\langle p\rangle = 0$ and $\langle q\rangle < 1$ you have asserted a statement $(p \to \bot)$ that is definitely true ($v(p \to \bot) = 1$), whereas my statement q is not definitely true ($v(q) < 1$).[6]

It is no accident that the above example involves the falsity constant $\bot$ as well as a value $\langle p\rangle = 0$. If we remove $\bot$ from the language and evaluate formulas as in P—using multiplication for conjunction and its residuum for implication—but over the left-open interval $(0, 1]$ instead of $[0, 1]$, then we arrive at a well-investigated logic, known as *cancellative hoop logic* CHL (see, e.g., Esteva et al., 2003; Metcalfe et al., 2004).[7]

It is easy to check[8] that the logical rules of system **HŁ** are sound and invertible not only for Ł, but also for CHL. Therefore, we can directly transfer the connection, described in Section 9.4, between d-strategies for Giles's game and **HŁ**-rules, to obtain the following.

Corollary 5. *F is valid in* CHL *iff for all assignments of values from* $(0, 1]$ *to the atomic formulas occurring in F, I have a winning strategy for F in the variant of Giles's game with the winning condition $W_{\prod}$.*

We have thus arrived at a game based characterization of CHL, which uses dialogue rules identical to those for Ł, but differs in the betting schemes determining the winning conditions.

We may justify the elimination of $\bot$ and 0 by the observation that the presence of elementary experiments, which *invariably* yield a negative result, spoils the whole idea of combining bets on positive results according to the schemes for P or G. On the other hand, however, the expressiveness of the language is considerably reduced by removing $\bot$, since negation is defined in terms of $\bot$. One may ask whether there is a characterization of P and G by a Giles/Lorenzen-style game. To address this problem we analyze the rules of the uniform calculus **rH** (Ciabattoni et al., 2005), mentioned at the end of Section 9.2. In contrast to **HŁ**, the component sequents of hypersequents in **rH** come in two versions: there are two different sequent signs, '$\leq$' and '$<$', instead of the one '$\vdash$' used in **HŁ**. More formally, an r-hypersequent is a finite multiset

$$\Gamma_1 \lhd_1 \Delta_1 \mid \ldots \mid \Gamma_n \lhd_n \Delta_n$$

where $\lhd_i \in \{<, \leq\}$ and Γ_i and Δ_i are finite multisets of formulas for $i = 1, \ldots, n$. The relational symbols indicate the intended semantics: the above

[6] The problem does not arise in logic Ł, since there the expected gain for state $[\bot\|p, q]$ is $\langle p, q\rangle_Ł - \langle\bot\rangle_Ł = 1 - (\langle p\rangle - 1) - (\langle q\rangle - 1) - (1 - 1) = \langle p\rangle + \langle q\rangle - 1$ and therefore, indeed, negative, as expected, if $\langle p\rangle = 0$ and $\langle q\rangle < 1$.

[7] Note that CHL is different from $\bot$-free P: e.g., $(A \to (A \,\&\, B)) \to B$ is valid in CHL, but not in P.

[8] For the rule $(\to, l)$ it suffices to observe that for all $a, b, c_i, d_j \in (0, 1]$: $(a \Rightarrow_P b)\cdot\prod_i g_i \leq \prod_j d_j$ iff $b\cdot\prod_i g_i \leq \prod_j d_j \cdot a$. For the rule $(\to, r)$ the relevant fact is that $\prod_i g_i \leq \prod_i d_i \cdot (a \Rightarrow_P b)$ iff both $a \cdot \prod_i g_i \leq \prod_j d_j \cdot b$ and $\prod_i g_i \leq \prod_j d_j$.

r-hypersequent is called *valid* for logic $\mathsf{X} \in \{\text{Ł},\mathsf{G},\mathsf{P}\}$ if for all valuations v, that refer to the corresponding t-norm $*_{\mathsf{X}}$, there is some i, $1 \leq i \leq n$, such that $\#^v_{\mathsf{X}}\Gamma_i \lhd_i \#^v_{\mathsf{X}}\Delta_i$, where $\#^v_{\mathsf{X}}\emptyset = 1$ and where

$$\#^v_{\text{Ł}}(\Gamma) = 1 + \sum_{A\in\Gamma}\{v(A) - 1\}, \quad \#^v_{\mathsf{G}}(\Gamma) = \min_{A\in\Gamma}\{v(A)\}, \quad \#^v_{\mathsf{P}}(\Gamma) = \prod_{A\in\Gamma}\{v(A)\}.$$

This allows us to check that the following **rH**-rules for introducing implication are sound and invertible for all three logics:

$$\frac{A, \Gamma \lhd \Delta, B \mid A \leq B \mid \mathcal{H} \qquad \Gamma \lhd \Delta \mid \mathcal{H}}{\Gamma \lhd \Delta, A \to B \mid \mathcal{H}}\ (\to, r)^*$$

$$\frac{B, \Gamma \lhd \Delta, A \mid \Gamma \lhd \Delta \mid \mathcal{H} \quad \Gamma \lhd \Delta \mid B < A \mid \mathcal{H}}{A \to B, \Gamma \lhd \Delta \mid \mathcal{H}}\ (\to, l)^*$$

where $\lhd$ is either $<$ or $\leq$, uniformly in each rule. Together with (also uniform, even simpler) rules for the other connectives and appropriate initial atomic r-hypersequents (that, of course, are different for each of the three logics) one obtains a sound and complete analytic system for Ł, G, and P, respectively (see Ciabattoni et al., 2005).

There are at least two different ways to translate these rules into rules for the construction of winning strategies in versions of our game. A rather direct interpretation of r-hypersequents in terms of disjunctions of states in a dialogue game is obtained by distinguishing two different types of states: One, corresponding to the sequent sign $\leq$, which is exactly as in the original game, and one corresponding to the sequent sign $<$, in which an additional flag ¶ is raised to announce that I will be declared the winner of the current run of the game, only if the evaluation of the final elementary state yields a *strictly positive* (and not just non-negative) expected gain for me.

Dialogue rules, replacing (R) in Giles's game, but directly corresponding to $(\to, r)^*$ and $(\to, l)^*$ can be formulated as follows:

(R^*_r) If I assert $A \to B$ then, whenever you choose to attack this statement by asserting A, I have the following choice: either I assert B in reply or I challenge your attack on $A \to B$ by replacing the current game with a new one in which you assert A and I assert B.

Note that the right-hand side premise of the rule $(\to, l)^*$ corresponds to the case in which you choose not to attack the exhibited occurrence of $A \to B$. As can be seen, the newly introduced flag plays no direct role. It is only needed in the rule corresponding to $(\to, l)^*$:

(R^*_l) If you assert $A \to B$ then, whenever I choose to attack this statement by asserting A, you have the following choice: either you assert B in reply

> or you challenge my attack on $A \to B$ by replacing the current game with a new one in which the flag ¶ is raised and I assert A while you assert B.

Note that I can also choose not to attack $A \to B$. This corresponds to the component sequents $\Gamma \triangleleft \Delta$ in the two premise r-hypersequents of rule $(\to, r)^*$. The flag ¶ is needed because the winning conditions are not fully complementary for me and you: we may both have a non-negative expected gain. Your 'attack-challenging' claim that I *cannot win* when starting in state $[A||B]$ is equivalent to the claim that I *can win* when starting in state $[B||A]$ only if the flag ¶, signalling a strictly positive expected gain as winning condition, is raised in the latter game.

The translation of the r-hypersequent rules for conjunction and disjunction in Ciabattoni et al. (2005) into dialogue game rules is also straightforward. Admittedly, these new versions of Lorenzen-style dialogue rules amount to ad hoc regulations to circumvent the problematic effects of bets on elementary results that always yield negative results. A different (but still ad hoc) way to deal with this problem has been described in Ciabattoni et al. (2005). Instead of using the additional flag, one imposes the following constraint on attacking implicative formulas:

> (Q) If I have a strategy for winning the run of the game starting in the state $[A||B]$, then I am not allowed to attack your assertion of $A \to B$. (And vice versa, i.e., for the roles of you and me switched.)[9]

Imposing (Q) also results in a game that characterizes Ł, P, and G, if the corresponding versions of the winning conditions are applied (cf. Ciabattoni et al., 2005). Here we only point out that applying rule (Q) involves the systematic development of full strategies for subformulas, before it can be judged whether an attack to a formula according to rule (R) is permitted. Whether more satisfying Giles/Lorenzen style characterizations of P and G in the presence of $\bot$ and 0 can be achieved remains an open problem.

9.6 Where is vagueness?

What has been achieved by the analysis of fuzzy logics in terms of dialogue games? Note that the rules for the stepwise reduction of arguments about compound formulas to arguments about their atomic subformulas are the same for Ł, CHL, P, and G. This opens a *unified view* of reasoning in t-norm based fuzzy logics. Moreover, the relation to classical logic is clarified: the dialogue part of the game coincides with a version of Lorenzen's original dialogue game

[9] Recall that the strategies mentioned in (Q) refer to a given assignment $\langle\cdot\rangle$ of values and thus appear at a more concrete level than d-strategies.

adapted to classical logic. If we trivialize the betting schemes by stipulating that all assigned probabilities are either 0 or 1—i.e., if each elementary experiment consistently shows the same result when repeated—then Giles's game, as well as the alternative games for P and G, discussed in Section 9.5, characterize *classical validity*. To see this, it suffices to check that for every elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ and $\mathsf{X} \in \{\text{Ł}, \mathsf{P}, \mathsf{G}\}$ we have:

$$\langle p_1, \ldots, p_m \rangle_{\mathsf{X}} \leq \langle q_1, \ldots, q_n \rangle_{\mathsf{X}} \quad \text{iff} \quad \{p_1, \ldots, p_m\} \cap \{q_1, \ldots, q_n\} \neq \emptyset,$$

for all assignments $\langle \cdot \rangle$ of values where $\langle p_i \rangle, \langle q_j \rangle \in \{0, 1\}$. If we denote elementary states in sequent notation

$$p_1, \ldots, p_m \vdash q_1, \ldots, q_n$$

it is clear that the latter condition corresponds to classical axiom sequents $p \vdash p$, up to (irrelevant) weakening. Moreover, it corresponds to the standard winning condition for Lorenzen style dialogues: I win the dialogue if you attack a statement that you have already asserted yourself ('*ipse dixisti* rule'). Indeed, it is straightforward to show that the winning of *d*-strategies for all versions of the game, described above, corresponds to cut-free proofs in versions of hypersequent calculi that are sound and complete for classical logic, if valuations are restricted to range over $\{0, 1\}$.

What is the significance of the betting schemes for the evaluation of atomic formulas? Obviously, the betting schemes allow us to characterize the *differences* between Ł, P, and G: different underlying *t*-norms correspond to different ways of combining bets on results of elementary experiments into a single bet. However, a closer look at this setting reveals a serious foundational problem. One would like to present the game based characterizations of Ł, CHL, P, and G as a derivation of fuzzy logics from first principles about reasoning with *vague propositions*, but all reference to vagueness and degrees of truth seems to have disappeared. More exactly: it has been replaced by references to classical reasoning combined with a *probabilistic* semantics for atomic statements. However, fuzziness should never be confused with probability (as has been emphasized in the literature, e.g., in Hájek, 1998, 2002; Dubois and Prade, 1980). Whereas fuzzy logic takes vague propositions to refer to *degrees of truth*, probability theory formalizes *degrees of rational belief*. Even without engaging in discussions on adequate interpretations of vagueness and probability, it should be clear that

1. 'The next throw of the die will result in 5 or 6'

is true only with some probability (1/3, if the die is fair), but does not involve vagueness; whereas,

2. 'Logicians are weird people'

is vague, but does not refer to probability. (2) may meaningfully be said to be true only to some degree (even in a fixed context), whereas (1), in the intended context, is either definitely true or definitely false, even if it is (not yet) known which of the two holds. Since Giles, in evaluating atomic statements, refers to elementary experiments that are of the same (probabilistic, but non-fuzzy) type as in statement (1), it might seem inadequate to interpret Giles's game as a model for proper reasoning with vague notions.

9.7 Connecting supervaluation, degrees of truth, and bets on positive results of experiments

There is prolific discourse in analytic philosophy about the nature of reasoning in the presence of vagueness. This is not the place to comment on these debates;[10] however, in order to connect the game-based analysis of Ł, P, and G with degrees of truth and to disentangle it from probabilistic logic, we refer to a particular approach to understanding vagueness, called *supervaluationism*—currently most popular among philosophers of vagueness (see, e.g., Keefe, 2000; Varzi, 2001; Weatherson, 2003).

Supervaluationism, as a theory of vagueness, is canonically developed by Kit Fine (1975). Since we are only interested in propositional logic without additional modal operators, only a simplified version of supervaluation will be needed here. The central idea is to formalize reasoning in vague contexts by reference to all *admissible precisifications* of vague expressions. More exactly, formulas are evaluated in reference to a *specification space* $\mathcal{S}$, which is simply a collection (multiset)[11] of partial models. A partial model is a possibly partial assignment of classical truth values, 0 or 1, to propositional variables. An element $w \in \mathcal{S}$ is called a *complete precisification* of $v \in \mathcal{S}$ if w is total and if $v(p) = w(p)$ for all propositional variables p, for which v is defined. A complete precisification of v is a classical model compatible with v. We are only interested in those elements of $\mathcal{S}$ that are complete precisifications of a fixed element ('actual world') $a \in \mathcal{S}$. This sub-multiset of $\mathcal{S}$ is denoted by C_a and is assumed to be non-empty. Three possibilities for the semantic status of a formula F arise:

- $v(F) = 1$ for all $v \in C_a$: in this case F is called *supertrue* in C_a.
- $v(F) = 0$ for all $v \in C_a$: in this case F is called *superfalse* in C_a.

[10] For an overview of theories of vagueness and their problematic relation to fuzzy logic we refer to Keefe (2000), Williamson (1994), Burns (1991) and Fermüller (2003b).

[11] As long as one is not interested in measuring the cardinality of precisifications that satisfy certain properties, the difference between precisifications spaces as sets and as multisets, respectively, disappears.

- $\exists v, w \in C_a$ such that $v(F) = 0$ and $w(F) = 1$: in this case the semantic status of F remains undefined.

Proponents of supervaluationism often contend that, in contrast to claims made by the degree-theorists, no revision of classical logic is necessary to cope with vagueness. (However, see Kremer and Kremer, 2003 for a criticism of the claim that supervaluationism does not deviate from classical logic.) Whereas, for example, the formula $A \vee \neg A$ is not valid in Ł, P, G, and related logics, it is evaluated true in all classical interpretations, and therefore is supertrue in all precisification spaces $\mathcal{S}$, even if A were evaluated true in some precisifications and false in other precisifications of the actual world of $\mathcal{S}$.

Given the coincidence of supertruth in all specification spaces and classical validity, it is understandable that supervaluationism is usually seen as incompatible with fuzzy logic. In contrast, we claim that the game-based interpretation reveals much common ground among these competing conceptions of reasoning under vagueness. Both supervaluationists and defendants of Ł, P, and G as logics of vagueness can agree on three principles:

1. An atomic statement is *definitely true* only if there is no admissible precisification of it that renders it false.

2. Arguments about compound statements F can be reduced to arguments involving only subformulas of F.

3. The rules used for (2) should only depend on the outmost connective of F and should be sound and complete for classical logic.

That the reduction rules should refer to classical logic seems, at a first glance, to be at variance with the standard t-norm-based interpretation of our fuzzy logics. However, the coincidence of the logical dialogue rules in Giles's game with those in versions of the game for classical logic makes shared intuitions about the meaning of connectives explicit.

Obviously, essential differences between supervaluation and t-norm based valuations remain. To facilitate a more detailed comparison, we interpret the truth value $\in [0, 1]$ that is assigned to a propositional variable p in fuzzy valuation in terms of the *proportion* of those complete precisifications that make p true. The simplest way to formalize this idea is to assume that the cardinality of $C_a \in \mathcal{S}$ is finite. We may then define the 'fuzzy valuation' $v_{\mathcal{S}}$ induced by a precisification space $\mathcal{S}$ via C_a as

$$v_{\mathcal{S}}(p) = \frac{|\{v \in C_a : v(p) = 1\}|}{|C_a|}$$

for all propositional variables p. In other words: with respect to a given precisification space, fuzzy valuations and supervaluation *agree* on the assignment of

classical truth values 1 and 0 to atomic formulas; but in the remaining cases, where supervaluation assigns no overall truth value, fuzzy logics assign a value that 'measures' the fraction of verifying precisifications.[12] For compound formulas, the difference between supervaluation and fuzzy valuation may be described in terms of the *syntactic level* at which a formula is tied to individual precisifications. For *supervaluation* the whole formula F is evaluated in each complete precisification to determine F's semantic status. Following the game-based characterization of Ł, P, and G, *fuzzy valuation* of F may be described as consisting of three stages:

1. An analysis—following classical principles—of F into arguments about its atomic components
2. A valuation of each of the resulting relevant occurrences of atomic formulas in F in reference to a specification space
3. A synthesis of the resulting values of the atomic subformulas of F into an overall value for F

The following table confronts the valuation function $v_{\mathcal{S}}^{sv}$ resulting from supervaluation with the valuation function $v_{\mathcal{S}}^{\mathsf{X}}$ of a t-norm based fuzzy logic X, where all valuations refer to the specification space $\mathcal{S}$ via the multiset C_a of its complete precisifications.

Supervaluation	**Valuation in logic X**
$v_{\mathcal{S}}^{sv}(p) = 1\,(0) \Longleftrightarrow \forall v \in C_a\colon v(A) = 1\,(0)$	$v_{\mathcal{S}}^{\mathsf{X}}(p) = \frac{\lvert\{v \in C_a\colon v(p)=1\}\rvert}{\lvert C_a\rvert}$
$v_{\mathcal{S}}^{sv}(\bot) = 0$	$v_{\mathcal{S}}^{\mathsf{X}}(\bot) = 0$
$v_{\mathcal{S}}^{sv}(F \to G) = 1\,(0) \Longleftrightarrow$ $\forall v \in C_a\colon (v(F) \Rightarrow_c v(G)) = 1\,(0)$	$v_{\mathcal{S}}^{\mathsf{X}}(F \to G) = (v_{\mathcal{S}}^{\mathsf{X}}(F) \Rightarrow_* v_{\mathcal{S}}^{\mathsf{X}}(G))$, where $\mathsf{X} = \mathbf{L}_*$

Remember that $\Rightarrow_*$ is the residuum of the t-norm $*$ that defines the logic $\mathbf{L}_*$. We have used $\Rightarrow_c$ to denote the classical truth function for implication (which, by the way, can be presented as the residuum of an arbitrary t-norm, restricted to $\{0, 1\}$). Also remember that all other logical connectives can be defined in terms of $\to$ and $\bot$, not only for Ł, but also for classical logic. Of course, one can easily extend the above list by the corresponding definitions for conjunction and disjunction (thus including also full P and G).

[12] At the propositional level, on which we focus here, it is not unreasonable to assume that only a finite number of different plausible precisifications is relevant when evaluating a given statement in a fixed context. Anyway, it is not difficult to extend the concept to more general situations. For example, one may wish to weight precisifications according to some measure of their individual plausibility. One may also take into account non-complete precisifications in different ways. In any case, an assignment of a truth value $\in [0, 1]$ to a propositional variable p in fuzzy logic can be interpreted as a way to *quantify the information* pertaining to p that is contained in a given specification space.

We think that supervaluation and fuzzy valuation capture contrasting, but individually coherent intuitions about the role of logical connectives in vague statements. Consider a sentence like

3. "The sky is blue and is not blue".

When formalized as $b \wedge \neg b$, (3) is *superfalse* in all specification spaces. This fits Fine's motivation (Fine, 1975) to capture 'penumbral connections' that prevent any mono-colored object from having two colors at the same time. According to his intuition the statement "The sky is blue" absolutely contradicts the statement "The sky is not blue", even if neither statement is definitely true or definitely false. Therefore (3) is judged as definitely false, even if admittedly vague. On the other hand, by asserting (3) one may intend to convey the information that both component assertions are true only to some degree. Under this reading (and a certain interpretation of 'and') $b \wedge \neg b$ is *not* definitely false, unless b is supertrue or superfalse. The latter intuition is directly captured in Łukasiewicz logic since $b \wedge \neg b$ may receive a value $\in [0, 0.5]$, where $\wedge$ denotes the 'weak conjunction', i.e., the minimum operator.[13]

As already indicated, the difference between the two interpretations of (3) can be described as a difference of the syntactic level at which the sentence is projected to admissible precisifications. In supervaluation it is checked whether the *whole sentence* $b \wedge \neg b$ is true in every complete precisification, whereas in fuzzy valuation each of the two occurrences of the subformula b is valuated separately with respect to the proportion of complete precisifications that make b true.

We claim that both kinds of intuitions should be accommodated in a full account of approximate reasoning.[14] Technically, supervaluation and various forms of fuzzy valuation can easily be embedded in a common semantic framework, as indicated above. For evaluating a formula F correspondingly, it suffices to mark syntactically—e.g., by using two different types of implication, conjunction, negation, etc.—whether an occurrence of a subformula of F should be supervaluated or valuated according to a certain t-norm-based scheme. In both cases, the valuation may refer to the same specification space.

9.8 Conclusion

Our presentation of Giles's game and its variants is meant to demonstrate that t-norm based fuzzy logics can be derived from first principles about approximate reasoning. As we have seen in Sections 9.4 and 9.5, rules for the

[13] Note that $b \,\&\, \neg b$ is always evaluated to 0, where $\&$ is the 'strong conjunction' (t-norm) of Ł. Thus one may argue that Ł is capable of representing both interpretations of a sentence like (3). Also remember that in P and G the value of $\neg b$ is 0 if the value of b is not equal to 1. Therefore $b \wedge \neg b$ is always evaluated to 0 in P and G.

[14] This is of particular significance for a successful analysis of *Sorites paradoxa* and the phenomena of *higher-order vagueness*, as we shall argue elsewhere.

systematic construction of winning strategies in the games for Ł, CHL, P, and G correspond to the logical rules of analytic calculi for these logics. This also partly clarifies the relation to classical logic: for all investigated logics the (dialogue-based) meaning of connectives adheres to constraints pertaining to classical logic. Moreover, the game-based analysis allows us to relate supervaluation to the seemingly opposite concept of 'degrees of truth': both models of approximate reasoning can be seen as referring to admissible precisifications in a given specification space.

Many interesting topics for further investigation arise. We conclude by explicitly posing a few relevant questions. Is there a similar analysis of other logics that have been suggested for approximate reasoning? In particular, can Hájek's 'basic logic' (Hájek, 1998)—the logic of *all* continuous t-norms—be characterized by an adequate game? What about quantifiers? How does the incompleteness of first-order Ł and P that contrasts with the existence of complete calculi for G (and classical logic), bear on the game-based semantics for these logics? How can we account for higher-order vagueness in dialogue games? Can one extend the analysis to logics equipped with a definiteness operator and other relevant modal operators? Can the game based characterization of fuzzy logics shed light on the relation to further conceptions of vagueness, like gap-theoretic, epistemic, pragmatic and information-based approaches?

References

Avron, A. (1991). Hypersequents, logical consequence and intermediate logics for concurrency. *Annals of Mathematics and AI*, 4(3–4):225–248.

Burns, L. C. (1991). *Vagueness: An Investigation Into Natural Language and the Sorites Paradox*. Kluwer, Dordrecht.

Ciabattoni, A. and Fermüller, C. G. (2003). From intuitionistic logic to gödel-dummett logic via parallel dialogue games. In *33rd Intl. Symp. on Multiple-Valued Logic*, pages 188–195. IEEE Computer Society Press, Tokyo.

Ciabattoni, A., Fermüller, C. G., and Metcalfe, G. (2005). Uniform rules and dialogue games for fuzzy logics. In *Logic for Programming, Artificial Intelligence, and Reasoning, LPAR 2004*, Springer LNAI 3452, 496–510, Dordrecht.

Cignoli, R., D'Ottaviano, I. M. L., and Mundici, D. (1999). *Algebraic Foundations of Many-Valued Reasoning*, volume 7 of *Trends in Logic*. Kluwer, Dordrecht.

Dubois, D. and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications*. Academic, New York.

Esteva, F., Godo, L., Hájek, P., and Montagna, F. (2003). Hoops and fuzzy logic. *Journal of Logic and Computation*, 13(4):532–555.

Felscher, W. (1985). Dialogues, strategies, and intuitionistic provability. *Annals of Pure and Applied Logic*, 28:217–254.

Fermüller, C. G. (2003a). Parallel dialogue games and hypersequents for intermediate logics. In *TABLEAUX 2003*, volume 2796 of *LNAI*, pages 48–64. Springer, Dordrecht.

Fermüller, C. G. (2003b). Theories of vagueness versus fuzzy logic: can logicians learn from philosophers? *Neural Network World Journal*, 13(5):455–466.

Fine, K. (1975). Vagueness, truth and logic. *Synthèse*, 30:265–300.

Giles, R. (1974). A non-classical logic for physics. *Studia Logica*, 4(33):399–417.

Giles, R. (1977). A non-classical logic for physics. In Wojcicki, R. and Malinkowski, G., editors, *Selected Papers on Łukasiewicz Sentential Calculi*, pages 13–51. Polish Academy of Sciences, Wrocław - Warszawa - Kraków - Gdańsk.

Hájek, P. (1998). *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht.

Hájek, P. (2002). Why fuzzy logic? In Jackquette, D., editor, *A Companion to Philosophical Logic*, pages 595–606. Blackwell, Oxford.

Keefe, R. (2000). *Theories of Vagueness*. Cambridge University Press, Cambridge.

Keefe, R. and Smith, P., editors (1987). *Vagueness: A Reader*. MIT Press, Cambridge, MA.

Krabbe, E. C. W. (1988). Dialogue sequents and quick proofs of completeness. In Hoepelman, J. P., editor, *Representation and Reasoning*, pages 135–140. Max Niemeyer Verlag, Tübingen.

Kremer, P. and Kremer, M. (2003). Some supervaluation-based consequence relations. *Journal of Philosophical Logic*, 32(3):225–244.

Lorenzen, P. (1960). Logik und agon. In *Atti Congr. Internat. di Filosofia*, volume 4, pages 187–194, Sansoni, Firenze.

Łukasiewicz, J. (1920). O logice tròjwartościowej. *Ruch Filozoficzny*, 5:169–171.

Metcalfe, G., Olivetti, N., and Gabbay, D. M. (2004). Analytic calculi for product logics. *Archive for Mathematical Logic*, 43(7):859–889.

Metcalfe, G., Olivetti, N., and Gabbay, D. M. (2005). Sequent and hypersequent calculi for Abelian and Łukasiewicz logics. To appear in ACM TOCL. Available at `http://www.dcs.kcl.ac.uk/pg/metcalfe/`.

Paris, J. (1997). A semantics for fuzzy logic. *Soft Computing*, 1:143–147.

Ruspini, E. H. (1991). On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5:45–88.

Varzi, A. (2001). Vagueness, logic, and ontology. *The Dialogue*, 1:135–154.

Weatherson, B. (2003). Many many problems. *Philosophical Quarterly*, 53:481–501.

Williamson, T. (1994). *Vagueness*. Routledge, London.

Chapter 10

IMPLICIT VERSUS EXPLICIT KNOWLEDGE IN DIALOGICAL LOGIC

Manuel Rebuschi
L. H. S. P. – Archives H. Poincaré
Manuel.Rebuschi@univ-nancy2.fr

Abstract A dialogical version of (modal) epistemic logic is outlined, with an intuitionistic variant. Another version of dialogical epistemic logic is then provided by means of the S4 mapping of intuitionistic logic. Both systems cast new light on the relationship between intuitionism, modal logic and dialogical games.

10.1 Introduction

Two main approaches to *knowledge* in logic can be distinguished (van Benthem, 1991). The first one is an implicit way of encoding knowledge and consists in an epistemic interpretation of usual propositional or first-order logic. This is, for instance, the case of intuitionistic logics, especially of the so-called Brouwer-Heyting-Kolmogorov (BHK) interpretation of it. There, assertion is assimilated to provability, negation to the provability of the implication of a contradiction, etc. The second approach is what is known, since Hintikka's seminal work (Hintikka, 1962), as (modal) epistemic logic. In this case, knowledge is explicitly supported by modal operators.

The aim of the present paper is to show the specific insight provided by dialogical games on this distinction. In Section 10.2, I will introduce dialogical versions of classical and intuitionistic Propositional Logic (PL) and a dialogical version of modal epistemic logic. In Sections 10.3 and 10.4, two combinations of implicit and explicit epistemic logics are accounted for: intuitionistic modal logic, and a modal embedding of intuitionistic logic. In Section 10.5, other issues connected with the implementation of epistemic logic in the dialogical frame are raised and briefly discussed.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

10.2 Dialogical epistemic logic (DEL) in a nutshell

Thanks to a straightforward adaptation of Rahman and Rückert's Dialogical Modal Logic (Rahman and Rückert, 1999), one obtains a Dialogical Epistemic Logic (hereafter DEL). For that purpose, several kinds of rules have to be stated: structural and particle rules for propositional logic and for modal epistemic logic. As will be shown, modal logic only requires a simple extension of rules for propositional logic.

10.2.1 Propositional logic

In a dialogical game, two players argue about a thesis: The proponent **P** defends it against the attacks of the opponent **O**. As in game semantics, something interesting appears when the proponent has a winning strategy, i.e. when she can defend the proposition against any attack from the opponent. Here the interesting result is that one is guaranteed that the proposition is logically true or valid—whereas in GTS, for instance, the existence of a winning strategy means that the challenged proposition is true simpliciter (materially true) in a given model.

Particle Rules. The meaning of each logical constant is given through a particle rule which determines how to attack and defend a formula whose main connective is the constant in question. The set, **PartRules**, of particle rules for disjunction, conjunction, implication and negation is recapitulated in the following table:

	Attack	**Defence**
$A \vee B$	?	A, or B (The defender chooses)
$A \wedge B$	$?_L$, or $?_R$ (The attacker chooses)	A, or B (respectively)
$A \rightarrow B$	A	B
$\neg A$	A	$\otimes$ (No possible defence)

The idea for disjunction is that the proposition $A \vee B$, when asserted by a player, is challenged by the question "*Which one?*"; the defender has then to choose one of the disjuncts and to defend it against any new attack. The rule is the same for the conjunction $A \wedge B$, except that the choice is now made by the attacker: "*Give me the left conjunct* ($?_L$)" or "*Give me the right one* ($?_R$)", and the defender has to assume the conjunct chosen by his or her challenger. For the conditional $A \rightarrow B$, the attacker assumes the antecedent A and the defender

continues with B. Finally negated formulas are attacked by the cancellation of negation, and cannot be defended. The defender in this case can thus only counterattack (if he or she can).

Structural Rules. In addition to the particle rules connected to each logical constant, one also needs structural rules to be able to play in such and such a way at the level of the whole game.

- (PL-0) **Starting Rule:** The initial formula (the *thesis* of the dialogical game) is asserted by **P**. Moves are numbered and alternatively uttered by **P** and **O**. Each move after the initial utterance is either an attack or a defence.
- (PL-1) **Winning Rule:** Player **X** wins iff it is **Y**'s turn to play and **Y** cannot perform any move.
- (PL-2) **No Delaying Tactics Rule:** Both players can only perform moves that change the situation.
- (PL-3) **Formal Rule:** (*In a given context*[1]) **P** cannot introduce any new atomic formula; new atomic formulas must be stated by **O** first. Atomic formulas can never be attacked.

These four rules are common to dialogical games for both classical and intuitionistic logics. The only difference resides in the following rule:

- (PL-4c) **Classical Rule:** In any move, each player may attack a complex formula uttered by the other player or defend him/herself against *any attack* (including those that have already been defended).
- (PL-4i) **Intuitionistic Rule:** In any move, each player may attack a complex formula uttered by the other player or defend him/herself against *the last attack that has not yet been defended*.

Now we can build two distinct sets of rules **DialPLc** and **DialPLi**, yielding respectively classical propositional logic and intuitionistic propositional logic:

$$\textbf{DialPLc} := \textbf{PartRules} \cup \{\text{PL-0, PL-1, PL-2, PL-3, PL-4c}\}$$
$$\textbf{DialPLi} := \textbf{PartRules} \cup \{\text{PL-0, PL-1, PL-2, PL-3, PL-4i}\}$$

For any set of rules Σ, I will use the notation $\Sigma \vDash A$ to say that there is a winning strategy for the proponent in the dialogical game about A played according to the rules of Σ. As PL-4i is more restrictive than PL-4c, we have for any propositional formula: $\textbf{DialPLi} \vDash A \Rightarrow \textbf{DialPLc} \vDash A$.

[1] In propositional logic contexts are not yet defined—this will be useful for dialogical games for modal logics.

As is shown in Rahman (1993), **DialPLi** $\vDash A$ iff A is intuitionistically valid, whereas **DialPLc** $\vDash A$ iff A is valid in classical logic. The difference between classical and intuitionistic logic is thus reducible to one structural rule, PL-4.

Example 1. As a first example of a dialogical game for propositional logic, let us consider a formula that is valid according to both classical and intuitionistic logic: $((a \to b) \wedge a) \to b$. In the dialogical frame, it means that there is a winning strategy for the proponent **P** when she plays according to both sets of rules. The rounds and the corresponding arguments, attacks or defences, are indicated by a number within brackets (n) in the external columns, whereas the arguments attacked by the players are referred to by their number m in the internal column. Defences are on the lines of the corresponding attacks. The reader can check the following game and see what is the winning strategy employed by **P**:

O			**P**		
				$((a \to b) \wedge a) \to b$	(0)
(1)	$(a \to b) \wedge a$	0		b	(8)
(3)	$a \to b$		1	$?_L$	(2)
(5)	a		1	$?_R$	(4)
(7)	b		3	a	(6)

Having stated the thesis (0), **P** cannot simply defend it against **O**'s first attack (1) since she would have to assert b which is an atom not yet stated by the opponent. But **P** can counterattack twice, with (2) and (4), and **O** is forced to defend himself with (3) and (5) respectively. Thanks to (5), **P** can use a at round (6) and attack **O**'s round (3) to oblige him to answer b (7). Now b is available to **P** who can answer the first attack and win the play (no further move being permitted for the opponent).

Example 2. The second example is provided by the dialogical games associated with the formula: $\neg\neg a \to a$. As can be expected, we will get **DialPLc** $\vDash \neg\neg a \to a$, but **DialPLi** $\nvDash \neg\neg a \to a$:

O			**P**		
				$\neg\neg a \to a$	(0)
(1)	$\neg\neg a$	0		a	(4)
	$\otimes$		1	$\neg a$	(2)
(3)	a	2		$\otimes$	

The difference between the games becomes manifest after round (3). Following the intuitionistic rule (PL-4i), the proponent should defend herself against the last attack not yet defended, i.e. against (3); but she cannot, since (3) is an

attack against a negation leaving no available defence. By contrast, according to the classical rule, the proponent can defend herself against a former attack of the opponent, so she can answer (4) to (1), and win the play.

10.2.2 (Modal) epistemic logic

As PL, also Modal Logic requires the introduction of particle and structural rules corresponding to the additional operators. We will moreover need a convention to designate the different contexts (or possible worlds) where propositions are stated by both players.

Particle Rules. The thesis of the dialogue is uttered in a given context w. The particle rules for the epistemic operator K and for its dual P enable the players to change the context.

	Attack	**Defence**
KA (in context w)	$?_{K/w'}$ (The attacker chooses an available context w')	A (in context w')
PA (in context w)	$?_P$	A (in an available context w' chosen by the defender)

Context numbering.

- The initial context is numbered 1. The n immediate successors of m are numbered $m.1, m.2, \ldots, m.n$.
- An immediate successor $m.n$ of a context m is said to be of rank $+1$ relative to m, and m is said to be of rank -1 relative to its immediate successors. A successor $m.n.p$ of a context m is said to be of rank $+2$ relative to m, etc.

Structural Rules. Modal structural rules correspond to restrictions on the accessibility relation $\mathcal{K}$ between contexts (and thus determine which contexts are available to players). The first two rules are obviously incompatible and should not be included together in the same set of rules.[2]

- (ML-frc) **Formal Rule for Contexts: P** *cannot introduce* a new context; new contexts must be introduced by **O**.

[2]Here I follow Rahman and Rückert's formulation of rules associated with specific modal *systems*. Structural rules could also be formulated in accordance with the specific *axioms* involved in those systems: the upshot would be the same.

- (ML-D) **Axiom D rule: P** *can introduce* a new context of rank +1 relative to the context she is playing in.

- (ML-K) **K Rule: P** cannot stay in the context she is playing in (as she attacks a formula of the form KA or defends a formula of the form PA). **P** can choose a (given) context of rank +1 relative to the context she is playing in.

- (ML-T) **T Rule: P** can either choose a (given) context of rank +1 relative to the context she is playing in, or stay in the context she is playing in.

- (ML-B) **B Rule: P** can either choose a (given) context of rank $-1/+1$ relative to the context she is playing in, or stay in the context she is playing in.

- (ML-S4) **S4 Rule: P** can either choose any (given) context of rank $+k$ relative to the context she is playing in, or stay in the context she is playing in.

- (ML-S5) **S5 Rule: P** can choose any (given) context.

Dialogical Epistemic Systems. Combining these new rules with those of **DialPLc**, one obtains sets of rules corresponding to different usual systems of propositional modal logic:

DialK	:=	**DialPLc** ∪ {ML-frc, ML-K}
DialD	:=	**DialPLc** ∪ {ML-D, ML-K}
DialT	:=	**DialPLc** ∪ {ML-frc, ML-T}
DialB	:=	**DialPLc** ∪ {ML-frc, ML-B}
DialS4	:=	**DialPLc** ∪ {ML-frc, ML-S4}
DialS5	:=	**DialPLc** ∪ {ML-frc, ML-S5}

Example. Let us consider a substitution instance of the Positive Introspection Property (also known as Axiom **4**): $K\phi \rightarrow KK\phi$, to be played according to **DialS4**. What is interesting here is the fact that the proponent resorts to the transitivity of $\mathcal{K}$ at round (6)—if the game was played in a non-transitive structure, there would be no more winning strategies available to **P**.

	O			**P**			
				$Ka \rightarrow KKa$	(0)		1
1	(1)	Ka	0	KKa	(2)		1
1	(3)	$?_{K/1.1}$	2	Ka	(4)		1.1
1.1	(5)	$?_{K/1.1.1}$	4	a	(8)		1.1.1
1.1.1	(7)	a		1 $?_{K/1.1.1}$	(6)		1.1.1

10.3 Intuitionistic DEL

Two kinds of intuitionistic epistemic logics can be provided using the dialogical frame. The first one is Intuitionistic Modal Logic to be sketched in this section. The second one is a dialogical version of the modal 'simulation' of intuitionistic logic, to be presented in the next section.

10.3.1 Intuitionistic modal logic

Within the Dialogical frame, Rahman and Rückert (1999) suggest just to change **DialPLc** into **DialPLi** in the set of structural rules. For instance, an intuitionistic version of **S5** is directly obtained by replacing (PL-4c) by (PL-4i): **DialS5i** := **DialPLi** ∪ {ML-frc, ML-S5}. For any dialogical system of modal logic **DialΣ**, I will use the notation **DialΣi** to designate the corresponding intuitionistic version obtained in this way.

Is it really intuitionistic modal logic? Such dialogical systems are obtained through a simple combinatorial step. It can be doubted that they yield "real" intuitionistic modal logics. Let $\mathcal{L}_M$ be the standard propositional language augmented by the modal connectives in a set M. Wolter and Zakharyaschev's general definition of an intuitionistic modal logic $\mathcal{L}$ in $\mathcal{L}_M$ is as follows (Wolter and Zakharyaschev, 1999): (1) $\mathcal{L} \subset \mathcal{L}_M$; (2) $\mathcal{L}$ contains propositional intuitionistic logic; (3) $\mathcal{L}$ is closed under: (i) Modus ponens, (ii) Substitution, (iii) Regularity Rule ($A \to B \,/\, \bigcirc A \to \bigcirc B$, for every $\bigcirc \in M$).

In the dialogical frame, it is easily seen that conditions (1) and (2) are automatically filled with the relations holding between the corresponding sets of rules **DialΣ** and **DialΣi**. Now, one can simply check that any dialogical epistemic system **DialΣi** is closed under the Regularity Rule:

	O			**P**		
				$A \to B$	(0)	1
1	(1)	A	0	B	(2)	1
		...		...		

$A \to B$ is valid iff there is a winning strategy for **P** to end this game.

	O			**P**		
				$KA \to KB$	(0)	1
1	(1)	KA	0	KB	(2)	1
1	(3)	$?_{K/1.1}$	2	B	(6)	1.1
1.1	(5)	A		1 $?_{K/1.1}$	(4)	1.1
		...		...		

	O			**P**		
				$PA \to PB$	(0)	1
1	(1)	PA	0	PB	(2)	1
1	(3)	$?_P$	2	B	(6)	1.1
1.1	(5)	A		1 $?_P$	(4)	1
		...		...		

What follows immediately from the last two dialogues is that they are to be continued in the same way as that corresponding to $A \to B$. In other words: if there is a winning strategy for **P** in the dialogue associated to $A \to B$, then there is one for the corresponding dialogue associated to $KA \to KB$ (or $PA \to PB$).

Moreover, the dialogical system **DialKi** at least encompasses Fischer Servi's intuitionistic modal logic **FS** (Fischer Servi, 1977), and the same for the corresponding extensions **S4**, **S5**, and so forth. (For more details, see the Appendix.)

10.3.2 Application to epistemic modalities

Knowledge Generalization. The necessitation rule **KG** (i.e., $\vdash A$ / $\vdash KA$), applies only to intuitionistic validities, not to classical ones:

	O			**P**		
				$K(a \vee \neg a)$	(0)	1
1	(1)	$?_{K/1.1}$	0	$(a \vee \neg a)$	(2)	1.1
		...		...		

After (2), the play goes on (in context 1.1) according to **DialPLi**. Hence **O** wins! Intuitionistic dialogical epistemic systems thus account for (explicit) knowledge of intuitionist agents.

Intuitionistic K and P. In intuitionistic modal systems **DialΣi**, K and P become genuine intuitionistic modal operators: they are no longer interdefinable.

For instance, according to **DialTi** (i.e., with a reflexive accessibility relation), one can see that: $\neg K\neg A \not\approx PA$—the following dialogue stops at round (7)—, whereas (as is expected): $\neg K\neg A \approx PA$ within **DialT**—the play goes on.

	O			**P**		
				$\neg Ka \to P\neg a$	(0)	1
1	(1)	$\neg Ka$	0	$P\neg a$	(2)	1
1	(3)	$?_P$	2	$\neg a$	(4)	1
1	(5)	a	4	$\otimes$		
		$\otimes$		1 Ka	(6)	1
1	(7)	$?_{K/1.1}$	6	a	(**10**)	1.1
	(3')	$?_P$	2	$\neg a$	(**8**)	1.1
1.1	(**9**)	a	8	$\otimes$		

It also can be shown that other properties of K and P still hold in **DialTi**, such as the Consistency Property (D)[3]:

	O			**P**			
				$Ka \rightarrow Pa$		(0)	1
1	(1)	Ka	0	Pa		(2)	1
1	(3)	$?_P$	2	a		(6)	1
1	(5)	a		1	$?_{K/1}$	(4)	1

Advantages of Intuitionistic DEL. To conclude this section, let us mention a few features of Intuitionistic DEL which make it a good tool for epistemic logic:

1. Intuitionistic DEL provides an interesting account of modalities K and P: ignoring a no longer implies considering $\neg a$ as a possibility.

2. Implicit epistemic logic is made explicit: the epistemic agent is described as an intuitionist agent (thanks to the aforementioned restriction of KG to intuitionistic valid formulas). One could change the rules of the underlying propositional logic (e.g., for more strictly constructive ones) and obtain a corresponding *explicit* epistemic version in the same straightforward manner.

3. With the intuitionistic operators K and P, not only the described agent but the describing one too is (implicitly) grasped as a cognitive agent. This may be illustrated by the rejection of the tertium non datur: $KA \vee \neg KA$, in **DialS5i**:

	O			**P**			
				$Ka \vee \neg Ka$		(0)	1
1	(1)	?	0	$\neg Ka$		(2)	1
1	(3)	Ka	2	$\otimes$			
1	(5)	a		3	$?_{K/1}$	(4)	1

10.4 Modal simulation of intuitionistic (non-modal) logic

Let us turn again to implicit epistemic logic, namely intuitionistic propositional logic **Int**. In the first section, I presented the system **DialPLi** which is the usual dialogical implementation of **Int**. In this section, I will propose a

[3]The Consistency Property is not valid according to **DialKi**, as can be seen at round (4): with no reflexive accessibility relation, **P** cannot choose the current context to attack **O**'s assertion of Ka.

new dialogical formulation of intuitionistic logic, grounded on Gödel's 1933 **S4** embedding of **Int**, and on Kripke's 1965 modal semantics for **Int**.

Such a formulation is based on a dynamic conception of knowledge: The accessibility relation between contexts corresponds to time and to the growth of information—in contrast to Hintikka's 1962 "static" conception of this relation.

10.4.1 Gödel's embedding and Kripke's semantics

Gödel's translation of Int into S4. The idea of Gödel's embedding is closely related to the BHK interpretation of intuitionistic logic. When a formula is known (one could say: 'proved'), it will persist through time. The underlying idea is that of an ever increasing knowledge with neither memory failure nor revision. Formally, the **S4**-translation A^T of an intuitionistic formula A is as follows:

$$\begin{array}{rcl} a^T & := & \Box a, \text{ for every atomic formula } a \\ (A \wedge B)^T & := & (A^T \wedge B^T) \\ (A \vee B)^T & := & (A^T \vee B^T) \\ (\neg A)^T & := & \Box\neg A^T \\ (A \rightarrow B)^T & := & \Box(A^T \rightarrow B^T) \end{array}$$

This translation leads to the expected equivalence: $\vDash_{\mathbf{Int}} A$ iff $\vDash_{\mathbf{S4}} A^T$.

Kripke's modal semantics for Int. Kripke's structures (Kripke, 1965) involve a reflexive and transitive relation $\leq$ between contexts. The idea is similar to Gödel's translation: there is a temporal ordering of worlds, propositions being established once and for all whenever they are. So before being known, a proposition is not true and neither is its negation.

Formally, a Kripke structure is thus a tuple $\mathcal{K} = \langle W, \leq, \Vdash\rangle$, where: (1) $\leq$ is a pre-ordering on W (i.e., a binary reflexive and transitive relation); (2) the forcing relation $\Vdash$ is such that: (2.1) For all $w \in W$, $w \nVdash \bot$ (2.2) For all $w, w' \in W$, if $w \leq w'$ and $w \Vdash a$, then $w' \Vdash a$ (where a is an atomic formula). The forcing relation is then extended to complex formulas according to the following requirements: (i) $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$; (ii) $w \Vdash A \vee B$ iff $w \Vdash A$ or $w \Vdash B$; (iii) $w \Vdash A \rightarrow B$ iff $\forall w' \in W$, if $w \leq w'$ and $w' \Vdash A$, then $w' \Vdash B$; (iv) $w \Vdash \neg A$ iff $\forall w' \in W$, if $w \leq w'$ then $w' \nVdash A$.

Now we have the following equivalence:

$$\vDash_{\mathbf{Int}} A \text{ iff } \mathcal{K} \Vdash A \text{ for any Kripke model } \mathcal{K}.$$

10.4.2 Dialogical simulation of Int

Can the modal simulation of **Int** be implemented in dialogical games? The idea is to consider propositional intuitionistic formulas as if they were modal formulas. A dialogical version of **S4** can be formulated thanks to the usual

structural rules. Of course, there will be a restriction on the formulas: we reach a **S4**-like dialogical version for the propositional fragment only (i.e., not for formulas including modal operators).

However, the usual **S4** structural rules are not enough for this implementation: we must take into account the *non*-standard interpretation of atoms, negation and implication. Eventually, our system $\mathbf{Int}^{\mathbf{S4}}$ will essentially differ from **S4** at the level of *particle rules*.

Atoms. According to Gödel's translation: $a^T := \Box a$ for every atomic formula a. Let us consider the following part of game involving $\Box a$ and played with **DialS4**:

	O			**P**			
m	(j)	$\Box a$			...		
		...			...		
					$\Box a$	(k)	n
n	(k+1)	$?_{\Box/n.1}$	k		a	(k+4)	$n.1$
$n.1$	(k+3)	a		j	$?_{\Box/n.1}$	(k+2)	m
		...			...		

Player **P** can defend her assertion of $\Box a$ only if $\Box a$ has been previously introduced by **O** in *any* context $m \leq n$ (thanks to the transitivity of $\leq$ in **S4**). So (**S4** translations of) **Int** atoms can be attacked since they are modal formulas. We could thus add a special particle rule for atoms in $\mathbf{Int}^{\mathbf{S4}}$, stating that if an atom a is asserted in a context m, then it can be attacked by $?_n$ where the attacker chooses an available context $n \geq m$, and defended by the assertion of a in the context n. Consequently, the formal rule (PL-3) should be modified to enable players to attack atomic formulas.

But we do not need to change the particle rule for atomic formula. A simple look at the situation makes it clear that the whole modification can be restricted to one structural rule:

- (PL-3*) $\mathbf{Int}^{\mathbf{S4}}$ **Formal Rule:** In a given context n **P** cannot introduce any new atomic formula that has not been introduced by **O** in any context $m \leq n$; new atomic formulas must be stated by **O** first. Atomic formulas can never be attacked.

Negation. Gödel's translation $(\neg A)^T := \Box\neg A^T$ indicates that a negated formula asserted in a context m can be challenged in any context $n \geq m$:

	O			**P**		
	...			...		
				$\Box\neg A$	(k)	m
m	(k+1)	$?_{\Box/m.1}$	k	$\neg A$	(k+2)	$m.1$
$m.1$	(k+3)	A	k+2	$\otimes$		
	...			...		

This leads naturally to the following Particle rule for negation in $\mathbf{Int}^{\mathbf{S4}}$:

	Attack	**Defence**
$\neg A$ (in context m)	A (in an available context $n \geq m$ chosen by the attacker)	$\otimes$

Implication. The case of implication $(A \to B)^T := \Box(A^T \to B^T)$ is similar to that of negation:

	O			**P**		
	...			...		
				$\Box(A \to B)$	(k)	m
m	(k+1)	$?_{\Box/m.1}$	k	$A \to B$	(k+2)	$m.1$
$m.1$	(k+3)	A	k+2	B	(k+4)	$m.1$
	...			...		

It leads to the following modified Particle rule in $\mathbf{Int}^{\mathbf{S4}}$:

	Attack	**Defence**
$A \to B$ (in context m)	A (The attacker chooses an available context $n \geq m$)	B (in context n)

Recapitulation. To sum up our new system, let us denote the new set of particle rules by $\mathbf{PartRules}^{\mathbf{S4}}$: it is thus identical to **PartRules** concerning conjunction and disjunction, and differs on implication and negation.

Now we get a new set of rules:

$$\mathbf{Int}^{\mathbf{S4}} := \mathbf{PartRules}^{\mathbf{S4}} \cup \{\text{ PL-0, PL-1, PL-2, PL-3*, PL-4c}\} \cup \{\text{ML-frc, ML-S4}\}$$

which is equivalent to **DialPLi** in the following sense:

$$\mathbf{Int}^{\mathbf{S4}} \vDash A \iff \mathbf{DialPLi} \vDash A$$

for any propositional formula A.

Example 1. $\mathbf{Int}^{\mathbf{S4}} \nvdash a \vee \neg a$

	O			**P**		
				$a \vee \neg a$	(0)	1
1	(1)	?	0	$\neg a$	(2)	1
1.1	(3)	a	2	$\otimes$		

Example 2. $\mathbf{Int}^{\mathbf{S4}} \vDash \neg\neg(a \vee \neg a)$

	O			**P**		
				$\neg\neg(a \vee \neg a)$	(0)	1
1.1	(1)	$\neg(a \vee \neg a)$	0	$\otimes$		
		$\otimes$		1 $a \vee \neg a$	(2)	1.1
1.1	(3)	?	2	$\neg a$	(4)	1.1
1.1.1	(5)	a	4	$\otimes$		
		$\otimes$		1 $a \vee \neg a$	(6)	1.1.1
1.1.1	(7)	?	6	a	(8)	1.1.1

10.5 Discussion

In Game-Theoretical Semantics (GTS), one can distinguish between two types of "knowledge" depending on whether one is concerned about the interpretation of the epistemic operators (i.e., the usual meaning of "knowledge") or about the knowledge of the players of evaluation games. Van Benthem (2001) strongly stresses the epistemic features involved in GTS and IF Logic, even though such an interpretation of imperfect information games is absent from Hintikka's original creation (about IF Logic, see Hintikka and Sandu, 1997).

A similar distinction can be made in the dialogical frame, between explicit knowledge (that is embedded in the operators) and implicit knowledge (of the players). However, the distinction is not exactly the same since the players of dialogical games are not assumed to have particular information sets at their disposal, but a set of *action rules*. The intuitionistic restriction (PL-4i) imposed on the set of classical rules for **PL** therefore leads to a modeling of an abstract agent with limited (i.e., intuitionistic) epistemic powers.

In the above sections, two competing accounts of knowledge using dialogical games are provided. The two implementations resort to both implicit and explicit knowledge. They consist of specific combinations of intuitionistic and modal logics. Whereas systems **DialΣi** inoculate an intuitionistic variation to standard dialogical systems of modal logic, $\mathbf{Int}^{\mathbf{S4}}$ (implicitly) involves a modal interpretation of intuitionistic logic.

The two approaches presented in this paper could be extended in several ways. Among the possible developments of what should be called in general Dialogical Epistemic Logic, we highlight the following:

1. As was already stressed, $\mathbf{Int}^{\mathbf{S4}}$ is only concerned with the propositional fragment of **S4**. One could easily consider modal extensions $\mathbf{Int}^{\mathbf{S4}}_{\mathbf{K}}$ of it, using the underlying Kripke-like modal structure, and consider assertions about the **S4**-knowledge of an intuitionistic agent, such as, for example, $\neg K(a \vee \neg a)$.

2. Dialogical systems of doxastic logic, for example the system **DialKD45**. Doxastic logic in the dialogical frame starts like a nice story. The structural rule (ML-D) which separates the logic of belief from the logic of knowledge specifically enables the proponent to *create* new contexts... In doxastic logic too, intuitionistic variations could be easily implemented.

3. Multi-agent epistemic logic: such a development would be naturally grounded in multi-modal dialogical logic.

4. Non-Normal Logics. These "deviant" modal logics are due to Lemmon (1957) and Kripke (1965). They are based on the rejection of the axiom **K** and/or of the Necessitation Rule (**KG** in epistemic logic: $\vdash A / \vdash KA$). A motivation for adopting such a logic is to escape logical omniscience. Several authors have supplied epistemic logics using non-normal logic:

   - Duc (1997) uses normal action modalities and non-normal epistemic modalities, where **KG** is replaced by: $\vdash A / \vdash \langle F_i \rangle KA$, $\langle F_i \rangle$ being a dynamic action-temporal modality.
   - Thomason's theory (Thomason, 2000) is a combination of a normal (**KD45**) multi-agent frame and of a non-normal (**E2**) intra-agent frame (between subagents).

   Rahman and Keiff's recent proposal of a dialogical implementation of non-normal logics (Rahman 2003; Rahman and Keiff, 2004) leads to an immediate epistemic interpretation. Their main idea consists in considering a kind of meta-modal logic, i.e., a frame in which it is possible to consider different modal systems *together*. For example, an intuitionist logician might want to consider the (possible) case where *tertium non datur* were valid in his or her logic: the case in question amounts to a context where another logic is assumed to hold. Transposing it to epistemic logic enables one to consider the following cases:

   - Standard interpretation of a deviant agent in multi-modal epistemic logic, e.g. $K_i K_j A$ where the agent j is crazy and the agent i is sane.
   - Deviant interpretations of standard epistemic agents.
   - Shifting positive introspection (this is made by a crazy agent who knows he or she is crazy, like in $K_i^{sane} K_i^{crazy} A$).

Let us give a quick illustration of this non-normal frame with the dialogical game of the formula $K^cK^i(a \vee \neg a)$—stating that one knows classically that she knows intuitionistically that a or not a, which is surely false. This is established according to **S0.5**:

	O			**P**		
				$KK(a \vee \neg a)$	(0)	1-PLc
1-PLc	(1)	$?_{K/1.1}$	0	$K(a \vee \neg a)$	(2)	1.1-PLc
1.1-PLc	(3)	$?_{K/1.1.1-PLi}$	2	$a \vee \neg a$	(4)	1.1.1-PLi
1.1.1-PLi	(5)	?	4	$\neg a$	(6)	1.1.1-PLi
1.1.1-PLi	(7)	a	6	$\otimes$		

(An underlying logic, here **DialPLc** or **DialPLi**, is associated to each context. For explanations of such dialogues, see the cited papers.)

10.6 Conclusion

The original formulation of dialogical logic by Lorenzen and Lorenz (1978) was strongly related to intuitionistic logic. However, this connection may be enriched as is shown by our implementation of Gödel's **S4** embedding of **Int**. With the system $\mathbf{Int}^{\mathbf{S4}}$ one gets a new version of intuitionistic logic where the interpretation is directly linked to the connectives, at the level of the particle rules—like in the BHK interpretation.

In this paper I have sketched two systems of epistemic logic, conceptually very different but implementationally very close. While presenting systems **DialΣi**, we have seen that genuine intuitionistic modal logics resulted easily. However, lots of perspectives in the field of dialogical epistemic logic go far beyond the scope of this paper, as for instance the exact delimitation of such systems, extensions to multi-modal logics or to non-normal systems.

A new insight on the relationship between intuitionistic and modal logics has been provided by dialogical logic. It is at least a confirmation of the fecundity of dialogical logic as a frame to compare logical theories.

Appendix: Dialogical games and intuitionistic ML

DialKi and (FS). **DialKi** apparently encompasses the well-known system of intuitionistic modal logic due to Fischer Servi (**FS**) (Fischer Servi, 1977): the axioms of (**FS**) are all valid according to the dialogical set of rules **DialKi**. This is shown through dialogical games; a real proof would require a demonstration exhibiting every possible strategy for the opponent–this can be accomplished with Beth-Smullyan-like tableaux, where formulas are prefixed with the name of the player, **P** or **O**. (See Fitting, 1969 and Rahman and Rückert, 1999 for details.)

The set of axioms of **FS** is the union of $\mathbf{IntK}_\Box$ (the result of extending propositional **Int** with the standard modal axioms K for $\Box$), $\mathbf{IntK}_\Diamond$ (the same except that the standard modal axioms K are given for $\Diamond$), and two specific axioms (see Celani, 2001; Wolter and Zakharyaschev, 1999).

We will need to add the symbol $\bot$ as a prime formula, and define $\neg$ and $\top$ in terms of it. Particle rules for $\bot$ and $\top$ immediately follow from these definitions:

- $\neg A := A \to \bot$, so $\bot$ can never be stated (if it could, negation would be defensible).
- $\top := \neg\bot$, so $\top$ can be stated by any player in any context (it is not an atomic formula) and it cannot be attacked.

In what follows, the games corresponding to each axiom are written down without comment. Every play is won by player **P**, according to some winning strategy. So for every axiom A of **FS**, we get: **DialKi** $\models A$.

IntK$_\Box$: **Int**, $\Box\top$, $\Box(a \wedge b) \leftrightarrow (\Box a \wedge \Box b)$.

	O			**P**		
				$\Box\top$	(0)	1
1	(1)	$?_{\Box/1.1}$	0	$\top$	(2)	1.1

	O			**P**		
				$\Box(a \wedge b) \to (\Box a \wedge \Box b)$	(0)	1
1	(1)	$\Box(a \wedge b)$	0	$\Box a \wedge \Box b$	(2)	1
1	(3)	$?_L$	2	$\Box a$	(4)	1
1	(5)	$?_{\Box/1.1}$	4	a	(10)	1.1
1.1	(7)	$a \wedge b$		1 $?_{\Box/1.1}$	(6)	1
1.1	(9)	a		7 $?_L$	(8)	1.1

	O			**P**		
				$(\Box a \wedge \Box b) \to \Box(a \wedge b)$	(0)	1
1	(1)	$\Box a \wedge \Box b$	0	$\Box(a \wedge b)$	(2)	1
1	(3)	$?_{\Box/1.1}$	2	$a \wedge b$	(4)	1.1
1.1	(5)	$?_L$	4	a	(10)	1.1
1	(7)	$\Box a$		1 $?_L$	(6)	1
1.1	(9)	a		7 $?_{\Box/1.1}$	(8)	1

IntK$_\Diamond$: **Int**, $\neg\Diamond\bot$, $\Diamond(a \vee b) \leftrightarrow (\Diamond a \vee \Diamond b)$

	O			**P**		
				$\neg\Diamond\bot$	(0)	1
1	(1)	$\Diamond\bot$	0	$\otimes$		
		$\otimes$		1 $?_\Diamond$	(2)	1

	O			**P**		
				$\Diamond(a \vee b) \to (\Diamond a \vee \Diamond b)$	(0)	1
1	(1)	$\Diamond(a \vee b)$	0	$\Diamond a \vee \Diamond b$	(2)	1
1	(3)	?	2	$\Diamond a$	(8)	1
1.1	(5)	$a \vee b$		1 $?_\Diamond$	(4)	1
1.1	(7)	a		5 ?	(6)	1.1
1	(9)	$?_\Diamond$	8	a	(10)	1.1

	O			**P**			
				$(\Diamond a \vee \Diamond b) \to \Diamond(a \vee b)$	(0)	1
1	(1)	$\Diamond a \vee \Diamond b$	0	$\Diamond(a \vee b)$	(2)	1
1	(3)	$?_\Diamond$	2	$a \vee b$	(8)	1.1
1	(5)	$\Diamond a$		1 ?	(4)	1
1.1	(7)	a		5 $?_\Diamond$	(6)	1
1.1	(9)	?	8	a	(10)	1.1

FS specific axioms: $\Diamond(a \to b) \to (\Box a \to \Diamond b)$, $(\Diamond a \to \Box b) \to \Box(a \to b)$.

	O			**P**			
				$\Diamond(a \to b) \to (\Box a \to \Diamond b)$	(0)	1
1	(1)	$\Diamond(a \to b)$	0	$\Box a \to \Diamond b$	(2)	1
1	(3)	$\Box a$	2	$\Diamond b$	(4)	1
1	(5)	$?_\Diamond$	4	b	(12)	1.1
1.1	(7)	$a \to b$		1 $?_\Diamond$	(6)	1.1
1.1	(9)	a		3 $?_{\Box/1.1}$	(8)	1
1.1	(11)	b		7 a	(10)	1.1

	O			**P**			
				$(\Diamond a \to \Box b) \to \Box(a \to b)$	(0)	1
1	(1)	$\Diamond a \to \Box b$	0	$\Box(a \to b)$	(2)	1
1	(3)	$?_{\Box/1.1}$	2	$a \to b$	(4)	1.1
1.1	(5)	a	4	b	(10)	1.1
1	(7)	$\Box b$		1 $\Diamond a$	(6)	1
1.1	(9)	b		7 $?_{\Box/1.1}$	(8)	1
1	(11)	$?_\Diamond$	6	a	(12)	1.1

DialKi $\neq$ DialK. This can be shown with the following game, played according to both sets of rules. With the intuitionistic version the proponent cannot answer to (9) and loses, whereas with the standard ones, she can go further and revise her defence against (3). Hence **DialK** $\models \Box(a \vee b) \to (\Box a \vee \Diamond b)$ but **DialKi** $\not\models \Box(a \vee b) \to (\Box a \vee \Diamond b)$.

	O			**P**			
				$\Box(a \vee b) \to (\Box a \vee \Diamond b)$	(0)	1
1	(1)	$\Box(a \vee b)$	0	$\Box a \vee \Diamond b$	(2)	1
1	(3)	?	2	$\Box a$	(4)	1
1	(5)	$?_{\Box/1.1}$	4			
1.1	(7)	$a \vee b$		1 $?_{\Box/1.1}$	(6)	1
1.1	(9)	b		7 ?	(8)	1.1
1	(3′)	?	2	$\Diamond b$	(**10**)	1
1	(**11**)	$?_\Diamond$	10	b	(**12**)	1.1

(Here (3′) is not a move but a repetition of (3) to let the reader see the attack the proponent answers to at round (10).)

References

Celani, S. (2001). Remarks on intuitionistic modal logics. *Divulgaciones Mathematicas*, 9(2):137–147.

Duc, H. N. (1997). Reasoning about rational, but not logically omniscient agents. *Journal of Logic and Computation*, 7(5):633–648.

Fischer Servi, G. (1977). On modal logics with an intuitionistic base. *Studia Logica*, 36: 141–149.

Fitting, M. C. (1969). *Intuitionistic Logic Model Theory and Forcing*. North Holland, Amsterdam, London.

Hintikka, J. (1962). *Knowledge and Belief*. Reidel, Dordrecht.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. MIT Press, Cambridge, MA.

Kripke, S. (1965). *Semantical Analysis of Intuitionistic Logic I*, pages 92–130. North-Holland, Amsterdam.

Lemmon, J. (1957). New foundations for Lewis modal systems. *Journal of Symbolic Logic* 22:176–186.

Lorenzen, P. and Lorenz, K. (1978). *Dialogische Logik*. WBG, Darmstadt.

Rahman, S. (1993). *Über Dialoge, Protologische Kategorien und andere Seltenheiten*. Peter Lang, Frankfurt a. M.

Rahman, S. (2003). Non-normal dialogics for a wonderful world and more. Preprint (2006). In J. van Benthem et al., editors, *The Age of Alternative Logics*, pages 311–334. Springer, Dordrecht.

Rahman, S. and Keiff, L. (2005). On how to be a dialogician. In Vandervecken, D., editor, *Logic, Thought and Action*, pages 359–408. Springer, Dordrecht.

Rahman, S. and Rückert, H. (1999). Dialogische Modallogik für T, B, S4 und S5. *Logique et Analyse*, 167–168:243–282.

Thomason, R. H. (2000). Modeling the beliefs of other agents. In Minker, J., editor, *Logic-Based Artificial Intelligence*. Kluwer, Dordrecht.

van Benthem, J. (1991). Reflections on epistemic logic. *Logique & Analyse*, 133–134:5–14.

van Benthem, J. (2001). *Logic in Games*. Lecture Notes. ILLC, Amsterdam.

Wolter, F. and Zakharyaschev, M. (1999). Intuitionistic modal logic. In Cantini, A., Cesari, E., and Minari, P., editors, *Logic and Foundations of Mathematics*, pages 227–238. Kluwer, Dordrecht.

Part IV

COMPUTATION AND MATHEMATICS

Chapter 11

IN THE BEGINNING WAS GAME SEMANTICS*

Giorgi Japaridze
Department of Computing Sciences, Villanova University
giorgi.japaridze@villanova.edu

Abstract This chapter presents an overview of *computability logic*—the game-semantically constructed logic of interactive computational tasks and resources. There is only one non-overview, technical section in it, devoted to a proof of the soundness of affine logic with respect to the semantics of computability logic.

11.1 Introduction

In the beginning was Semantics, and Semantics was Game Semantics, and Game Semantics was Logic.[1] Through it all concepts were conceived; for it all axioms are written, and to it all deductive systems should serve...

This is not an evangelical story, but the story and philosophy of *computability logic* (CL), the recently introduced (Japaridze, 2003) mini-religion within logic. According to its philosophy, *syntax*—the study of axiomatizations or any other, deductive or nondeductive string-manipulation systems—exclusively owes its right on existence to *semantics*, and is thus secondary to it. CL believes that logic is meant to be the most basic, general-purpose formal tool potentially usable by intelligent agents in successfully navigating real life. And it is semantics that establishes that ultimate real-life meaning of logic. Syntax is important, yet it is so not in its own right but only as much as it serves a meaningful semantics, allowing us to realize the potential of that semantics in some systematic and perhaps convenient or efficient way. Not passing the test for soundness with respect to the underlying semantics would fully disqualify any syntax, no matter how otherwise appealing it is. Note—disqualify the syntax

*This material is based upon work supported by the National Science Foundation under Grant No. 0208816, and 2005 Summer Research Grant from Villanova University.

[1] "In the beginning was the Word, and the Word was with God, and the Word was God... Through him all things were made; without him nothing was made that has been made." — John's Gospel.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

and not the semantics. Why this is so hardly requires any explanation: relying on an unsound syntax might result in wrong beliefs, misdiagnosed patients or crashed spaceships.

Unlike soundness, completeness is a desirable but not necessary condition. Sometimes—as, say, in the case of pure second-order logic, or first-order applied number theory with + and ×—completeness is impossible to achieve in principle. In such cases we may still benefit from continuing working with various reasonably strong syntactic constructions. A good example of such a "reasonable" yet incomplete syntax is Peano arithmetic. Another example, as we are going to see later, is affine logic, which turns out to be sound but incomplete with respect to the semantics of CL. And even when complete axiomatizations are known, it is not fully unusual for them to be sometimes artificially downsized and made incomplete for efficiency, simplicity, convenience or even esthetic considerations. Ample examples of this can be found in applied computer science. But again, while there might be meaningful trade-offs between (the degrees of) completeness, efficiency and other desirable-but-not-necessary properties of a syntax, the underlying semantics remains untouchable, and the condition of soundness unnegotiable. It is that very untouchable core that should be the point of departure for logic as a fundamental science.

A separate question, of course, is what counts as a semantics. The model example of a semantics with a capital 'S' is that of classical logic. But in the logical literature this term often has a more generous meaning than what CL is ready to settle for. As pointed out, CL views logic as a universal-utility tool. So, a capital 'S' semantics should be non-specific enough, and applicable to the world in general rather than some very special and artificially selected fragment of it. Often what is called a semantics is just a special-purpose apparatus designed to help analyze a given syntactic construction rather than understand and navigate the outside world. The usage of Kripke models as a derivability test for intuitionistic formulas, or as a validity criterion in various systems of modal logic is an example. An attempt to see more than a technical, syntax-serving instrument (which, as such, may be indeed very important and useful) in this type of lowercase 's' semantics might create a vicious circle: a deductive system L under question is "right" because it derives exactly the formulas that are valid in a such and such Kripke semantics; and then it turns out that the reason why we are considering the such and such Kripke semantics is that . . . it validates exactly what L derives.

This was about why *in the beginning was Semantics*. Now a few words about why *Semantics was Game Semantics*. For CL, game is not just a game. It is a foundational mathematical concept on which a powerful enough logic (=semantics) should be based. This is so because, as noted, CL sees logic as a "real-life navigational tool", and it is games that appear to offer the most comprehensive, coherent, natural, adequate and convenient mathematical models

for the very essence of all "navigational" activities of agents: their interactions with the surrounding world. An *agent* and its *environment* translate into game-theoretic terms as two *players*; their *actions* as *moves*; *situations* arising in the course of interaction as *positions*; and *success* or *failure* as *wins* or *losses*.

It is natural to require that the interaction strategies of the party that we have referred to as an "agent" be limited to *algorithmic* ones, allowing us to henceforth call that player a *machine*. This is a minimum condition that any non-esoteric game semantics would have to satisfy. On the other hand, no restrictions can or should be imposed on the environment, who represents 'the blind forces of nature, or the devil himself' (Japaridze, 2003). Algorithmic activities being synonymous to *computations*, games thus represent *computational problems*—interactive tasks performed by computing agents, with *computability* meaning *winnability*, i.e. existence of a machine that wins the game against any possible (behavior of the) environment.

In the 1930s mankind came up with what has been perceived as an ultimate mathematical definition of the precise meaning of algorithmic solvability. Curiously or not, such a definition was set forth and embraced before really having attempted to answer the seemingly more basic question about what *computational problems* are—the very entities that may or may not have algorithmic solutions in the first place. The tradition established since then in theoretical computer science by computability simply means Turing computability of *functions*, as the task performed by every Turing machine is nothing but receiving an input x and generating the output $f(x)$ for some function f. Turing himself (Turing, 1936), however, was more cautious about making overly broad philosophical conclusions, acknowledging that not everything one would potentially call a computational problem might necessarily be a function, or reducible to such. Most tasks that computers and computer networks perform are interactive. And nowadays more and more voices are being heard (Goldin et al., 2004; Japaridze, 2006e; Milner, 1993; Wegner, 1998) pointing out that true interaction might be going beyond what functions and hence ordinary Turing machines are meant to capture.

Two main concepts on which the semantics of CL is based are those of *static games* and their *winnability* (defined later in Sections 11.5 and 11.6). Correspondingly, the philosophy of CL relies on two beliefs that, together, present what can be considered an interactive version of the Church-Turing thesis:

Belief 1. *The concept of static games is an adequate formal counterpart of our intuition of* ("*pure*", *speed-independent*) *interactive computational problems.*

Belief 2. *The concept of winnability is an adequate formal counterpart of our intuition of algorithmic solvability of such problems.*

As will be seen later, one of the main features distinguishing the CL games from more traditional game models is the absence of *procedural rules* (van Benthem, 2001)—rules strictly regulating which player is to move in any given situation. Here, in a general case, either player is free to move. It is exactly this feature that makes players' strategies no longer definable as functions (functions from positions to moves). And it is this highly relaxed nature that makes the CL games apparently most flexible and general of all two-player, two-outcome games.

Trying to understand strategies as functions would not be a good idea even if the type of games we consider naturally allowed us to do so. Because, when it comes to long or infinite games, functional strategies would be disastrously inefficient, making it hardly possible to develop any reasonable complexity theory for interactive computation (the next important frontier for CL or theoretical computer science in general). To understand this, it would be sufficient to just reflect on the behavior of one's personal computer. The job of your computer is to play one long—potentially infinite—game against you. Now, have you noticed your faithful servant getting slower every time you use it? Probably not. That is because the computer is smart enough to follow a non-functional strategy in this game. If its strategy was a function from positions (interaction histories) to moves, the response time would inevitably keep worsening due to the need to read the entire—continuously lengthening and, in fact, practically infinite—interaction history every time before responding. Defining strategies as functions of only the latest moves (rather than entire interaction histories) in Abramsky and Jagadeesan's (1994) tradition is also not a way out, as typically more than just the last move matters. Back to your personal computer, its actions certainly depend on more than your last keystroke.

Computability in the traditional Church-Turing sense is a special case of winnability—winnability restricted to two-step (input/output, question/answer) interactive problems. So is the classical concept of truth, which is nothing but winnability restricted to propositions, viewed by CL as zero-step problems, i.e. games with no moves that are automatically won or lost depending on whether they are true or false. This way, the semantics of CL is a generalization, refinement and conservative extension of that of classical logic.

Thinking of a human user in the role of the environment, computational problems are synonymous to computational tasks—tasks performed by a machine for the user/environment. What is a task for a machine is then a resource for the environment, and vice versa. So the CL games, at the same time, formalize our intuition of *computational resources*. Logical operators are understood as operations on such tasks/ resources/games, atoms as variables ranging over tasks/resources/games, and validity of a logical formula as being "always winnable", i.e. as existence—under every particular interpretation of atoms—of a machine that successfully accomplishes/provides/wins the corresponding

task/resource/game no matter how the environment behaves. With this semantics, 'computability logic is a formal theory of computability in the same sense as classical logic is a formal theory of truth' (Japaridze, 2006c). Furthermore, as mentioned, the classical concept of truth is a special case of winnability, which eventually translates into classical logic's being nothing but a special fragment of computability logic.

CL is a semantically constructed logic and, at this young age, its syntax is only just starting to develop, with open problems and unverified conjecture prevailing over answered questions. In a sense, this situation is opposite to the case with some other non-classical traditions such as intuitionistic or linear logics where, as most logicians would probably agree, "in the beginning was Syntax", and really good formal semantics convincingly justifying the proposed syntactic constructions are still being looked for. In fact, the semantics of CL can be seen to be providing such a justification, although, for linear logic, this is only in a limited sense explained below.

The set of valid formulas in a certain fragment of the otherwise more expressive language of CL forms a logic that is similar to but by no means the same as linear logic. The two logics typically agree on short and simple formulas (perhaps with the exception for those involving exponentials, where disagreements may start already on some rather short formulas). Say, both logics reject $P \rightarrow P \wedge P$ and accept $P \rightarrow P \sqcap P$, with classical-shape propositional connectives here and later understood as the corresponding multiplicative operators of linear logic, and square-shape operators as additives ($\sqcap$="with", $\sqcup$="plus"). Similarly, both logics reject $P \sqcup \neg P$ and accept $P \vee \neg P$. On the other hand, CL disagrees with linear logic on many more evolved formulas. For example, CL validates the following two principles rejected even by *affine logic* **AL**—linear logic with the weakening rule:

$$((P \wedge Q) \vee (R \wedge S)) \rightarrow ((P \vee R) \wedge (Q \vee S));$$

$$(P \wedge (R \sqcap S)) \sqcap (Q \wedge (R \sqcap S)) \sqcap ((P \sqcap Q) \wedge R) \sqcap ((P \sqcap Q) \wedge S) \rightarrow (P \sqcap Q) \wedge (R \sqcap S).$$

Neither the similarities nor the discrepancies are a surprise. The philosophies of CL and linear logic overlap in their striving to develop a logic of resources. But the ways this philosophy is materialized are rather different. CL starts with a mathematically strict and intuitively convincing semantics, and only after that, as a natural second step, asks what the corresponding logic and its possible axiomatizations are. On the other hand, it would be accurate to say that linear logic started directly from the second step. Even though certain companion semantics were provided for it from the very beginning, those are not quite what we earlier agreed to call capital-'S'. As a formal theory of resources (rather than that of phases or coherent spaces), linear logic has been motivated and introduced syntactically rather than semantically, essentially by

taking classical sequent calculus and deleting the rules that seemed unacceptable from a certain intuitive, naive resource point of view. Hence, in the absence of a clear formal concept of resource-semantical truth or validity, the question about whether the resulting system was complete could not even be meaningfully asked. In this process of syntactically rewriting classical logic some innocent, deeply hidden principles could have easily gotten victimized. CL believes that this is exactly what happened, with the above formulas separating it from linear logic—and more such formulas to be seen later—viewed as babies thrown out with the bath water. Of course, many retroactive attempts have been made to find semantical (often game-semantical) justifications for linear logic. Technically it is always possible to come up with some sort of a formal semantics that matches a given target syntactic construction, but the whole question is how natural and meaningful such a semantics is in its own rights, and how adequately it corresponds to the logic's underlying philosophy and ambitions. 'Unless, by good luck, the target system really *is* "the right logic", the chances of a decisive success when following the odd scheme *from syntax to semantics* could be rather slim' (Japaridze, 2003). The natural scheme is *from semantics to syntax*. It matches the way classical logic evolved and climaxed in Gödel's completeness theorem. And, as we now know, this is exactly the scheme that computability logic, too, follows.

Intuitionistic logic is another example of a syntactically conceived logic. Despite decades of efforts, no fully convincing semantics has been found for it. Lorenzen's game semantics (Felscher, 1985; Lorenzen, 1959), which has a concept of validity without having a concept of truth, has been perceived as a technical supplement to the existing syntax rather than as having independent importance. Some other semantics, such as Kleene's realizability (Kleene, 1952) or Gödel's Dialectica interpretation (Gödel, 1958), are closer to what we might qualify as capital-'S'. But, unfortunately, they validate certain principles unnegotiably rejected by intuitionistic logic. From our perspective, the situation here is much better than with linear logic though. In Japaridze (2006b), Heyting's first-order intuitionistic calculus has been shown to be sound with respect to the CL semantics. And the propositional fragment of Heyting's calculus has also been shown to be complete (Japaridze, 2007c, d, 2008b; Vereshchagin, 2006). This signifies success—at least at the propositional level—in semantically justifying intuitionistic logic, and a materialization of Kolmogorov's (1932) well known yet so far rather abstract thesis according to which intuitionistic logic is a logic of problems. Just as the resource philosophy of CL overlaps with that of linear logic, so does its algorithmic philosophy with the constructivistic philosophy of intuitionism. The difference, again, is in the ways this philosophy is materialized. Intuitionistic logic has come up with a "constructive syntax" without having an adequate underlying formal semantics, such as a clear concept of truth in some constructive

sense. This sort of a syntax was essentially obtained from the classical one by banning the offending law of the excluded middle. But, as in the case of linear logic, the critical question immediately springs out: where is a guarantee that together with excluded middle some innocent principles would not be expelled as well? The constructivistic claims of CL, on the other hand, are based on the fact that it defines truth as algorithmic solvability. The philosophy of CL does not find the term *constructive syntax* meaningful unless it is understood as soundness with respect to some *constructive semantics*, for only a semantics may or may not be constructive in a reasonable sense. The reason for the failure of $P \sqcup \neg P$ in CL is not that this principle ... is not included in its axioms. Rather, the failure of this principle is exactly the reason why this principle, or anything else entailing it, would not be among the axioms of a sound system for CL. Here "failure" has a precise semantical meaning. It is non-validity, i.e. existence of a problem A for which $A \sqcup \neg A$ is not algorithmically solvable.

It is also worth noting that, while intuitionistic logic irreconcilably defies classical logic, computability logic comes up with a peaceful solution acceptable for everyone. The key is the expressiveness of its language, that has (at least) two versions for each traditionally controversial logical operator, and particularly the two versions $\vee$ and $\sqcup$ of disjunction. As will be seen later, the semantical meaning of $\vee$ conservatively extends—from moveless games to all games—its classical meaning, and the principle $P \vee \neg P$ survives as it represents an always-algorithmically-solvable combination of problems, even if solvable in a sense that some constructivistically-minded might fail—or pretend to fail—to understand. And the semantics of $\sqcup$, on the other hand, formalizes and conservatively extends a different, stronger meaning which apparently every constructivist associates with disjunction. As expected, then $P \sqcup \neg P$ turns out to be semantically invalid. CL's proposal for settlement between classical and constructivistic logics then reads: 'If you are open (=classically) minded, take advantage of the full expressive power of CL; and if you are constructivistically minded, just identify a collection of the operators whose meanings seem constructive enough to you, mechanically disregard everything containing the other operators, and put an end to those fruitless fights about what deductive methods or principles should be considered right and what should be deemed wrong' (Japaridze, 2003).

Back to linear—more precisely, affine—logic. As mentioned, **AL** is sound with respect to the CL semantics, a proof of which is the main new technical contribution of the present paper. This is definitely good news from the "better something than nothing" standpoint. **AL** is simple and, even though incomplete, still reasonably strong. What is worth noting is that our soundness proof for **AL**, just as all other soundness proofs known so far in CL, including that for the intuitionistic fragment (Japaridze, 2007d), or the **CL4** fragment (Japaridze, 2007a) that will be discussed in Section 11.9, is *constructive*. This is in the sense that, whenever a formula F is provable in a given deductive

system, an algorithmic solution for the problem(s) represented by F can be automatically extracted from the proof of F. The persistence of this phenomenon for various fragments of CL carries another piece of good news: CL provides a systematic answer not only to the theoretical question "*what* can be computed?" but, as it happens, also to the more terrestrial question "*how* can be computed?".

The main practical import of the constructive soundness result for **AL** (just as for any other sublogic of CL) is related to the potential of basing applied theories or knowledge base systems on that logic, the latter being a reasonable, computationally meaningful alternative to classical logic. The non-logical axioms—or knowledge base—of an **AL**-based applied system/theory would be any collection of (formulas representing) problems whose algorithmic solutions are known. Then our soundness result for **AL** guarantees that every theorem T of the theory also has an algorithmic solution and that, furthermore, such a solution, itself, can be effectively constructed from a proof of T. This makes **AL** a systematic problem-solving tool: finding a solution for a given problem reduces to finding a proof of that problem in an **AL**-based theory. The incompleteness of **AL** only means that, in its language, this logic is not as perfect/strong as a formal tool could possibly be, and that, depending on needs, it makes sense to continue looking for further sound extensions (up to a complete one) of it. As pointed out earlier, when it comes to applications, unlike soundness, completeness is a desirable but not necessary condition.

With the two logics in a sense competing for the same market, the main—or perhaps only—advantage of linear logic over CL is its having a nice and simple syntax. In fact, linear logic *is* (rather than *has*) a beautiful syntax; and computability logic *is* (rather than *has*) a meaningful semantics. At this point it is not clear what a CL-semantically complete extension of **AL** would look like syntactically. As a matter of fact, while the set of valid formulas of the exponential-free fragment of the language of linear logic has been shown to be decidable (Japaridze, 2007a), so far it is not even known whether that set in the full language is recursively enumerable. If it is, finding a complete axiomatization for it would most likely require a substantially new syntactic approach, going far beyond the traditional sequent-calculus framework within which linear logic is constructed (a possible candidate here is cirquent calculus, briefly discussed at the end of this section). And, in any case, such an axiomatization would hardly be as simple as that of **AL**, so the syntactic simplicity advantage of linear logic will always remain. Well, CL has one thing to say: simplicity is good, yet, if it is most important, then nothing can ever beat . . . the empty logic.

The rest of this paper is organized as follows. Sections 11.2–11.8 provide a detailed introduction to the basic semantical concepts of computability logic: games and operations on them, two equivalent models of interactive computation (algorithmic strategies), and validity. The coverage of most of these concepts is more detailed here than in the earlier survey-style papers (Japaridze,

2003, 2006e) on CL, and is supported with ample examples and illustrations. Section 11.9 provides an overview, without a proof, of the strongest technical result obtained so far in computability logic, specifically, the soundness and completeness of system **CL4**, whose logical vocabulary contains negation $\neg$, parallel ("multiplicative") connectives $\wedge, \vee, \rightarrow$, choice ("additive") connectives $\sqcap, \sqcup$ with their quantifier counterparts $\bigsqcap, \bigsqcup$, and blind ("classical") quantifiers $\forall, \exists$. Section 11.10 outlines potential applications of computability logic in knowledge base systems, systems for planning and action, and constructive applied theories. There the survey part of the paper ends, and the following two sections are devoted to a formulation (Section 11.11) and proof (Section 11.12) of the new result—the soundness of affine logic with respect to the semantics of CL. The final Section 11.13 outlines some possible future developments in the area.

This paper gives an overview of most results known in computability logic as of the end of 2005, by the time when the main body of the text was written. The present paragraph is a last-minute addition made at the beginning of 2008. Below is a list of the most important developments that, due to being very recent, have received no coverage in this chapter:

- As already mentioned, the earlier conjecture about the completeness of Heyting's propositional intuitionistic calculus with respect to the semantics of CL has been resolved positively. A completeness proof for the implicative fragment of intuitionistic logic was given in Japaridze (2007c), and that proof was later extended to the full propositional intuitionistic calculus in Japaridze (2007d). With a couple of months' delay, Vereshchagin (2006) came up with an alternative proof of the same result.

- In Japaridze (2009), the implicative fragment of affine logic has been proven to be complete with respect to the semantics of computability logic. The former is nothing but implicative intuitionistic logic without the rule of contraction. Thus, both the implication of intuitionistic logic and the implication of affine logic have adequate interpretations in CL—specifically, as the operations $\circ\!\!-$ and $\rightarrow$, respectively. Intuitively, as will be shown later in Section 11.4, these are two natural versions of the operation of reduction, with the difference between $A \circ\!\!- B$ and $A \rightarrow B$ being that in the former A can be "reused" while in the latter it cannot. Japaridze (2009) also introduced a series of intermediate-strength natural versions of reduction operations.

- Section 11.4.6 briefly mentions *sequential operations*. The recent paper (Japaridze, 2008b) has provided an elaboration of this new group of operations (formal definitions, associated computational intuitions, motivations, etc.), making them full-fledged citizens of computability

logic. It has also constructed a sound and complete axiomatization of the fragment of computability logic whose logical vocabulary, together with negation, contains three—parallel, choice and sequential—sorts of conjunction and disjunction.

- Probably the most significant of the relevant recent developments is the invention of *cirquent calculus* in Japaridze (2007b, c). Roughly, this is a deductive approach based on circuits instead of formulas or sequents. It can be seen as a refinement of Gentzen's methodology, and correspondingly the methodology of linear logic based on the latter, achieved by allowing shared resources between different parts of sequents and proof trees. Thanks to the sharing mechanism, cirquent calculus, being more general and flexible than sequent calculus, appears to be the only reasonable proof-theoretic approach capable of syntactically taming the otherwise wild computability logic. Sharing also makes it possible to achieve exponential-magnitude compressions of formulas and proofs, whether it be in computability logic or the kind old classical logic.

11.2 Constant games

The symbolic names used in CL for the two players *machine* and *environment* are $\top$ and $\bot$, respectively. $\wp$ is always a variable ranging over $\{\top, \bot\}$, with

$$\neg\wp$$

meaning $\wp$'s adversary, i.e. the player which is not $\wp$. Even though it is often a human user that acts in the role of $\bot$, our sympathies are with $\top$ rather than $\bot$, and by just saying "won" or "lost" without specifying by whom we always mean won or lost by $\top$.

The reason why I should be a fan of the machine even—in fact especially—when it is playing against me is that the machine is a tool, and what makes it valuable as such is exactly its winning the game, i.e. its not being malfunctional (it is precisely losing by a machine the game that it was supposed to win what in everyday language is called *malfunctioning*). Let me imagine myself using a computer for computing the "28% of x" function in the process of preparing my federal tax return. This is a game where the first move is mine, consisting in inputting a number m and meaning asking $\top$ the question "what is 28% of m?". The machine wins iff it answers by the move/output n such that $n = 0.28m$. Of course, I do not want the machine to tell me that $27,000$ is 28% of 100,000. In other words, I do not want to win against the machine. For then I could lose the more important game against Uncle Sam.

Before getting to a formal definition of games, let us agree without loss of generality that a **move** is always a string over the standard keyboard alphabet. One of the non-numeric and non-punctuation symbols of this alphabet, denoted

♠, is designated as a special-status move, intuitively meaning a move that is always illegal to make. A **labeled move** (**labmove**) is a move prefixed with $\top$ or $\bot$, with its prefix (**label**) indicating which player has made the move. A **run** is a (finite or infinite) sequence of labmoves, and a **position** is a finite run.

We will be exclusively using the letters $\Gamma, \Delta, \Theta, \Phi, \Psi, \Upsilon, \Lambda, \Sigma, \Pi$ for runs, $\alpha, \beta, \gamma, \delta$ for moves, and λ for labmoves. Runs will be often delimited by "⟨" and "⟩", with ⟨⟩ thus denoting the **empty run**. The meaning of an expression such as $\langle \Phi, \wp\alpha, \Gamma\rangle$ must be clear: this is the result of appending to position Φ the labmove $\wp\alpha$ and then the run Γ. We write

$$\neg\Gamma$$

for the result of simultaneously replacing every label $\wp$ in every labmove of Γ by $\neg\wp$.

Our ultimate definition of games will be given later in terms of the simpler and more basic class of games called *constant*. The following is a formal definition of constant games combined with some less formal conventions regarding the usage of certain terminology.

Definition 1. *A* **constant game** *is a pair* $A = (\mathbf{Lr}^A, \mathbf{Wn}^A)$*, where:*

1. $\mathbf{Lr}^A$*, called the* **structure** *of A, is a set of runs not containing (whatever-labeled) move* ♠*, satisfying the condition that a finite or infinite run is in* $\mathbf{Lr}^A$ *iff all of its nonempty finite—not necessarily proper—initial segments are in* $\mathbf{Lr}^A$ *(notice that this implies* $\langle\rangle \in \mathbf{Lr}^A$*). The elements of* $\mathbf{Lr}^A$ *are said to be* **legal runs** *of A, and all other runs are said to be* **illegal runs** *of A. We say that* α *is a* **legal move** *for* $\wp$ *in a position* Φ *of A iff* $\langle \Phi, \wp\alpha\rangle \in \mathbf{Lr}^A$*; otherwise* α *is an* **illegal move***. When the last move of the shortest illegal initial segment of* Γ *is* $\wp$*-labeled, we say that* Γ *is a* $\wp$**-illegal run** *of A.*

2. $\mathbf{Wn}^A$*, called the* **content** *of A, is a function that sends every run* Γ *to one of the players* $\top$ *or* $\bot$*, satisfying the condition that if* Γ *is a* $\wp$*-illegal run of A, then* $\mathbf{Wn}^A\langle\Gamma\rangle \neq \wp$*. When* $\mathbf{Wn}^A\langle\Gamma\rangle = \wp$*, we say that* Γ *is a* $\wp$**-won** *(or* **won by** $\wp$*) run of A; otherwise* Γ *is* **lost by** $\wp$*. Thus, an illegal run is always lost by the player who has made the first illegal move in it.*

Let A be a constant game. A is said to be **finite-depth** iff there is a (smallest) integer d, called the **depth** of A, such that the length of every legal run of A is $\leq d$. And A is **perifinite-depth** iff every legal run of it is finite, even if there are arbitrarily long legal runs. Japaridze (2003) defines the depths of perifinite-depth games in terms of ordinal numbers, which are finite for finite-depth games and transfinite for all other perifinite-depth games. Let us call a legal run Γ of A **maximal** iff Γ is not a proper initial segment of any other legal

run of A. Then we say that A is **finite-breadth** if the total number of maximal legal runs of A, called the **breadth** of A, is finite. Note that, in a general case, the breadth of a game may be not only infinite, but even uncountable. A is said to be (simply) **finite** iff it only has a finite number of legal runs. Of course, A is finite only if it is finite-breadth, and when A is finite-breadth, it is finite iff it is finite-depth iff it is perifinite-depth.

The structure component of a constant game can be visualized as a tree whose arcs are labeled with labmoves, as shown in Figure 11.1. Every branch of such a tree represents a legal run, specifically, the sequence of the labels of the arcs of that branch in the top-down direction starting from the root. For instance, the rightmost branch (in its full length) of the tree of Figure 11.1 corresponds to the run $\langle \bot\gamma, \top\gamma, \top\alpha \rangle$. Thus the nodes of a tree, identified with the (sub)branches that end in those nodes, represent legal positions; the root stands for the empty position, and leaves for maximal positions.

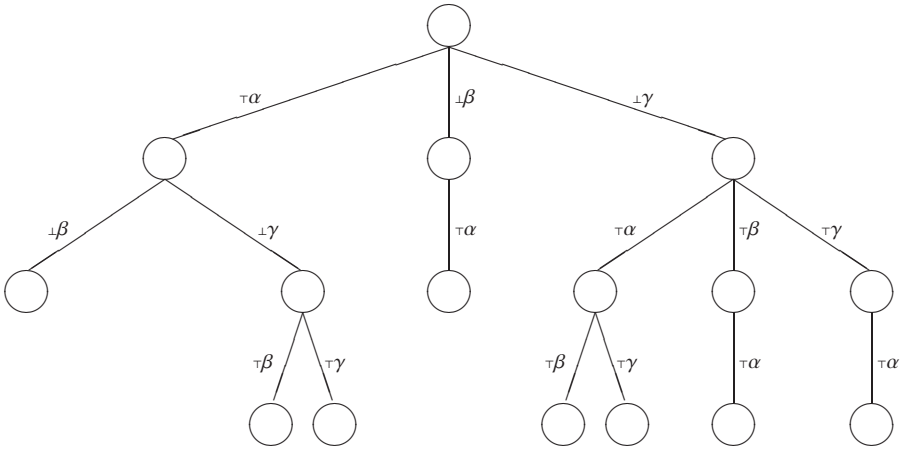


Figure 11.1: A structure

Notice the relaxed nature of our games. In the empty position of the above-depicted structure, both players have legal moves. This can be seen from the two ($\top$-labeled and $\bot$-labeled) sorts of labmoves on the outgoing arcs of the root. Let us call such positions/nodes **heterogenous**. Generally any non-leaf nodes can be heterogenous, even though in our particular example only the root is so. As we are going to see later, in heterogenous positions indeed either player is free to move. Based on this liberal attitude, our games can be called **free**, as opposed to *strict* games where, in every situation, at most one of the players is allowed to move. Of course, strict games can be considered special cases of our free games—the cases with no heterogenous nodes. Even though not having legal moves does not formally preclude the "wrong" player to move in a given position, such an action, as we remember, results in an immediate

loss for that player and hence amounts to not being permitted to move. There are good reasons in favor of the free-game approach. Hardly many tasks that humans, computers or robots perform in real life are strict. Imagine you are playing chess over the Internet on two boards against two independent adversaries that, together, form the (one) environment for you. Let us say you play white on both boards. Certainly the initial position of this game is not heterogenous. However, once you make your first move—say, on board #1—the picture changes. Now both you and the environment have legal moves, and who will be the next to move depends on who can or wants to act sooner. Namely, you are free to make another opening move on board #2, while the environment—adversary #1—can make a reply move on board #1. A strict-game approach would have to impose some not-very-adequate supplemental conditions uniquely determining the next player to move, such as not allowing you to move again until receiving a response to your previous move. Let alone that this is not how the real two-board game would proceed, such regulations defeat the very purpose of the idea of parallel/distributed computations with all the known benefits it offers.

While the above discussion used the term "strict game" in a perhaps somewhat more general sense, let us agree that from now on we will stick to the following meaning of that term:

Definition 2. *A constant game A is said to be* **strict** *iff, for every legal position* Φ *of A, we have* $\{\alpha \mid \langle \Phi, \top\alpha\rangle \in \mathbf{Lr}^A\} = \emptyset$ *or* $\{\alpha \mid \langle \Phi, \bot\alpha\rangle \in \mathbf{Lr}^A\} = \emptyset$.

Figure 11.2 adds a content to the structure of Figure 11.1, thus turning it into a constant game:

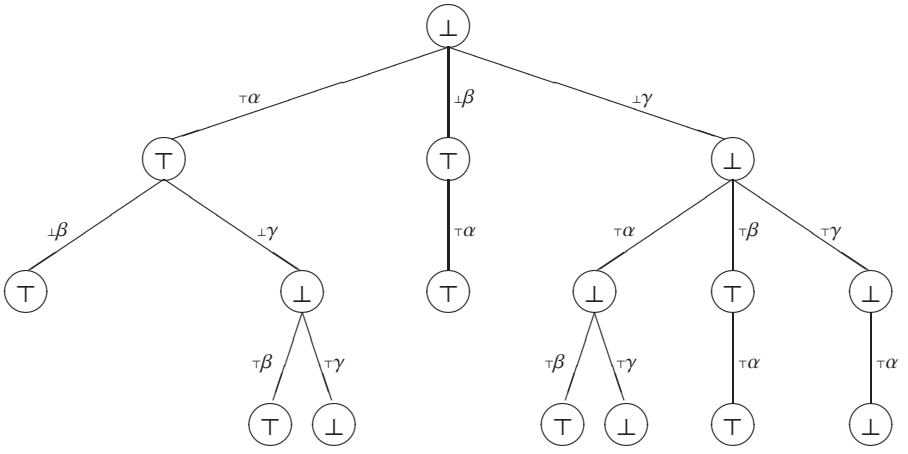


Figure 11.2: Constant game = structure + content

Here the label of each node indicates the winner in the corresponding position. For example, we see that the empty run is won by $\bot$, and the run $\langle \top\alpha, \bot\gamma, \top\beta \rangle$ won by $\top$. There is no need to indicate winners for illegal runs: as we remember, such runs are lost by the player responsible for making them illegal, so we can tell at once that, say, $\langle \top\alpha, \bot\gamma, \top\alpha, \top\beta, \bot\gamma \rangle$ is lost by $\top$ because the offending third move of it is $\top$-labeled. Generally, every perifinite-depth constant game can be fully represented in the style of Figure 11.2 by labeling the nodes of the corresponding structure tree. To capture a non-perifinite-depth game, we will need some additional way to indicate the winners in infinite branches, for no particular (end)nodes represent such branches.

The traditional, strict-game approach usually defines a player $\wp$'s strategy as a function that sends every position in which $\wp$ has legal moves to one of those moves. As pointed out earlier, such a functional view is no longer applicable in the case of properly free games. Indeed, if $f_\top$ and $f_\bot$ are the two players' functional strategies for the game of Figure 11.2 with $f_\top(\langle\rangle) = \alpha$ and $f_\bot(\langle\rangle) = \beta$, then it is not clear whether the first move of the corresponding run will be $\top\alpha$ or $\bot\beta$. Yet, even if not functional, $\top$ does have a winning strategy for that game. What, exactly, a *strategy* means will be explained in Section 11.6. For now, in our currently available ad hoc terms, one of $\top$'s winning strategies sounds as follows: "Regardless of what the adversary is doing or has done, go ahead and make move α; make β as your second move if and when you see that the adversary has made move γ, no matter whether this happened before or after your first move". Which of the runs consistent with this strategy will become the actual one depends on how (and how fast) $\bot$ acts, yet every such run will be a success for $\top$. It is left as an exercise for the reader to see that there are exactly five possible legal runs consistent with $\top$'s above strategy, all won by $\top$: $\langle \top\alpha \rangle$, $\langle \top\alpha, \bot\beta \rangle$, $\langle \top\alpha, \bot\gamma, \top\beta \rangle$, $\langle \bot\beta, \top\alpha \rangle$ and $\langle \bot\gamma, \top\alpha, \top\beta \rangle$. As for illegal runs consistent with that strategy, it can be seen that every such run would be $\bot$-illegal and hence, again, won by $\top$.

Below comes our first formal definition of a game operation. This operation, called **prefixation**, is somewhat reminiscent of the modal operator(s) of dynamic logic. It takes two arguments: a (here constant) game A and a legal position Φ of A, and generates the game $\langle\Phi\rangle A$ that, with A visualized as a tree in the style of Figure 11.2, is nothing but the subtree rooted at the node corresponding to position Φ. This operation is undefined when Φ is an illegal position of A.

Definition 3. *Let A be a constant game and Φ a legal position of A. The game $\langle\Phi\rangle A$ is defined by:*

- $\mathbf{Lr}^{\langle\Phi\rangle A} = \{\Gamma \mid \langle\Phi, \Gamma\rangle \in \mathbf{Lr}^A\}$.
- $\mathbf{Wn}^{\langle\Phi\rangle A}\langle\Gamma\rangle = \mathbf{Wn}^A\langle\Phi, \Gamma\rangle$.

Intuitively, $\langle\Phi\rangle A$ is the game playing which means playing A starting (continuing) from position Φ. That is, $\langle\Phi\rangle A$ is the game to which A **evolves** (will be "**brought down**") after the moves of Φ have been made.

11.3 Games in general, and nothing but games

Computational problems in the traditional, Church-Turing sense can be seen as strict, depth-2 games of the special type shown in Figure 11.3. The first-level arcs of such a game represent inputs, i.e., $\bot$'s moves; and the second-level arcs represent outputs, i.e., $\top$'s moves. The root of this sort of a game is always $\top$-labeled as it corresponds to the situation when there was no input, in which case the machine is considered the winner because the absence of an input removes any further responsibility from it. All second-level nodes, on the other hand, are $\bot$-labeled, for they represent the situations when there was an input but the machine failed to generate any output. Finally, each group of siblings of the third-level nodes has exactly one $\top$-labeled member. This is so because traditional problems are about computing functions, meaning that there is exactly one "right" output per given input. What particular nodes of those groups will have label $\top$—and only this part of the game tree—depends on what particular function is the one under question. The game of Figure 11.3 is about computing the successor function.

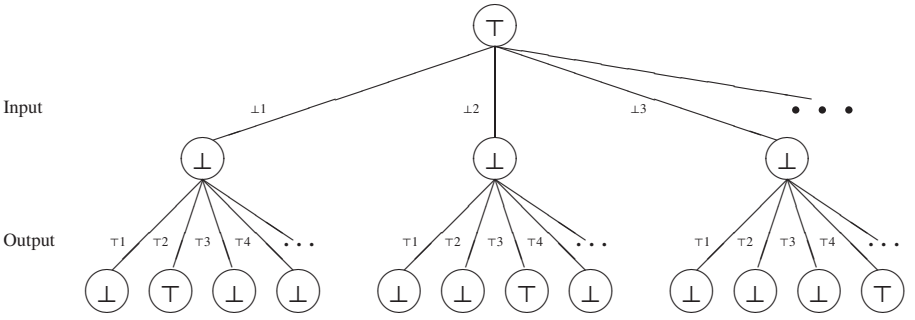


Figure 11.3: The problem of computing $n + 1$

Once we agree that computational problems are nothing but games, the difference in the degrees of generality and flexibility between the traditional approach to computational problems and our approach becomes apparent and appreciable. What we see in Figure 11.3 is indeed a very special sort of game, and there is no good call for confining ourselves to its limits. In fact, staying within those limits would seriously retard any more or less advanced and systematic study of computability. First of all, one would want to get rid of the "one $\top$-labeled node per sibling group" restriction for the third-level nodes.

Many natural problems, such as the problem of finding a prime integer between n and $2n$, or finding an integral root of $x^2 - 2n = 0$, may have more than one as well as less than one solution. That is, there can be more than one as well as less than one "right" output on a given input n. And why not further get rid of any remaining restrictions on the labels of whatever-level nodes and whatever-level arcs. One can easily think of natural situations when, say, some inputs do not obligate the machine to generate an output and thus the corresponding second-level nodes should be $\top$-labeled. An example would be the case when the machine is computing a partially-defined function f and receives an input n on which f is undefined. So far we have been talking about generalizations within the depth-2 restriction, corresponding to viewing computational problems as very short dialogues between the machine and its environment. Permitting longer-than-2 or even infinitely long branches would allow us to capture problems with arbitrarily high degrees of interactivity and arbitrarily complex interaction protocols. The task performed by a network server is a tangible example of an infinite dialogue between the server and its environment—the collection of clients, or let us just say the rest of the network. Notice that such a dialogue is usually a properly free game with a much more sophisticated interface between the interacting parties than the simple input/output interface offered by the ordinary Turing machine model, where the whole story starts by the environment asking a question (input) and ends by the machine generating an answer (output), with no interaction whatsoever in-between these two steps.

Removing restrictions on depths yields a meaningful generalization not only in the upward, but in the downward direction as well: it does make perfect sense to consider "dialogues" of lengths less than 2. Constant games of depth 0 we call **elementary**. There are exactly two elementary constant games, for which we use the same symbols $\top$ and $\bot$ as for the two players (Figure 11.4):

Figure 11.4: Elementary constant games

We identify these with the two propositions of classical logic: $\top$ (*true*) and $\bot$ (*false*). "Snow is white" is thus a moveless game automatically won by the machine, while "Snow is black" is automatically lost. So, not only traditional computational problems are special cases of our games, but traditional propositions as well. This is exactly what eventually makes classical logic a natural—elementary—fragment of computability logic.

As we know, however, propositions are not sufficient to build a reasonably expressive logic. For higher expressiveness, classical logic generalizes propositions to predicates. Let us fix two infinite sets of expressions: the set $\{v_1, v_2, \ldots\}$ of **variables** and the set $\{1, 2, \ldots\}$ of **constants**. Without loss of generality here we assume that this collection of constants is exactly the universe of discourse in all cases that we consider. By a **valuation** we mean a function e that sends each variable x to a constant $e(x)$. In these terms, a classical **predicate** p can be understood as a function that sends each valuation e to either $\top$ (meaning that p is true at e) or $\bot$ (meaning that p is false at e). Propositions can thus be thought of as special, *constant* cases of predicates—predicates that return the same proposition for every valuation.

The concept of games that we define below generalizes constant games in exactly the same sense as the above classical concept of predicates generalizes propositions:

Definition 4. *A* **game** *is a function from valuations to constant games.*

We write $e[A]$ *(rather than* $A(e)$*) to denote the constant game returned by game* A *for valuation* e*. Such a constant game* $e[A]$ *is said to be an* **instance** *of* A*.*

We also typically write $\mathbf{Lr}^A_e$ *and* $\mathbf{Wn}^A_e$ *instead of* $\mathbf{Lr}^{e[A]}$ *and* $\mathbf{Wn}^{e[A]}$*.*

Throughout this paper, x, y, z will be usually used as metavariables for variables, c for constants and e for valuations.

Just as this is the case with propositions versus predicates, we think of constant games in the sense of Definition 1 as special, *constant* cases of games in the sense of Definition 4. In particular, each constant game A' is the game A such that, for every valuation e, $e[A] = A'$. From now on we will no longer distinguish between such A and A', so that, if A is a constant game, it is its own instance, with $A = e[A]$ for every e.

The notion of elementary game that we defined for constant games naturally generalizes to all games by stipulating that a given game is **elementary** iff all of its instances are so. Hence, just as we identified classical propositions with constant elementary games, classical predicates from now on will be identified with elementary games. For instance, $Even(x)$ is the elementary game such that $e[Even(x)]$ is the game $\top$ if $e(x)$ is even, and the game $\bot$ if $e(x)$ is odd. Many other concepts originally defined only for constant games—including the properties *strict*, *finite*, *(peri)finite-depth* and *finite-breadth*—can be extended to all games in a similar way.

We say that a game A **depends on** a variable x iff there are two valuations e_1, e_2 which agree on all variables except x such that $e_1[A] \neq e_2[A]$. Constant games thus do not depend on any variables. A is said to be **finitary** iff there is a finite set $\vec{x}$ of variables such that, for every two valuations e_1 and e_2 that agree on all variables of $\vec{x}$, we have $e_1[A] = e_2[A]$. The cardinality of (the smallest) such $\vec{x}$ is said to be the **arity** of A. So, "constant game" and "0-ary game" are synonyms.

To generalize the standard operation of substitution of variables to games, let us agree that by a **term** we mean either a variable or a constant. The domain of each valuation e is extended to all terms by stipulating that,

$$\textit{for any constant } c,\ e(c) = c.$$

Definition 5. *Let A be a game, $x_1, \ldots, x_n$ pairwise distinct variables, and $t_1, \ldots, t_n$ any (not necessarily distinct) terms. The result of* **substituting** *$x_1, \ldots, x_n$ by $t_1, \ldots, t_n$ in A, denoted $A(x_1/t_1, \ldots, x_n/t_n)$, is defined by stipulating that, for every valuation e, $e[A(x_1/t_1, \ldots, x_n/t_n)] = e'[A]$, where e' is the valuation for which we have:*

1. $e'(x_1) = e(t_1),\ \ldots,\ e'(x_n) = e(t_n)$;
2. *For every variable $y \notin \{x_1, \ldots, x_n\}$, $e'(y) = e(y)$.*

Intuitively $A(x_1/t_1, \ldots, x_n/t_n)$ is A with $x_1, \ldots, x_n$ remapped to $t_1, \ldots, t_n$, respectively. For instance, if A is the predicate/elementary game $x < y$, then $A(x/y, y/x)$ is $y < x$, $A(x/y)$ is $y < y$, $A(y/3)$ is $x < 3$, and $A(z/3)$—where z is different from x, y—remains $x < y$ because A does not depend on z.

Following the standard readability-improving practice established in the literature for predicates, we will often fix a tuple $(x_1, \ldots, x_n)$ of pairwise distinct variables for a game A and write A as $A(x_1, \ldots, x_n)$. It should be noted that when doing so, by no means do we imply that $x_1, \ldots, x_n$ are all of (or only) the variables on which A depends. Representing A in the form $A(x_1, \ldots, x_n)$ sets a context in which we can write $A(t_1, \ldots, t_n)$ to mean the same as the more clumsy expression $A(x_1/t_1, \ldots, x_n/t_n)$. So, if the game $x < y$ is represented as $A(x)$, then $A(3)$ will mean $3 < y$ and $A(y)$ mean $y < y$. And if the same game is represented as $A(y, z)$ (where $z \neq x, y$), then $A(z, 3)$ means $x < z$ while $A(y, 3)$ again means $x < y$.

The entities that in common language we call games are at least as often non-constant as constant. Chess is a classical example of a constant game. On the other hand, many of the card games—including solitaire games where only one player is active—are more naturally represented as non-constant games: each session/instance of such a game is set by a particular permutation of the card deck, and thus the game can be understood as a game that depends on a variable x ranging over the possible settings of the deck. Even the game of checkers—another "classical example" of a constant game—has a natural non-constant generalization $\mathit{Checkers}(x)$ (with x ranging over $\{8, 10, 12, 14, \ldots\}$), meaning a play on the board of size $x \times x$ where, in the initial position, the first $\frac{3}{2}x$ black cells are filled with white pieces and the last $\frac{3}{2}x$ black cells with black pieces. Then the ordinary checkers can be written as $\mathit{Checkers}(8)$. Furthermore, the numbers of pieces of either color also can be made variable, getting an even more general game $\mathit{Checkers}(x, y, z)$, with the ordinary checkers being the instance $\mathit{Checkers}(8, 12, 12)$ of it. By further allowing rectangular- (rather

than just square-) shape boards, we would get a game that depends on four variables, etc. Computability theory texts also often appeal to non-constant games to illustrate certain complexity-theory concepts such as alternating computation or PSPACE-completeness. The *Formula Game* or *Generalized Geography* (Sipser, 2006, Section 8.3) are typical examples. Both can be understood as games that depend on a variable x, with x ranging over quantified Boolean formulas in Formula Game and over directed graphs in Generalized Geography.

A game A is said to be **unistructural in** a variable x—or simply x**-unistructural**—iff, for every two valuations e_1 and e_2 that agree on all variables except x, we have $\mathbf{Lr}^A_{e_1} = \mathbf{Lr}^A_{e_2}$. And A is (simply) **unistructural** iff $\mathbf{Lr}^A_{e_1} = \mathbf{Lr}^A_{e_2}$ for any two valuations e_1 and e_2. A unistructural game is thus a game whose every instance has the same structure (the **Lr** component). And A is unistructural in x iff the structure of any instance $e[A]$ of A does not depend on how e evaluates the variable x. Of course, every constant or elementary game is unistructural, and every unistructural game is unistructural in all variables. While natural examples of non-unistructural games exist such as the games mentioned in the above paragraph, all examples of particular games discussed elsewhere in the present paper are unistructural. In fact, every non-unistructural game can be rather easily rewritten into an equivalent (in a certain reasonable sense) unistructural game. One of the standard ways to convert a non-unistructural game A into a corresponding unistructural game A' is to take the union (or anything bigger) U of the structures of all instances of A to be the common-for-all-instances structure of A', and then extend the (relevant part of the) **Wn** function of each instance $e[A]$ of A to U by stipulating that, if $\Gamma \in (U - \mathbf{Lr}^A_e)$, then the player who made the first illegal (in the sense of $e[A]$) move is the loser in $e[A']$. So, say, in the unistructural version of generalized checkers, an attempt by a player to move to a non-existing cell would result in a loss for that player but otherwise considered a legal move. The class of naturally emerging unistructural games is very wide. All elementary games are trivially there, and Theorem 14.1 of Japaridze (2003) establishes that all of the game operations studied in CL preserve the unistructural property of games. In view of these remarks, if the reader feels more comfortable this way, without much loss of generality (s)he can always understand "game" as "unistructural game".

What makes unistructural games nice is that, even when non-constant, they can still be visualized in the style of Figures 11.2 and 11.3. The difference will be that whereas the nodes of a game tree of a constant game are always labeled by propositions ($\top$ or $\bot$), now such labels can be any predicates. The constant game of Figure 11.3 was about the problem of computing $n+1$. We can generalize it to the problem of computing $n + z$, where z is a (the only) variable on which the game depends. The corresponding non-constant game then can be drawn by modifying the labels of the bottom-level nodes of Figure 11.3 as follows:

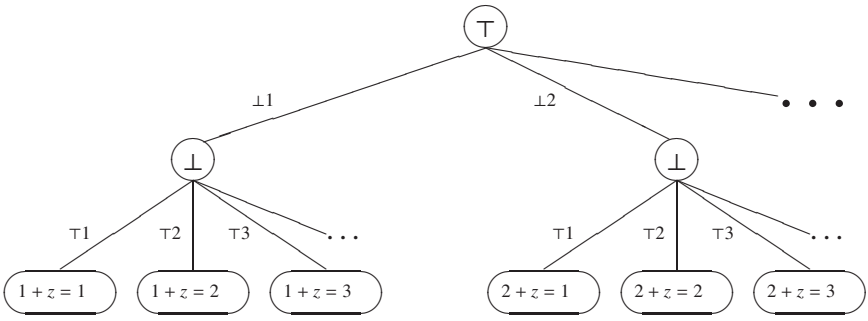


Figure 11.5: The problem of computing $n + z$

Denoting the above game by $A(z)$, the game of Figure 11.3 becomes the instance $A(1)$ of it. The latter results from replacing z by 1 in the tree of Figure 11.5. This replacement turns every label $n + z = m$ into the constant game/proposition $n + 1 = m$, i.e.—depending on its truth value—into $\top$ or $\bot$.

Let A be an arbitrary game. We say that Γ is a **unilegal run** (position if finite) of A iff, for every valuation e, Γ is a legal run of $e[A]$. The set of all unilegal runs of A is denoted by $\mathbf{LR}^A$. Of course, for unistructural games, "legal" and "unilegal" mean the same. The operation of prefixation defined in Section 11.2 only for constant games naturally extends to all games. For $\langle\Phi\rangle A$ to be defined, Φ should be a unilegal position of A. Once this condition is satisfied, we define $\langle\Phi\rangle A$ as the unique game such that, for every valuation e, $e[\langle\Phi\rangle A] = \langle\Phi\rangle e[A]$. For example, where $A(z)$ is the game of Figure 11.5, $\langle\bot 1\rangle A(z)$ is the subtree rooted at the first (leftmost) child of the root, and $\langle\bot 1, \top 2\rangle A(z)$ is the subtree rooted at the second grandchild from the first child, i.e. simply the predicate $1 + z = 2$.

Computability logic can be seen as an approach that generalizes both the traditional theory of computation and traditional logic, and unifies them on the basis of one general formal framework. The main objects of study of the traditional theory of computation are traditional computational problems, and the main objects of study of traditional logic are predicates. Both of these sorts of objects turn out to be special cases of our games. So, one can characterize classical logic as the elementary—non-interactive—fragment of computability logic. And characterize (the core of) the traditional theory of computation as the fragment of computability logic where interaction is limited to its simplest, two-step— input/output, or question/answer—form. The basic entities on which such a unifying framework needs to focus are thus games, and nothing but games.

11.4 Game operations

As we already know, logical operators in CL stand for operations on games. There is an open-ended pool of operations of potential interest, and which of those to study may depend on particular needs and taste. Yet, there is a core collection of the most basic and natural game operations, to the definitions of which the present section is devoted: the **propositional connectives**[2] $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\sqcap$, $\sqcup$, $\curlywedge$, $\curlyvee$, $\succ\!\!-$, $\diamondsuit$, $\Diamond$, $\circ\!\!-$ and the **quantifiers** $\bigsqcap$, $\bigsqcup$, $\bigwedge$, $\bigvee$, $\forall$, $\exists$. Among these we see all operators of classical logic, and our choice of the classical notation for them is no accident. It was pointed out earlier that classical logic is nothing but the elementary, zero-interactivity fragment of computability logic. Indeed, after analyzing the relevant definitions, each of the classically-shaped operations, *when restricted to elementary games*, can be easily seen to be virtually the same as the corresponding operator of classical logic. For instance, if A and B are elementary games, then so is $A \wedge B$, and the latter is exactly the classical conjunction of A and B understood as an (elementary) game. In a general—not-necessarily-elementary—case, however, $\neg$, $\wedge$, $\vee$, $\rightarrow$ become more reminiscent of (yet not the same as) the corresponding multiplicative operators of linear logic. Of course, here we are essentially comparing apples with oranges for, as noted earlier, linear logic is a syntax while computability logic is a semantics, and it may be not clear in what precise sense one can talk about similarities or differences. In the same apples and oranges style, our operations $\sqcap$, $\sqcup$, $\bigsqcap$, $\bigsqcup$ can be perceived as relatives of the additive connectives and quantifiers of linear logic, $\bigwedge$, $\bigvee$ as "multiplicative quantifiers", and $\curlywedge$, $\curlyvee$, $\diamondsuit$, $\Diamond$ as "exponentials", even though it is hard to guess which of the two groups—$\curlywedge$, $\curlyvee$ or $\diamondsuit$, $\Diamond$—would be closer to an orthodox linear logician's heart. The quantifiers $\forall$, $\exists$, on the other hand, hardly have any reasonable linear-logic counterparts.

Let us agree that in every definition of this section x stands for an arbitrary variable, $A, B, A(x), A_1, A_2, \ldots$ for arbitrary games, e for an arbitrary valuation, and Γ for an arbitrary run. Note that it is sufficient to define the content (**Wn** component) of a given constant game only for its legal runs, for then it uniquely extends to all runs. Furthermore, as usually done in logic textbooks and as we already did with the operation of prefixation, propositional connectives can be initially defined just as operations on constant games; then they automatically extend to all games by stipulating that $e[\ldots]$ simply commutes with all of those operations. That is, $\neg A$ is the unique game such that, for every e, $e[\neg A] = \neg e[A]$; $e[A_1 \wedge A_2]$ is the unique game such that, for every e, $e[A_1 \wedge A_2] = e[A_1] \wedge e[A_2]$, etc. With this remark in mind, in each of our definitions of propositional

[2]The term "propositional" is not very adequate here, and we use it only by inertia from classical logic. Propositions are very special—elementary and constant—cases of games. On the other hand, our "propositional" operations are applicable to all games, and not all of them preserve the elementary property of their arguments, even though they do preserve the constant property.

connectives that follow in this section, games $A, B, A_1, A_2, \ldots$ are implicitly assumed to be constant. Alternatively, this assumption can be dropped; all one needs to change in the corresponding definitions in this case is to write $\mathbf{Lr}_e^A$ and $\mathbf{Wn}_e^A$ instead of simply $\mathbf{Lr}^A$ and $\mathbf{Wn}^A$.

For similar reasons, it would be sufficient to define QxA (where Q is a quantifier) just for 1-ary games A that only depend on x. Since we are lazy to explain how, exactly, Qx would then extend to all games, our definitions of quantifiers given in this section, unlike those of propositional connectives, neither explicitly nor implicitly assume any conditions on the arity of A.

11.4.1 Negation

Negation $\neg$ is the role-switch operation: it turns $\top$'s wins and legal moves into $\bot$'s wins and legal moves, and vice versa. For instance, if *Chess* is the game of chess from the point of view of the white player, then $\neg$*Chess* is the same game as seen by the black player. Figure 11.6 illustrates how applying $\neg$ to a game A generates the exact "negative image" of A, with $\top$ and $\bot$ interchanged both in the nodes and the arcs of the game tree.

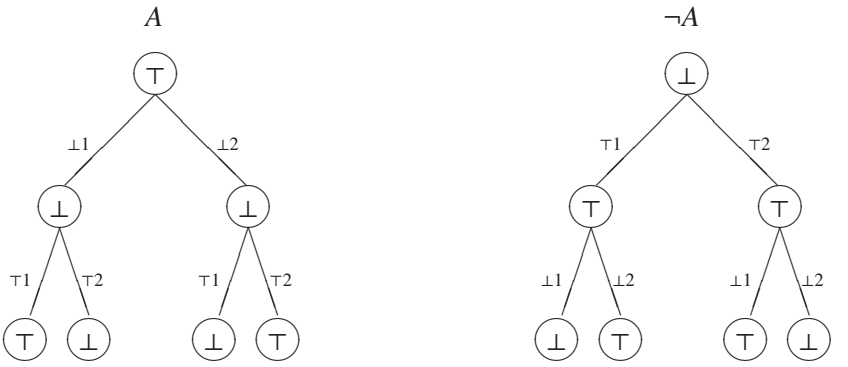


Figure 11.6: Negation

Notice the three different meanings that we associate with symbol $\neg$. In Section 11.2 we agreed to use $\neg$ as an operation on players (turning $\top$ into $\bot$ and vice versa), and an operation on runs (interchanging $\top$ with $\bot$ in every labmove). Below comes our formal definition of the third meaning of $\neg$ as an operation on games:

Definition 6. Negation $\neg A$*:*

- $\Gamma \in \mathbf{Lr}^{\neg A}$ *iff* $\neg\Gamma \in \mathbf{Lr}^A$.
- $\mathbf{Wn}^{\neg A}\langle\Gamma\rangle = \top$ *iff* $\mathbf{Wn}^A\langle\neg\Gamma\rangle = \bot$.

Even from the informal explanation of $\neg$ it is clear that $\neg\neg A$ is always the same as A, for interchanging in A the payers' roles twice brings the players to their original roles. It would also be easy to show that we always have $\neg(\langle\Phi\rangle A) = \langle\neg\Phi\rangle\neg A$. So, say, if α is $\top$'s legal move in the empty position of A that brings A down to B, then the same α is $\bot$'s legal move in the empty position of $\neg A$, and it brings $\neg A$ down to $\neg B$. Test the game A of Figure 11.6 to see that this is indeed so.

11.4.2 Choice operations

$\sqcap, \sqcup, \bigsqcap$ and $\bigsqcup$ are called **choice operations**. $A_1 \sqcap A_2$ is the game where, in the initial position, $\bot$ has two legal moves (choices): 1 and 2. Once such a choice i is made, the game continues as the chosen component A_i, meaning that $\langle\bot i\rangle(A_1 \sqcap A_2) = A_i$; if a choice is never made, $\bot$ loses. $A_1 \sqcup A_2$ is similar/symmetric, with $\top$ and $\bot$ interchanged; that is, in $A_1 \sqcup A_2$ it is $\top$ who makes an initial choice and who loses if such a choice is never made. Figure 11.7 helps us visualize the way $\sqcap$ and $\sqcup$ combine two games A and B:

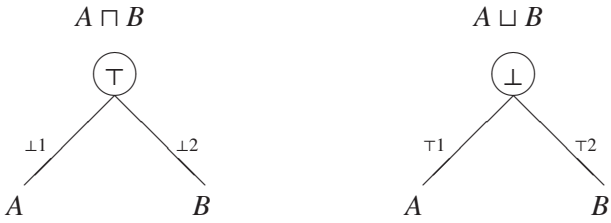


Figure 11.7: Choice propositional connectives

The game A of Figure 11.6 can now be easily seen to be $(\top\sqcup\bot)\sqcap(\bot\sqcup\top)$, and its negation be $(\bot\sqcap\top)\sqcup(\top\sqcap\bot)$. The symmetry/duality familiar from classical logic persists: we always have $\neg(A \sqcap B) = \neg A \sqcup \neg B$ and $\neg(A \sqcup B) = \neg A \sqcap \neg B$. Similarly for the quantifier counterparts $\bigsqcap$ and $\bigsqcup$ of $\sqcap$ and $\sqcup$. We might have already guessed that $\bigsqcap xA(x)$ is nothing but the infinite $\sqcap$-conjunction $A(1) \sqcap A(2) \sqcap A(3) \sqcap \ldots$ and $\bigsqcup xA(x)$ is $A(1) \sqcup A(2) \sqcup A(3) \sqcup \ldots$, as can be seen from Figure 11.8.

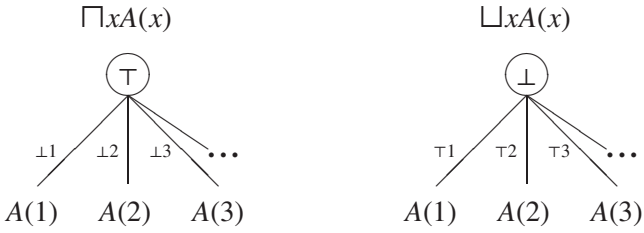


Figure 11.8: Choice quantifiers

So, we always have $\langle \bot c \rangle \sqcap x A(x) = A(c)$ and $\langle \top c \rangle \sqcup x A(x) = A(c)$. The meaning of such a labmove $\wp c$ can be characterized as that player $\wp$ selects/specifies the particular value c for x, after which the game continues—and the winner is determined—according to the rules of $A(c)$.

Now we are already able to express traditional computational problems using formulas. Traditional problems come in two forms: the problem of computing a function $f(x)$, or the problem of deciding a predicate $p(x)$. The former can be captured by $\sqcap x \sqcup y(f(x) = y)$, and the latter (which, of course, can be seen as a special case of the former) by $\sqcap x(p(x) \sqcup \neg p(x))$. So, the game of Figure 11.3 will be written as $\sqcap x \sqcup y(x + 1 = y)$, and the game of Figure 11.5 as $\sqcap x \sqcup y(x + z = y)$.

The following Definition 7 summarizes the above-said, and generalizes $\sqcap, \sqcup$ from binary to any $\geq$ 2-ary operations. Note the perfect symmetry in it: the definition of each choice operation can be obtained from that of its dual by just interchanging $\top$ with $\bot$.

Definition 7. *In clauses 1 and 2, n is* 2 *or any greater integer.*

1. **Choice conjunction** $A_1 \sqcap \ldots \sqcap A_n$*:*

   - $\mathbf{Lr}^{A_1 \sqcap \ldots \sqcap A_n} = \{\langle\rangle\} \cup \{\langle \bot i, \Gamma \rangle \mid i \in \{1, \ldots, n\},\ \Gamma \in \mathbf{Lr}^{A_i}\}$.

   - $\mathbf{Wn}^{A_1 \sqcap \ldots \sqcap A_n}\langle\rangle = \top$*;*
     where $i \in \{1, \ldots, n\}$, $\mathbf{Wn}^{A_1 \sqcap \ldots \sqcap A_n}\langle \bot i, \Gamma \rangle = \mathbf{Wn}^{A_i}\langle \Gamma \rangle$.

2. **Choice disjunction** $A_1 \sqcup \ldots \sqcup A_n$*:*

   - $\mathbf{Lr}^{A_1 \sqcup \ldots \sqcup A_n} = \{\langle\rangle\} \cup \{\langle \top i, \Gamma \rangle \mid i \in \{1, \ldots, n\},\ \Gamma \in \mathbf{Lr}^{A_i}\}$.

   - $\mathbf{Wn}^{A_1 \sqcup \ldots \sqcup A_n}\langle\rangle = \bot$*;*
     where $i \in \{1, \ldots, n\}$, $\mathbf{Wn}^{A_1 \sqcup \ldots \sqcup A_n}\langle \top i, \Gamma \rangle = \mathbf{Wn}^{A_i}\langle \Gamma \rangle$.

3. **Choice universal quantification** $\sqcap x A(x)$*:*

   - $\mathbf{Lr}^{e[\sqcap x A(x)]} = \{\langle\rangle\} \cup \{\langle \bot c, \Gamma \rangle \mid c \in \{1, 2, 3, \ldots\},\ \Gamma \in \mathbf{Lr}^{e[A(c)]}\}$.

   - $\mathbf{Wn}^{e[\sqcap x A(x)]}\langle\rangle = \top$*;*
     where $c \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{e[\sqcap x A(x)]}\langle \bot c, \Gamma \rangle = \mathbf{Wn}^{e[A(c)]}\langle \Gamma \rangle$.

4. **Choice existential quantification** $\sqcup x A(x)$*:*

   - $\mathbf{Lr}^{e[\sqcup x A(x)]} = \{\langle\rangle\} \cup \{\langle \top c, \Gamma \rangle \mid c \in \{1, 2, 3, \ldots\},\ \Gamma \in \mathbf{Lr}^{e[A(c)]}\}$.

   - $\mathbf{Wn}^{e[\sqcup x A(x)]}\langle\rangle = \bot$*;*
     where $c \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{e[\sqcup x A(x)]}\langle \top c, \Gamma \rangle = \mathbf{Wn}^{e[A(c)]}\langle \Gamma \rangle$.

11.4.3 Parallel operations

The operations $\wedge, \vee, \bigwedge, \bigvee$ combine games in a way that corresponds to our intuition of parallel computations. For this reason we call such operations **parallel**. Playing $A_1 \wedge A_2$ (resp. $A_1 \vee A_2$) means playing the two games simultaneously where, in order to win, $\top$ needs to win in both (resp. at least one) of the components A_i. Back to our chess example, the two-board game $\neg Chess \vee Chess$ can be easily won by just mimicking in *Chess* the moves that the adversary makes in $\neg Chess$, and vice versa. This is very different from the situation with $\neg Chess \sqcup Chess$, winning which is not easy at all: there $\top$ needs to choose between $\neg Chess$ and *Chess* (i.e. between playing black or white), and then win the chosen one-board game. Technically, a move α in the kth $\wedge$-conjunct or $\vee$-disjunct is made by prefixing α with 'k.'. For instance, in (the initial position of) $(A \sqcup B) \vee (C \sqcap D)$, the move '2.1' is legal for $\bot$, meaning choosing the first $\sqcap$-conjunct in the second $\vee$-disjunct of the game. If such a move is made, the game will continue as $(A \sqcup B) \vee C$. The player $\top$, too, has initial legal moves in $(A \sqcup B) \vee (C \sqcap D)$, which are '1.1' and '1.2'. As we may guess, $\bigwedge x A(x)$ is nothing but $A(1) \wedge A(2) \wedge A(3) \wedge \ldots$, and $\bigvee x A(x)$ is nothing but $A(1) \vee A(2) \vee A(3) \vee \ldots$.

The following formal definition summarizes this meaning of parallel operations, generalizing the arity of $\wedge, \vee$ to any $n \geq 2$. In that definition and throughout the rest of this paper, we use the important notational convention according to which, for a string/move α,

$$\Gamma^{\alpha}$$

means the result of removing from Γ all (lab)moves except those of the form $\wp\alpha\beta$, and then deleting the prefix[3] 'α' in the remaining moves, i.e. replacing each such $\wp\alpha\beta$ by $\wp\beta$. For example, where Γ is the leftmost branch of the tree for $(\top \sqcap \bot) \vee (\bot \sqcup \top)$ shown in Figure 11.9, we have $\Gamma^{1\cdot} = \langle \bot 1 \rangle$ and $\Gamma^{2\cdot} = \langle \top 1 \rangle$. Intuitively, we view this Γ as consisting of two subruns, one ($\Gamma^{1\cdot}$) being a run in the first $\vee$-disjunct of $(\top \sqcap \bot) \vee (\bot \sqcup \top)$, and the other ($\Gamma^{2\cdot}$) being a run in the second disjunct.

Definition 8. *In clauses 1 and 2, n is 2 or any greater integer.*

1. **Parallel conjunction** $A_1 \wedge \ldots \wedge A_n$*:*

- $\Gamma \in \mathbf{Lr}^{A_1 \wedge \ldots \wedge A_n}$ *iff every move of* Γ *has the prefix 'i.' for some* $i \in \{1, \ldots, n\}$ *and, for each such* i, $\Gamma^{i\cdot} \in \mathbf{Lr}^{A_i}$.
- *Whenever* $\Gamma \in \mathbf{Lr}^{A_1 \wedge \ldots \wedge A_n}$, $\mathbf{Wn}^{A_1 \wedge \ldots \wedge A_n}\langle \Gamma \rangle = \top$ *iff, for each* $i \in \{1, \ldots, n\}$, $\mathbf{Wn}^{A_i}\langle \Gamma^{i\cdot} \rangle = \top$.

[3] Here and later, when talking about a prefix of a labmove $\wp\gamma$, we do not count the label $\wp$ as a part of the prefix.

2. **Parallel disjunction** $A_1 \vee \ldots \vee A_n$*:*

- $\Gamma \in \mathbf{Lr}^{A_1 \vee \ldots \vee A_n}$ *iff every move of* Γ *has the prefix 'i.' for some* $i \in \{1, \ldots, n\}$ *and, for each such* i, $\Gamma^{i\cdot} \in \mathbf{Lr}^{A_i}$.
- *Whenever* $\Gamma \in \mathbf{Lr}^{A_1 \vee \ldots \vee A_n}$, $\mathbf{Wn}^{A_1 \vee \ldots \vee A_n}\langle\Gamma\rangle = \bot$ *iff, for each* $i \in \{1, \ldots, n\}$, $\mathbf{Wn}^{A_i}\langle\Gamma^{i\cdot}\rangle = \bot$.

3. **Parallel universal quantification** $\bigwedge x A(x)$*:*

- $\Gamma \in \mathbf{Lr}^{e[\bigwedge x A(x)]}$ *iff every move of* Γ *has the prefix 'c.' for some* $c \in \{1, 2, 3, \ldots\}$ *and, for each such* c, $\Gamma^{c\cdot} \in \mathbf{Lr}^{e[A(c)]}$.
- *Whenever* $\Gamma \in \mathbf{Lr}^{e[\bigwedge x A(x)]}$, $\mathbf{Wn}^{e[\bigwedge x A(x)]}\langle\Gamma\rangle = \top$ *iff, for each* $c \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{e[A(c)]}\langle\Gamma^{c\cdot}\rangle = \top$.

4. **Parallel existential quantification** $\bigvee x A(x)$*:*

- $\Gamma \in \mathbf{Lr}^{e[\bigvee x A(x)]}$ *iff every move of* Γ *has the prefix 'c.' for some* $c \in \{1, 2, 3, \ldots\}$ *and, for each such* c, $\Gamma^{c\cdot} \in \mathbf{Lr}^{e[A(c)]}$.
- *Whenever* $\Gamma \in \mathbf{Lr}^{e[\bigvee x A(x)]}$, $\mathbf{Wn}^{e[\bigvee x A(x)]}\langle\Gamma\rangle = \bot$ *iff, for each* $c \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{e[A(c)]}\langle\Gamma^{c\cdot}\rangle = \bot$.

As was the case with choice operations, we can see that the definition of each of the parallel operations can be obtained from the definition of its dual by just interchanging $\top$ with $\bot$. Hence it is easy to verify that we always have $\neg(A \wedge B) = \neg A \vee \neg B$, $\neg(A \vee B) = \neg A \wedge \neg B$, $\neg \bigwedge x A(x) = \bigvee x \neg A(x)$, $\neg \bigvee x A(x) = \bigwedge x \neg A(x)$.

Note also that just like negation (and unlike choice operations), parallel operations preserve the elementary property of games and, when restricted to elementary games, the meanings of $\wedge$ and $\vee$ coincide with those of classical conjunction and disjunction, while the meanings of $\bigwedge$ and $\bigvee$ coincide with those of classical universal quantifier and existential quantifier. The same conservation of classical meaning is going to be the case with the blind quantifiers $\forall, \exists$ defined later; so, at the elementary level, $\bigwedge$ and $\bigvee$ are indistinguishable from $\forall$ and $\exists$.

A strict definition of our understanding of validity—which, as we may guess, conserves the classical meaning of this concept in the context of elementary games—will be given later in Section 11.7. For now, let us adopt an intuitive explanation according to which validity means being "always winnable by a machine". While all classical tautologies automatically remain valid when parallel operators are applied to elementary games, in the general case the class of valid principles shrinks. For example, $\neg P \vee (P \wedge P)$ is not valid. Proving this

might require some thought, but at least we can see that the earlier "mimicking" ("copy-cat") strategy successful for $\neg Chess \vee Chess$ would be inapplicable to $\neg Chess \vee (Chess \wedge Chess)$. The best that $\top$ can do in this three-board game is to pair $\neg Chess$ with one of the two conjuncts of $Chess \wedge Chess$. It is possible that then $\neg Chess$ and the unmatched $Chess$ are both lost, in which case the whole game will be lost.

When A and B are finite (or finite-depth) games, the depth of $A \wedge B$ or $A \vee B$ is the sum of the depths of A and B, which signifies an exponential growth of the breadth. Figure 11.9 illustrates this growth, suggesting that once we have reached the level of parallel operations—let alone recurrence operations that will be defined shortly—continuing drawing trees in the earlier style becomes no fun. Not to be disappointed though: making it possible to express large- or infinite-size game trees in a compact way is what our game operators are all about after all.

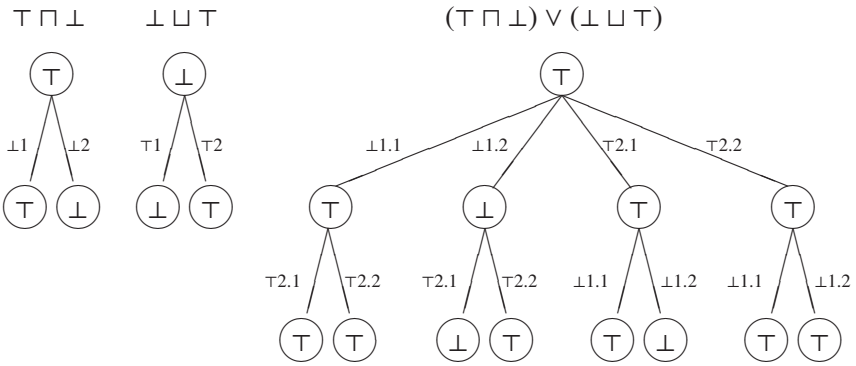


Figure 11.9: Parallel disjunction

An alternative approach to graphically representing $A \vee B$ (or $A \wedge B$) would be to just draw two trees—one for A and one for B—next to each other rather than draw one tree for $A \vee B$. The legal positions of $A \vee B$ can then be visualized as pairs (Φ, Ψ), where Φ is a node of the A-tree and Ψ a node of the B-tree; the "label" of each such position (Φ, Ψ) will be $\top$ iff the label of at least one (or both if we are dealing with $A \wedge B$) of the positions/nodes Φ, Ψ in the corresponding tree is $\top$. For instance, the root of the $(\top \sqcap \bot) \vee (\bot \sqcup \top)$-tree of Figure 11.9 can just be thought of as the pair consisting of the roots of the $(\top \sqcap \bot)$- and $(\bot \sqcup \top)$-trees; child #1 of the root of the $(\top \sqcap \bot) \vee (\bot \sqcup \top)$-tree as the pair whose first node is the left child of the root of the $(\top \sqcap \bot)$-tree and the second node is the root of the $(\bot \sqcup \top)$-tree, etc. It is true that, under this approach, a pair (Φ, Ψ) might correspond to more than one position of $A \vee B$. For example, grandchildren #1 and #5 of the root of the $(\top \sqcap \bot) \vee (\bot \sqcup \top)$-tree,

i.e. the positions $\langle \bot 1.1, \top 2.1\rangle$ and $\langle \top 2.1, \bot 1.1\rangle$, would become indistinguishable. This, however, is OK, because such two positions would always be equivalent, in the sense that

$$\langle \bot 1.1, \top 2.1\rangle((\top \sqcap \bot) \vee (\bot \sqcup \top)) = \langle \top 2.1, \bot 1.1\rangle((\top \sqcap \bot) \vee (\bot \sqcup \top)).$$

Whether trees are or are not helpful in visualizing parallel combinations of games, prefixation is still very much so if we think of each (uni)legal position Φ of A as the game $\langle \Phi\rangle A$. This way, every (uni)legal run Γ of A becomes a sequence of games.

Example 9. To the legal run $\langle \bot 2.7, \top 1.7, \bot 1.49, \top 2.49\rangle$—call it Γ—of game $\sqcup x \sqcap y(y \neq x^2) \vee \sqcap x \sqcup y(y = x^2)$—call it A—corresponds the following sequence, showing how things evolve as Γ runs, i.e. how the moves of Γ affect/modify the game that is being played:

A_0: $\sqcup x \sqcap y(y \neq x^2) \vee \sqcap x \sqcup y(y = x^2)$, i.e. A,
i.e. $\langle\rangle A$;

A_1: $\sqcup x \sqcap y(y \neq x^2) \vee \sqcup y(y = 7^2)$, i.e. $\langle \bot 2.7\rangle A_0$,
i.e. $\langle \bot 2.7\rangle A$;

A_2: $\sqcap y(y \neq 7^2) \vee \sqcup y(y = 7^2)$, i.e. $\langle \top 1.7\rangle A_1$,
i.e. $\langle \bot 2.7, \top 1.7\rangle A$;

A_3: $49 \neq 7^2 \vee \sqcup y(y = 7^2)$, i.e. $\langle \bot 1.49\rangle A_2$,
i.e. $\langle \bot 2.7, \top 1.7, \bot 1.49\rangle A$;

A_4: $49 \neq 7^2 \vee 49 = 7^2$, i.e. $\langle \top 2.49\rangle A_3$,
i.e. $\langle \bot 2.7, \top 1.7, \bot 1.49, \top 2.49\rangle A$.

The run hits the true proposition A_4, and hence is won by $\top$.

When visualizing $\wedge, \vee$-games in a similar style, we are better off representing them as infinite conjunctions/disjunctions. Of course, putting infinitely many conjuncts/disjuncts on paper would be no fun. But, luckily, in every position of $\wedge x A(x)$ or $\vee x A(x)$ only a finite number of conjuncts/disjuncts would be "activated", i.e. have a non-$A(c)$ form, so that all of the other, uniform, conjuncts can be combined into blocks and represented, say, through an ellipsis, or through expressions such as $\wedge m \leq x \leq n A(x)$ or $\wedge x \geq m A(x)$.

Example 10. Let $Odd(x)$ be the predicate "x is odd". The $\top$-won legal run $\langle \top 7.1\rangle$ of $\vee x(Odd(x) \sqcup \neg Odd(x))$ will be represented as follows:

$\vee x \geq 1(Odd(x) \sqcup \neg Odd(x))$;
$\vee 1 \leq x \leq 6(Odd(x) \sqcup \neg Odd(x)) \vee Odd(7) \vee \vee x \geq 8(Odd(x) \sqcup \neg Odd(x))$.

And the infinite legal run $\Gamma = \langle \top 1.1, \top 2.2, \top 3.1, \top 4.2, \top 5.1, \top 6.2, \ldots\rangle$ of $\wedge x(Odd(x) \sqcup \neg Odd(x))$ will be represented as follows:

$\bigwedge x{\geq}1(Odd(x) \sqcup \neg Odd(x))$;
$Odd(1) \wedge \bigwedge x{\geq}2(Odd(x) \sqcup \neg Odd(x))$;
$Odd(1) \wedge \neg Odd(2) \wedge \bigwedge x{\geq}3(Odd(x) \sqcup \neg Odd(x))$;
$Odd(1) \wedge \neg Odd(2) \wedge Odd(3) \wedge \bigwedge x{\geq}4(Odd(x) \sqcup \neg Odd(x))$;
. . . etc.

Note that Γ is won by $\top$ but every finite initial segment of it is lost.

11.4.4 Reduction

What we call **reduction** $\rightarrow$ is perhaps most interesting of all operations, yet we do not introduce $\rightarrow$ as a primitive operation as it can be formally defined by

$$B \rightarrow A \;=\; (\neg B) \vee A.$$

From this definition we see that, when applied to elementary games, $\rightarrow$ has its ordinary classical meaning, because so do $\neg$ and $\vee$.

Intuitively, $B \rightarrow A$ is (indeed) the problem of *reducing* A to B: solving $B \rightarrow A$ means solving A while having B as a *computational resource*. Resources are symmetric to problems: what is a problem to solve for one player is a resource that the other player can use, and vice versa. Since B is negated in $\neg B \vee A$ and negation means switching the roles, B appears as a resource rather than problem for $\top$ in $B \rightarrow A$. For example, the game of Example 9 can be written as $\sqcap x \sqcup y(y = x^2) \rightarrow \sqcap x \sqcup y(y = x^2)$. For $\top$, $\sqcap x \sqcup y(y = x^2)$ is the problem of computing square, which can be seen as a task (telling the square of any given number) performed by $\top$ for $\bot$. But in the antecedent it turns into a square-computing resource—a task performed by $\bot$ for $\top$. In the run Γ of Example 9, $\top$ took advantage of this fact, and solved problem $\sqcap x \sqcup y(y = x^2)$ in the consequent using $\bot$'s solution to the same problem in the antecedent. That is, $\top$ reduced $\sqcap x \sqcup y(y = x^2)$ to $\sqcap x \sqcup y(y = x^2)$.

To get a better appreciation of $\rightarrow$ as a problem reduction operation, let us look a less trivial—already "classical" in CL—example. Let $A(x, y)$ be the predicate "Turing machine (whose code is) x accepts input y", and $H(x, y)$ the predicate "Turing machine x halts on input y". Note that then $\sqcap x \sqcap y(A(x, y) \sqcup \neg A(x, y))$ expresses the acceptance problem as a decision problem: in order to win, $\top$ should be able to tell which of the disjuncts—$A(x, y)$ or $\neg A(x, y)$—is true for any particular values for x and y selected by the environment. Similarly, $\sqcap x \sqcap y(H(x, y) \sqcup \neg H(x, y))$ expresses the halting problem as a decision problem. No machine can (always) win $\sqcap x \sqcap y(A(x, y) \sqcup \neg A(x, y))$ because the acceptance problem, just as the halting problem, is known to be undecidable. Yet, the acceptance problem is algorithmically reducible to the halting problem. Into our terms, this fact translates as existence of a machine that always wins the game

$$\sqcap x \sqcap y(H(x, y) \sqcup \neg H(x, y)) \;\rightarrow\; \sqcap x \sqcap y(A(x, y) \sqcup \neg A(x, y)). \qquad (1)$$

A successful strategy for such a machine ($\top$) is as follows. At the beginning, $\top$ waits till $\bot$ specifies some values m and n for x and y in the consequent, i.e. makes the moves '2.m' and '2.n'. Such moves, bringing the consequent down to $A(m,n) \sqcup \neg A(m,n)$, can be seen as asking $\top$ the question "does Turing machine m accept input n?". To this question $\top$ replies by the counterquestion "does m halt on n?", i.e. makes the moves '1.m and '1.n', bringing the antecedent down to $H(m,n) \sqcup \neg H(m,n)$. The environment has to answer this counterquestion, or else it loses. If it answers "no" (i.e. makes the move '1.2' and thus further brings the antecedent down to $\neg H(m,n)$), $\top$ also answers "no" to the original question in the consequent (i.e. makes the move '2.2'), with the overall game having evolved to the true and hence $\top$-won proposition/elementary game $\neg H(m,n) \rightarrow \neg A(m,n)$. Otherwise, if the environment's answer is "yes" (move '1.1'), $\top$ simulates Turing machine m on input n until it halts, and then makes the move '2.1' or '2.2' depending whether the simulation accepted or rejected. Of course, it is a possibility that such a simulation goes on forever and thus no moves will be made by $\top$ in the consequent. This, however, will only happen when $H(m,n)$—the $\sqcup$-disjunct selected by the environment in the antecedent—is false, and having lied in the antecedent makes $\bot$ the loser no matter what happens in the consequent.

Again, what the machine did in the above strategy indeed was a reduction: it used an (external) solution to the halting problem to solve the acceptance problem. There are various natural concepts of reduction, and a strong case can be made in favor of the thesis that the sort of reduction captured by $\rightarrow$ is most basic among them, with a great variety of other reasonable concepts of reduction being expressible in terms of $\rightarrow$. Among those is *Turing reduction*. It will be discussed a little later when we get to recurrence operations. Another frequently used concept of reduction is *mapping reduction* that we are going to look at in the following paragraph. And yet some other natural concepts of reduction expressible in terms of $\rightarrow$ may or may not have established names. For example, from the above discussion it can be seen that a certain reducibility-style relation holds between the predicates $A(x,y)$ and $H(x,y)$ in an even stronger sense than the algorithmic winnability of (1). In fact, not only (1) is winnable, but also the generally harder-to-win game

$$\sqcap x \sqcap y\big(H(x,y) \sqcup \neg H(x,y) \rightarrow A(x,y) \sqcup \neg A(x,y)\big). \tag{2}$$

This is so because $\top$'s above-described strategy for (1) did not use (while could have used) any values for x and y others than the values chosen for these variables by $\bot$ in the consequent. So, the $\sqcap x \sqcap y$ prefix can be just made external as this is done in (2). It will be seen later that semantically our approach treats free variables as if they were (externally) bound by $\sqcap$. Hence, the winnability of (2), in turn, is the same as simply the winnability of

$$H(x,y) \sqcup \neg H(x,y) \rightarrow A(x,y) \sqcup \neg A(x,y).$$

A predicate $p(\vec{x})$ is said to be **mapping reducible**[4] to a predicate $q(\vec{y})$ iff there is an effective function f such that, for any constants $\vec{c}$, $p(\vec{c})$ is true iff $q(f(\vec{c}))$ is so. Here $\vec{x}$ abbreviates any n-tuple of pairwise distinct variables, $\vec{c}$ any n-tuple of constants, $\vec{y}$ any m-tuple of pairwise distinct variables, and f is a function that sends n-tuples of constants to m-tuples of constants. Using $A \leftrightarrow B$ as an abbreviation for $(A \rightarrow B) \wedge (B \rightarrow A)$ and $\sqcap\vec{z}$ for $\sqcap z_1 \ldots \sqcap z_k$ where $\vec{z} = z_1, \ldots, z_k$ (and similarly for $\sqcup\vec{z}$), it is not hard to see that mapping reducibility of $p(\vec{x})$ to $q(\vec{y})$ means nothing but existence of an algorithmic winning strategy for

$$\sqcap\vec{x}\sqcup\vec{y}(p(\vec{x}) \leftrightarrow q(\vec{y})).$$

Our acceptance predicate $A(x, y)$ can be shown to be mapping reducible to the halting predicate $H(x, y)$, i.e. the game

$$\sqcap x\sqcap y\sqcup x'\sqcup y'(A(x, y) \leftrightarrow H(x', y'))$$

shown to be winnable by a machine. An algorithmic strategy for $\top$ is the following. After $\bot$ selects values m and n for x and y, select the values m' and (the same) n for x' and y', and rest your case. Here m' is the Turing machine that works exactly as m does, with the only difference that whenever m enters its reject state, m' goes into an infinite loop instead, so that m accepts if and only if m' halts. Such an m', of course, can be effectively constructed from m.

Notice that while the standard approaches only allow us to talk about (a whatever sort of) reducibility as a *relation* between problems, in our approach reduction becomes an *operation* on problems, with reducibility as a relation simply meaning computability of the corresponding combination (such as $\sqcap\vec{x}\sqcup\vec{y}(p(\vec{x}) \leftrightarrow q(\vec{y}))$ or $A \rightarrow B$) of games. Similarly for other relations or properties such as the property of *decidability*. The latter becomes the operation of *deciding* if we define the problem of deciding a predicate (or any computational problem) $p(\vec{x})$ as the game $\sqcap\vec{x}(p(\vec{x}) \sqcup \neg p(\vec{x}))$. So, now we can meaningfully ask questions such as "is the reduction of the problem of deciding $p(x)$ to the problem of deciding $q(x)$ reducible to the mapping reduction of $p(x)$ to $q(x)$?". This question would be equivalent to whether the following game has an algorithmic winning strategy:

$$\begin{array}{l}\Big(\sqcap x\sqcup y(p(x) \leftrightarrow q(y))\Big) \rightarrow \\ \Big(\sqcap x(q(x) \sqcup \neg q(x)) \rightarrow \sqcap x(p(x) \sqcup \neg p(x))\Big).\end{array} \tag{3}$$

This problem is indeed algorithmically solvable no matter what particular predicates $p(x)$ and $q(x)$ are, which means that mapping reduction is at least as

[4]This term is adopted from Sipser (2006). The more common but less adequate name for what we call mapping reducibility is **many-one reducibility**.

strong as reduction. Here is a strategy for $\top$: Wait till $\bot$ selects a value k for x in the consequent of the consequent of (3). Then specify the same value k for x in the antecedent of (3), and wait till $\bot$ replies there by selecting a value n for y. Then select the same value n for x in the antecedent of the consequent of (3). $\bot$ will have to respond by 1 or 2 in that component of the game. Repeat that very response in the consequent of the consequent of (3), and celebrate victory.

We are going to see in Section 11.9 that (3) is a legal formula of the language of system **CL4**, which, according to Theorem 34, is sound and complete with respect to the semantics of computability logic. So, had our ad hoc methods failed to find an answer (and this would certainly be the case if we dealt with a more complex computational problem), the existence of a successful algorithmic strategy could have been established by showing that (3) is provable in **CL4**. Moreover, by the first clause of Theorem 34, after finding a **CL4**-proof of (3), we would not only know that an algorithmic solution for (3) exists, but we would also be able to constructively extract such a solution from the proof. On the other hand, the fact that reduction is not as strong as mapping reduction could be established by showing that **CL4** does not prove

$$\Big(\sqcap x(q(x) \sqcup \neg q(x)) \rightarrow \sqcap x(p(x) \sqcup \neg p(x))\Big) \rightarrow \Big(\sqcap x \sqcup y(p(x) \leftrightarrow q(y))\Big). \tag{4}$$

This negative fact, too, can be established effectively as, according to Theorem 32, the relevant fragment of **CL4** is decidable. In fact, the completeness proof for **CL4** given in Japaridze (2007a) shows a way how to actually construct particular predicates—$p(x)$ and $q(x)$ in our case—for which the problem represented by a given **CL4**-unprovable formula has no algorithmic solution.

11.4.5 Blind operations

Another group of core game operations, called **blind**, comprises $\forall$ and its dual $\exists$. Intuitively, playing $\forall xA(x)$ or $\exists xA(x)$ means just playing $A(x)$ "blindly", without knowing the value of x. In $\forall xA(x)$, $\top$ wins iff the play it generates is successful for every possible value of x, while in $\exists xA(x)$ being successful for just one value is sufficient. $\forall$ and $\exists$ thus essentially produce what is called games with *imperfect information* (see Pietarinen, 2002). This sort of a blind play is meaningful or possible—and hence $\forall xA(x)$, $\exists xA(x)$ defined—only when what moves are available (legal) does not depend on the unknown value of x; in other words, when $A(x)$ is unistructural in x.

Definition 11. *Assume $A(x)$ is unistructural in x.*

1. **Blind universal quantification** $\forall xA(x)$*:*

- $\mathbf{Lr}^{e[\forall xA(x)]} = \mathbf{Lr}^{e[A(x)]}$.

- $\mathbf{Wn}^{e[\forall xA(x)]}\langle\Gamma\rangle = \top$ *iff, for every constant* c, $\mathbf{Wn}^{e[A(c)]}\langle\Gamma\rangle = \top$.

2. **Blind existential quantification** $\exists xA(x)$*:*

- $\mathbf{Lr}^{e[\exists xA(x)]} = \mathbf{Lr}^{e[A(x)]}$.
- $\mathbf{Wn}^{e[\exists xA(x)]}\langle\Gamma\rangle = \bot$ *iff, for every constant* c, $\mathbf{Wn}^{e[A(c)]}\langle\Gamma\rangle = \bot$.

As with the other pairs of quantifiers, one can see that we always have $\neg\forall xA(x) = \exists x\neg A(x)$ and $\neg\exists xA(x) = \forall x\neg A(x)$.

To feel the difference between $\forall$ and $\sqcap$, compare the games

$$\sqcap x(Even(x) \sqcup Odd(x))$$

and

$$\forall x(Even(x) \sqcup Odd(x)).$$

Both are about telling whether a given number is even or odd; the difference is only in whether that "given number" is specified (made as a move by $\bot$) or not. The first problem is an easy-to-win, depth-2 game of a structure that we have already seen. The depth of the second game, on the other hand, is 1, with only by the machine to make a move—select the "true" disjunct, which is hardly possible to do as the value of x remains unspecified.

Figure 11.10 shows trees for $Even(x) \sqcup Odd(x)$ and $\forall x(Even(x) \sqcup Odd(x))$ next to each other. Notice that applying $\forall x$ does not change the structure of a (unistructural) game. What it does is that it simply prefixes every node with a $\forall x$ (we do not explicitly see such a prefix on the root because $\forall x\bot = \bot$). This means that we always have $\langle\Phi\rangle\forall xA(x) = \forall x\langle\Phi\rangle A(x)$. Similarly for $\exists x$.

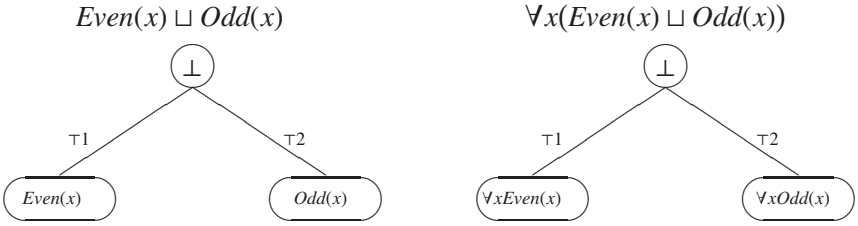


Figure 11.10: Blind universal quantification

Of course, not all nonelementary $\forall$-games will be unwinnable. Here is an example:

$$\forall x\Big(Even(x) \sqcup Odd(x) \rightarrow \sqcap y(Even(x \times y) \sqcup Odd(x \times y))\Big).$$

Solving this problem, which means reducing the consequent to the antecedent without knowing the value of x, is easy: $\top$ waits till $\bot$ selects a value n for y. If n is even, then $\top$ selects the first $\sqcup$-disjunct in the consequent. Otherwise, if n is odd, $\top$ continues waiting until $\bot$ selects one of the $\sqcup$-disjuncts in the antecedent (if $\bot$ has not already done so), and then $\top$ makes the same move 1 or 2 in the consequent as $\bot$ made in the antecedent. One of the possible runs such a strategy can yield is $\langle \bot 1.2, \bot 2.5, \top 2.2\rangle$, which can be visualized as the following sequence of games:

$$\forall x\Big(\mathit{Even}(x) \sqcup \mathit{Odd}(x) \rightarrow \sqcap y(\mathit{Even}(x \times y) \sqcup \mathit{Odd}(x \times y))\Big);$$
$$\forall x\Big(\mathit{Odd}(x) \rightarrow \sqcap y(\mathit{Even}(x \times y) \sqcup \mathit{Odd}(x \times y))\Big);$$
$$\forall x\Big(\mathit{Odd}(x) \rightarrow \mathit{Even}(x \times 5) \sqcup \mathit{Odd}(x \times 5)\Big);$$
$$\forall x\Big(\mathit{Odd}(x) \rightarrow \mathit{Odd}(x \times 5)\Big).$$

By now we have seen three—choice, parallel and blind—natural sorts of quantifiers. The idea of a forth—*sequential*—sort, which we will not discuss here, was outlined in Japaridze (2006e). It is worthwhile to take a brief look at how different quantifiers relate. Both $\forall xA(x)$ and $\wedge xA(x)$ can be shown to be properly stronger than $\sqcap xA(x)$, in the sense that $\forall xP(x) \rightarrow \sqcap xP(x)$ and $\wedge xP(x) \rightarrow \sqcap xP(x)$ are valid while $\sqcap xP(x) \rightarrow \forall xP(x)$ and $\sqcap xP(x) \rightarrow \wedge xP(x)$ are not. On the other hand, the strengths of $\forall xP(x)$ and $\wedge xP(x)$ are mutually incomparable: neither $\forall xP(x) \rightarrow \wedge xP(x)$ nor $\wedge xP(x) \rightarrow \forall xP(x)$ is valid. The big difference between $\forall$ and $\wedge$ is that, while playing $\forall xA(x)$ means playing one "common" play for all possible $A(c)$ and thus $\forall xA(x)$ is a one-board game, $\wedge xA(x)$ is an infinitely-many-board game: playing it means playing, in parallel, game $A(1)$ on board #1, game $A(2)$ on board #2, etc. When restricted to elementary games, however, the distinction between the blind and the parallel groups of quantifiers disappears as already noted and, just like $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\bigwedge$, $\bigvee$, the blind quantifiers behave exactly in the classical way. Having this collection of operators makes CL a conservative extension of classical first-order logic: the latter is nothing but CL restricted to elementary problems and the logical vocabulary $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\forall$ (and/or $\bigwedge$), $\exists$ (and/or $\bigvee$).

11.4.6 Recurrence operations

What is common to the members of the family of game operations called **recurrence operations** is that, when applied to A, they turn it into a game playing which means repeatedly playing A. In terms of resources, recurrence operations generate multiple "copies" of A, thus making A a reusable/recyclable resource. The difference between various sorts of recurrences is how "reusage" is exactly understood. Imagine a computer that has a program successfully playing *Chess*. The resource that such a computer provides is obviously something stronger than just *Chess*, for it permits to play *Chess* as many times as the user

wishes, while *Chess*, as such, only assumes one play. The simplest operating system would allow to start a session of *Chess*, then—after finishing or abandoning and destroying it—start a new play again, and so on. The game that such a system plays—i.e. the resource that it supports/provides—is $⍋Chess$, which assumes an unbounded number of plays of *Chess* in a sequential fashion. We call $⍋$ **sequential recurrence**. A more advanced operating system, however, would not require to destroy the old sessions before starting new ones; rather, it would allow to run as many parallel sessions as the user needs. This is what is captured by $⍲Chess$, meaning nothing but the infinite parallel conjunction $Chess \wedge Chess \wedge Chess \wedge \ldots$. Hence $⍲$ is called **parallel recurrence**. As a resource, $⍲Chess$ is obviously stronger than $⍋Chess$ as it gives the user more flexibility. But $⍲$ is still not the strongest form of reusage. A really good operating system would not only allow the user to start new sessions of *Chess* without destroying old ones; it would also make it possible to branch/replicate each particular stage of each particular session, i.e. create any number of "copies" of any already reached position of the multiple parallel plays of *Chess*, thus giving the user the possibility to try different continuations from the same position. What corresponds to this intuition is $⍛Chess$, where $⍛$ is called **branching recurrence**.[5] As all of the operations (except $\neg$, $\rightarrow$) seen in this section, each sort of recurrence comes with its dual operation, called **corecurrence**. Say, the **branching corecurrence** $⍰A$ of A is nothing but $⍛\neg A$ as seen by the environment, so $⍰A$ can be defined as $\neg⍛\neg A$; similarly for **parallel corecurrence** $⍱$ and **sequential corecurrence** $⍒$. $⍱A$ thus means the infinite parallel disjunction $A \vee A \vee A \vee \ldots$. The sequential recurrence and sequential corecurrence of A, on the other hand, can be defined as infinite **sequential conjunction** $A \triangle A \triangle A \triangle \ldots$ and infinite **sequential disjunction** $A \triangledown A \triangledown A \triangledown \ldots$, respectively. An idea of the sequential versions of conjunction/disjunction, quantifiers and recurrence/corecurrence was informally outlined in a footnote of Section 8 of Japaridze (2006e), and then fully elaborated in Japaridze (2008b). Out of laziness, in this paper we are not going to go any farther than the above intuitive explanation of sequential recurrence, just as we have not attempted and will not attempt to define the sequential versions of propositional connectives or quantifiers.[6] Only the parallel and branching sorts of recurrence will receive our full attention.

Definition 12.
1. **Parallel recurrence** $⍲A$*:*

- $\Gamma \in \mathbf{Lr}^{⍲}$ *iff every move of* Γ *has the prefix 'i.' for some* $i \in \{1, 2, 3, \ldots\}$ *and, for each such i,* $\Gamma^{i.} \in \mathbf{Lr}^{A}$.

[5]The term "branching recurrence" and the symbol $⍛$ were established in Japaridze (2006e). The earlier (foundational) paper (Japaridze, 2003) uses "branching conjunction" and ! instead. Similarly, Japaridze (2003) uses the term "branching disjunction" instead of our present "branching corecurrence", and symbol ? instead of $⍰$. Finally, to our present symbol $\circ\!\!-$ in Japaridze (2003) corresponds $\Rightarrow$.

[6]There are similar and even more serious reasons for not attempting to introduce blind versions of conjunction and disjunction.

- *Whenever* $\Gamma \in \mathbf{Lr}^{\curlywedge A}$*, we have* $\mathbf{Wn}^{\curlywedge A}\langle\Gamma\rangle = \top$ *iff, for each* $i \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{A}\langle\Gamma^{i.}\rangle = \top$.

2. **Parallel corecurrence** $\curlyvee A$*:*

- $\Gamma \in \mathbf{Lr}^{\curlyvee A}$ *iff every move of* Γ *has the prefix 'i.' for some* $i \in \{1, 2, 3, \ldots\}$ *and, for each such* i, $\Gamma^{i.} \in \mathbf{Lr}^{A}$.

- *Whenever* $\Gamma \in \mathbf{Lr}^{\curlyvee A}$*, we have* $\mathbf{Wn}^{\curlyvee A}\langle\Gamma\rangle = \bot$ *iff, for each* $i \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}^{A}\langle\Gamma^{i.}\rangle = \bot$.

Thus, from the machine's perspective, $\curlywedge\sqcap x\sqcup y(y = x^2)$ is the problem of computing the square function, and—unlike the case with $\sqcap x\sqcup y(y = x^2)$—doing so repeatedly, i.e. as many times as the environment asks a question "what is the square of m?". In the style of Example 10, a unilegal position of $\curlywedge A$ (resp. $\curlyvee A$) can be represented as an infinite parallel conjunction (resp. disjunction), with the infinite contiguous block of "not-yet-activated" conjuncts (resp. disjuncts) starting from item #n combined together and written as $\curlywedge nA$ (resp. $\curlyvee nA$). Below is an illustration.

Example 13. The $\top$-won run $\langle\bot 1.3, \top 1.9, \bot 2.1, \top 2.1\rangle$ of the game $\curlywedge\sqcap x\sqcup y(y = x^2)$ generates the following sequence:

$\curlywedge\sqcap x\sqcup y(y = x^2)$ (or $\curlywedge 1\sqcap x\sqcup y(y = x^2)$);
$\sqcup y(y = 3^2) \wedge \curlywedge 2\sqcap x\sqcup y(y = x^2)$;
$9 = 3^2 \wedge \curlywedge 2\sqcap x\sqcup y(y = x^2)$;
$9 = 3^2 \wedge \sqcup y(y = 1^2) \wedge \curlywedge 3\sqcap x\sqcup y(y = x^2)$;
$9 = 3^2 \wedge 1 = 1^2 \wedge \curlywedge 3\sqcap x\sqcup y(y = x^2)$.

Among the best known and most natural concepts of reducibility in traditional computability theory is that of **Turing reducibility** of a problem A to a problem B, meaning existence of a Turing machine that solves A with an oracle for B. In this definition "problem", of course, is understood in the traditional sense, meaning a two-step, question-answer problem such as computing a function or deciding a predicate. This is so because both the oracle and the Turing machine offer a simple, question-answer interface, unable to handle problems with higher degrees or more sophisticated forms of interactivity. Our approach allows us to get rid of the "amateurish" concept of an oracle, and reformulate the above definition of Turing reducibility of A to B as computability of $B \succ\!- A$, where $\succ\!-$ is defined by

$$B \succ\!- A \;=\; (\curlywedge B) \to A.$$

This newborn concept of $\succ$**-reducibility** then not only adequately rephrases Turing reducibility, but also generalizes it, for $\succ$-reducibility is defined for all games A and B rather than only those representing traditional computational problems.

To get a better feel of $\succ$ and appreciate the difference between it and the ordinary reduction $\rightarrow$, remember our example of $\rightarrow$-reducing the acceptance problem to the halting problem. The reduction that $\top$ used in its successful strategy for

$$\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y)) \rightarrow \sqcap x \sqcap y(A(x,y) \sqcup \neg A(x,y))$$

was in fact a Turing reduction, as $\top$'s moves $1.m$ and $1.n$ can be seen as querying an oracle (with the environment acting in the role of such) regarding whether m halts on n. The potential usage of an "oracle", however, was limited there, as it could be employed only once. If, for some reason, $\top$ needed to repeat the same question with some different parameters m' and n', it would not be able to do so, for this would require having two "copies" of the resource $\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y))$ in the antecedent, i.e. having

$$\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y)) \wedge \sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y))$$

rather than $\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y))$ there. On the other hand, Turing reduction assumes an unlimited oracle-querying capability. Such a capability is precisely accounted for by prefixing the antecedent with a $!$, i.e. replacing $\rightarrow$ with $\succ$. As an example of a problem $\succ$-reducible but not $\rightarrow$-reducible to the halting problem, consider the relative Kolmogorov complexity problem. It can be expressed as $\sqcap x \sqcap y \sqcup z K(x,y,z)$, where $K(x,y,z)$ is the predicate "z is the Kolmogorov complexity of x relative to y", i.e. "z is the smallest (code of a) Turing machine that returns x on input y". The problem of Turing-reducing the relative Kolmogorov complexity problem to the halting problem translates into our terms as

$$!\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y)) \rightarrow \sqcap x \sqcap y \sqcup z K(x,y,z).$$

Seeing the antecedent as an infinite $\wedge$-conjunction, here is $\top$'s algorithmic winning strategy for the above game. $\top$ waits till $\bot$ selects some values m and n for x and y in the consequent, signifying asking $\top$ about the Kolmogorov complexity of m relative to n. Then, starting from $i = 1$, $\top$ does the following. In the ith $\wedge$-conjunct of the antecedent, it makes two consecutive moves by specifying x and y as i and n, respectively, thus asking $\bot$ whether machine i halts on input n. If $\bot$ responds there by "no", $\top$ increments i by one and repeats the step. Otherwise, if $\bot$ responds by "yes", $\top$ simulates machine i on input n until it halts; if the simulation shows that machine i on input n returns m, $\top$ makes

the move i in the consequent, thus saying that i is the Kolmogorov complexity of m relative to n; otherwise, $\top$ increments i by one and repeats the step.

Turing reducibility has well-justified claims to be a formalization of our weakest intuition of algorithmic reducibility of one traditional problem to another, and $\succ\!\!-$-reducibility, as we now know, conservatively extends Turing reducibility to all games. This may suggest that $\succ\!\!-$-reducibility could be an adequate formal counterpart of our weakest intuition of algorithmic reducibility of one interactive problem to another. Such a guess would be wrong though. As claimed earlier, it is $\mathbin{⫰}A$ rather than $\mathbin{⫯}A$ that corresponds to our strongest intuition of using/reusing A. This automatically translates into another claim: it is $\circ\!\!-$-reducibility rather than $\succ\!\!-$-reducibility that (in full interactive generality) captures the weakest form of reducibility. Here $\circ\!\!-$ is defined by

$$B \circ\!\!- A \;=\; (\mathbin{⫰}B) \to A.^{7}$$

It was mentioned in Section 11.1 that Heyting's intuitionistic calculus is sound and, in the propositional case, also complete with respect to the semantics of computability logic. This is so when intuitionistic implication is understood as $\circ\!\!-$, and intuitionistic conjunction, disjunction, universal quantifier and existential quantifier as $\sqcap, \sqcup, \bigsqcap$ and $\bigsqcup$, respectively. With intuitionistic implication read as $\succ\!\!-$, intuitionistic calculus is unsound as, for example, it proves

$$(P \succ\!\!- R) \sqcap (Q \succ\!\!- R) \succ\!\!- (P \sqcup Q \succ\!\!- R)$$

which fails to be a valid principle of computability.

$\succ\!\!-$-reducibility and $\circ\!\!-$-reducibility, while being substantially different in the general case, turn out to be equivalent when restricted to certain special sorts of problems with low degrees of interactivity such as what we have been referring to as "traditional problems", examples of which being the halting, acceptance or relative Kolmogorov complexity problems. For this reason, both $\succ\!\!-$-reducibility and $\circ\!\!-$-reducibility are equally adequate as (conservative) generalizations of the traditional concept of Turing reducibility.

It is now time to get down to a formal definition of branching recurrence $⫰$. This is not just as easy as defining $⫯$, and requires a number of auxiliary concepts and conventions. Let us start with a closer look at the associated intuitions. One of the ways to view both $\mathbin{⫯}A$ and $\mathbin{⫰}A$ is to think of them as games where $\bot$ is allowed to restart A an unlimited number of times without terminating the already-in-progress runs of A, creating, this way, more and more parallel plays of A with the possibility to try different strategies in them and become the winner as long as one of those strategies succeeds. What makes $\mathbin{⫰}A$

[7] Now we may guess that, if it ever comes to studying the sequential-recurrence-based reduction $(\mathbin{⩲}B) \to A$, the symbol for it would be $\mapsto\!\!-$.

stronger (as a resource) than $⅄A$, however, is that, as noted earlier, in $⅁A$, $\bot$ does not have to really restart A from the very beginning every time it "restarts" it; rather, it can select to continue A from any of the previous positions, thus depriving $\top$ of the possibility to reconsider the moves it has already made. A little more precisely, at any time $\bot$ is allowed to replicate (backup) any of the currently reached parallel positions of A before further modifying it, which gives it the possibility to come back later and continue playing A from the backed-up position. This way, we get a tree of labmoves (each branch spelling a legal run of A) rather than just multiple parallel sequences of labmoves. Then $⅄A$ can be thought of as a weak version of $⅁A$ where only the empty position of A can be replicated, that is, where branching in the tree only happens at its root. A discussion of how $⅁$ relates to Blass's (1972, 1992) repetition operator is given in Section 13 of Japaridze (2003).

To visualize the scheme that lies under our definition of $⅁$, consider a play over $⅁Chess$. The play takes place between a computer ($\top$) and a user ($\bot$), and its positions are displayed on the screen. In accordance with the earlier elaborated intuitions, we think of each such position Φ as the game $\langle\Phi\rangle Chess$, and vice versa. At the beginning, there is a window on the screen—call it Window ϵ—that displays the initial position of *Chess*:

Window ϵ
| *Chess* |

We denote this position by *Chess*, but the designer would probably make the window show a colorful image of a chess board with 32 chess pieces in their initial locations. The play starts and proceeds in an ordinary fashion: the players are making legal moves of *Chess*, which correspondingly update the position displayed in the window. At some (any) point, when the current position in the window is $\langle\Phi\rangle Chess$, $\bot$ may decide to replicate the position, perhaps because he wants to try different continuations in different copies of it. This splits Window ϵ into two children windows named 0 and 1, each containing the same position $\langle\Phi\rangle Chess$ as the mother window contained at the time of split. The mother window disappears, and the picture on the screen becomes

Window 0
| $\langle\Phi\rangle Chess$ |

Window 1
| $\langle\Phi\rangle Chess$ |

(again, let us try to imagine a real chess position colorfully depicted in the two windows instead of the bleak expression "$\langle\Phi\rangle Chess$").

From now on the play continues on two boards/in two windows. Either player can make a legal move in either window. After some time, when

the game in Window 0 has evolved to $\langle\Phi,\Psi\rangle$*Chess* and in Window 1 to $\langle\Phi,\Theta\rangle$*Chess*, $\bot$ can, again, decide to split one of the windows—say, Window 0. The mother window 0 will be replaced by two children windows: 00 and 01, each having the same content as their mother had at the moment of split, so that now the screen will be showing three windows:

Window 00	Window 01	Window 1
$\langle\Phi,\Psi\rangle$*Chess*	$\langle\Phi,\Psi\rangle$*Chess*	$\langle\Phi,\Theta\rangle$*Chess*

If and when, at some later moment, $\bot$ decides to make a third split—say, in Window 01—the picture on the screen becomes

Window 00	Window 010	Window 011	Window 1
$\langle\Phi,\Psi,\Lambda\rangle$*Chess*	$\langle\Phi,\Psi,\Sigma\rangle$*Chess*	$\langle\Phi,\Psi,\Sigma\rangle$*Chess*	$\langle\Phi,\Theta,\Pi\rangle$*Chess*

etc. At the end, the game will be won by $\top$ if and only if each of the windows contains a winning position of *Chess*.

The above four-window position can also be represented as the following binary tree, where the name of each window is uniquely determined by its location in the tree:

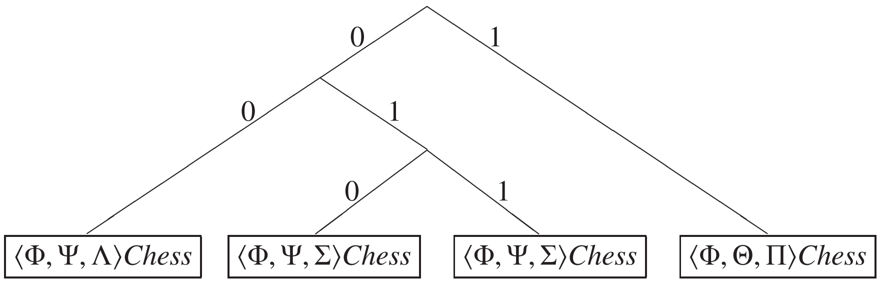


Figure 11.11: A position of ⫰*Chess*

Window names will be used by the players to indicate in which of the windows they are making a move. Specifically, '$w.\alpha$' is the move meaning making move α in Window w; and the move (by $\bot$) that splits/replicates Window w is 'w:'. Sometimes the window in which a player is trying to make a move may no longer exist. For example, in the position preceding the position of Figure 11.11, $\top$ might have decided to make move α in Window 01. However, before $\top$ actually made this move, $\bot$ made a replicative move in the same window, which took us to the four-window position of Figure 11.11. $\top$ may not notice this replicative move and complete its move $01.\alpha$ by the time when

Window 01 no longer exists. This kind of a move is still considered legal, and its effect is making the move α in all (in our case both) of the descendants of the no-longer-existing Window 01. The result will be

Window 00	Window 010	Window 011	Window 1
$\langle \Phi, \Psi, \Lambda \rangle$*Chess*	$\langle \Phi, \Psi, \Sigma, \top\alpha \rangle$*Chess*	$\langle \Phi, \Psi, \Sigma, \top\alpha \rangle$*Chess*	$\langle \Phi, \Theta, \Pi \rangle$*Chess*

The initial position in the example that we have just discussed was one-window. This, generally, is not necessary. The operation ⫰ can be applied to any construction in the above style, such as, say,

Window 0	Window 1
Chess	*Checkers*

A play over this game, which our later-introduced notational conventions would denote by ⫰(*Chess* ∘ *Checkers*), will proceed in a way similar to what we saw, where more and more windows can be created, some (those whose names are 0-prefixed) displaying positions of chess, and some (with 1-prefixed names) displaying positions of checkers. In order to win, the machine will have to win all of the multiple parallel plays of chess and checkers that will be generated.

As the dual of ⫰, ⫯ can be characterized in exactly the same way as ⫰, only, in a ⫯-game, it is $\top$ who has the privilege of splitting windows, and for whom winning in just one of the multiple parallel plays is sufficient.

To put together the above intuitions, let us agree that by a **bitstring** we mean a string of 0s and 1s, including infinite strings and the **empty string** ϵ. We will be using the letters w, u, v as metavariables for bitstrings. ϵ will exclusively stand for the empty bitstring. The expression uw, meaningful only when u is finite, will stand for the concatenation of strings u and w. We write $u \preceq w$ to mean that u is an initial segment of w. And $u \prec w$ means that u is a proper initial segment of w, i.e. that $u \preceq w$ and $u \neq w$.

Definition 14.

1. *A* **bitstring tree** *(***BT***) is a nonempty set T of bitstrings, called the* **branches** *of the tree (with finite branches also called* **nodes***), such that, for all bitstrings w, u, the following three conditions*[8] *are satisfied:*

   (a) *If $w \in T$ and $u \prec w$, then $u \in T$.*
   (b) *$w0 \in T$ iff $w1 \in T$ (finite w).*
   (c) *If w is infinite and all u with $u \prec w$ are in T, then so is w.*

[8] Due to a mechanical error, the third condition was lost in the published version of Japaridze (2003).

2. *A* **complete branch** *of a BT T is a branch w of T such that for no bitstring u with $w \prec u$ do we have $u \in T$. A finite complete branch of T is also said to be a* **leaf** *of T. Notice that T (as a set of strings) is finite iff all of its branches (as strings) are so. Hence, the terms "complete branch" and "leaf" are synonymic for finite BTs, as are "branch" and "node".*

3. *A* **decoration for** *a finite BT T is a function d that sends each leaf of T to some game.*

4. *A* **decorated bitstring tree** *(***DBT***) $\mathcal{T}$ is a pair (T, d), where T—called the* **BT-structure** *of $\mathcal{T}$—is a finite BT, and d— called the* **decoration of** *$\mathcal{T}$—is a decoration for T. Such a $\mathcal{T}$ is said to be a* **singleton** *iff $T = \{\epsilon\}$. We identify each singleton DBT $(\{\epsilon\}, d)$ with the game $d(\epsilon)$, and vice versa: think of each game A as the singleton DBT $(\{\epsilon\}, d)$ with $d(\epsilon) = A$. In some contexts, on the other hand, we may identify a DBT $\mathcal{T}$ with its treestructure T, and say "branch (leaf, etc.) of $\mathcal{T}$" to mean "branch (leaf, etc.) of T".*

In Figure 11.11 we see an example of a DBT whose BT-structure is $\{\epsilon, 0, 1, 00, 01, 010, 011\}$ and whose decoration is the function d given by $d(00) = \langle \Phi, \Psi, \Lambda \rangle Chess$, $d(010) = d(011) = \langle \Phi, \Psi, \Sigma \rangle Chess$, $d(1) = \langle \Phi, \Theta, \Pi \rangle Chess$.

Drawing actual trees for DBTs is not very convenient, and an alternative way to represent a DBT $\mathcal{T} = (T, d)$ is the following:

- If $\mathcal{T}$ is a singleton with $d(\epsilon) = A$, then $\mathcal{T}$ is simply written as A.
- Otherwise, $\mathcal{T}$ is written as $E_0 \circ E_1$, where E_0 and E_1 represent the sub-DBTs of $\mathcal{T}$ rooted at 0 and 1, respectively.

For example, the DBT of Figure 11.11 will be written as

$$\Big(\big(\langle \Phi, \Psi, \Lambda \rangle Chess\big) \circ \big(\big(\langle \Phi, \Psi, \Sigma \rangle Chess\big) \circ \big(\langle \Phi, \Psi, \Sigma \rangle Chess\big)\big)\Big) \circ \Big(\langle \Phi, \Theta, \Pi \rangle Chess\Big).$$

We are going to define $\natural$ and $\flat$ as operations applicable not only to games, i.e. singleton DBTs, but to any DBTs as well.

Definition 15. *Let $\mathcal{T} = (T, d)$ be DBT. We define the notion of a* **prelegal position** *of $\natural\mathcal{T}$ (resp. $\flat\mathcal{T}$), together with the function $Tree^{\natural\mathcal{T}}$ (resp. $Tree^{\flat\mathcal{T}}$) that associates a BT $Tree^{\natural\mathcal{T}}\langle\Phi\rangle$ (resp. $Tree^{\flat\mathcal{T}}\langle\Phi\rangle$) with each such position Φ, by the following induction:*

(a) $\langle\rangle$ *is a prelegal position of $\natural\mathcal{T}$ (resp. $\flat\mathcal{T}$), and*

$$Tree^{\natural\mathcal{T}}\langle\rangle = T$$
$$(resp.\quad Tree^{\flat\mathcal{T}}\langle\rangle = T).$$

(b) $\langle \Phi, \lambda \rangle$ *is a prelegal position of $\natural\mathcal{T}$ (resp. $\flat\mathcal{T}$) iff Φ is so and one of the following two conditions is satisfied:*

1. $\lambda = \bot w$*: (resp.* $\lambda = \top w$*:) for some leaf* w *of* $Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle$ *(resp.* $Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle$*). We call this sort of a labmove* λ *or move* w*:* **replicative**. *In this case*

$$Tree^{\downarrow\mathcal{T}}\langle\Phi, \bot w:\rangle = Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle \cup \{w0, w1\}$$
$$(resp. \quad Tree^{\uparrow\mathcal{T}}\langle\Phi, \top w:\rangle = Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle \cup \{w0, w1\}).$$

2. $\lambda = \wp w.\alpha$ *for some node* w *of* $Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle$ *(resp.* $Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle$*), player* $\wp$ *and move* α*. We call this sort of a labmove* λ *or move* $w.\alpha$ **nonreplicative**. *In this case*

$$Tree^{\downarrow\mathcal{T}}\langle\Phi, \wp w.\alpha\rangle = Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle$$
$$(resp. \quad Tree^{\uparrow\mathcal{T}}\langle\Phi, \wp w.\alpha\rangle = Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle).$$

As mentioned earlier, with a visualization in the style of Figure 11.11 in mind, the meaning of a replicative labmove $\wp w$: is that player $\wp$ splits leaf/window w into two children windows $w0$ and $w1$; and the meaning of a nonreplicative labmove $\wp w.\alpha$ is that $\wp$ makes the move α in all windows whose names start with w. Prelegality is a minimum condition that every legal run of a $\downarrow$- or $\uparrow$-game should satisfy. In particular, prelegality means that new windows have only been created by the "right player" (i.e. $\bot$ in a $\downarrow$-game, and $\top$ in a $\uparrow$-game), and that no moves have been made in not-yet-created windows. As for $Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle$ or $Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle$, it shows to what we will be referring as the **underlying BT-structure** of the position to which Φ brings the game down. Note that, as can be seen from the definition, whether Φ is a prelegal position of $\downarrow\mathcal{T}$ or $\uparrow\mathcal{T}$ and what the value of $Tree^{\downarrow\mathcal{T}}\langle\Phi\rangle$ or $Tree^{\uparrow\mathcal{T}}\langle\Phi\rangle$ is, only depends on the BT-structure of $\mathcal{T}$ and not on its decoration.

The concept of a prelegal position of $\downarrow\mathcal{T}$ can be generalized to all runs by stipulating that a **prelegal run of** $\downarrow\mathcal{T}$ is a run whose every finite initial segment is a prelegal position of $\downarrow\mathcal{T}$. Similarly, the function $Tree^{\downarrow\mathcal{T}}$ can be extended to all prelegal runs of $\downarrow\mathcal{T}$ by stipulating that $Tree^{\downarrow\mathcal{T}}\langle\lambda_1, \lambda_2, \lambda_3, \ldots\rangle$ is the smallest BT containing all elements of the set

$$Tree^{\downarrow\mathcal{T}}\langle\rangle \cup Tree^{\downarrow\mathcal{T}}\langle\lambda_1\rangle \cup Tree^{\downarrow\mathcal{T}}\langle\lambda_1, \lambda_2\rangle \cup Tree^{\downarrow\mathcal{T}}\langle\lambda_1, \lambda_2, \lambda_3\rangle \cup \ldots \qquad (5)$$

of bitstrings. In other words, $Tree^{\downarrow\mathcal{T}}\langle\lambda_1, \lambda_2, \lambda_3, \ldots\rangle$ is the result of adding to (5) every infinite bitstring w such that all finite initial segments of w are in (5). The concept of a prelegal position of $\uparrow\mathcal{T}$ and the function $Tree^{\uparrow\mathcal{T}}$ generalize to infinite runs in a similar way.

We now introduce an important notational convention that should be remembered. Let u be a bitstring and Γ any run. Then

$$\Gamma^{\preceq u}$$

will stand for the result of first removing from Γ all labmoves except those that look like $\wp w.\alpha$ for some bitstring w with $w \preceq u$, and then deleting this

sort of prefixes 'w.' in the remaining labmoves, i.e. replacing each remaining labmove $\wp w.\alpha$ (where w is a bitstring) by $\wp\alpha$. Example: If $u = 101000\ldots$ and $\Gamma = \langle \top\epsilon.\alpha_1, \bot\epsilon{:}, \bot 1.\alpha_2, \top 0.\alpha_3, \bot 1{:}, \top 10.\alpha_4\rangle$, then $\Gamma^{\preceq u} = \langle \top\alpha_1, \bot\alpha_2, \top\alpha_4\rangle$.

Being a prelegal run of $\mathfrak{b}\mathcal{T}$ is a necessary but not a sufficient condition for being a legal run of this game. For simplicity, let us consider the case when $\mathcal{T}$ is singleton DBT A, where A is a constant game. It was noted earlier that a legal run Γ of $\mathfrak{b}A$ can be thought of as consisting of multiple legal runs of A. In particular, these runs will be the runs $\Gamma^{\preceq u}$, where u is a complete branch of $\mathit{Tree}^{\mathfrak{b}A}\langle\Gamma\rangle$. The labmoves of $\Gamma^{\preceq u}$ for such a u are those $\wp\alpha$ that emerged as a result of making (nonreplicative) labmoves of the form $\wp w.\alpha$ with $w \preceq u$. For example, to branch 010 in Figure 11.11 corresponds run $\langle\Phi, \Psi, \Sigma\rangle$, where the labmoves of Φ originate from the nonreplicative labmoves of the form $\wp\epsilon.\alpha$ (i.e. $\wp.\alpha$) made before the first replicative move, the labmoves of Ψ originate from the nonreplicative labmoves of the form $\wp w.\alpha$ with $w \preceq 0$ made between the first and the second replicative moves, and the labmoves of Σ originate from the nonreplicative labmoves of the form $\wp w.\alpha$ with $w \preceq 01$ made between the second and the third replicative moves. Generally, for a prelegal run Γ of $\mathfrak{b}A$ to be a legal run, it is necessary and sufficient that all of the runs $\Gamma^{\preceq u}$, where u is a complete branch of $\mathit{Tree}^{\mathfrak{b}A}\langle\Gamma\rangle$, be legal runs of A. And for such a Γ to be a $\top$-won run, it is necessary and sufficient that all of those $\Gamma^{\preceq u}$ be $\top$-won runs of A.

When $\mathcal{T}$ is a non-singleton DBT, the situation is similar. For example, for Γ to be a legal (resp. $\top$-won) run of $\mathfrak{b}(\mathit{Chess} \circ \mathit{Checkers})$, along with being a prelegal run, it is necessary that, for every complete branch $0u$ of $\mathit{Tree}^{\mathfrak{b}(\mathit{Chess}\circ\mathit{Checkers})}\langle\Gamma\rangle$, $\Gamma^{\preceq 0u}$ be a legal (resp. $\top$-won) run of *Chess* and, for every complete branch $1u$ of the same tree, $\Gamma^{\preceq 1u}$ be a legal (resp. $\top$-won) run of *Checkers*.

Finally, the case with $\mathfrak{p}\mathcal{T}$, of course, is symmetric to that with $\mathfrak{b}\mathcal{T}$.

All of the above intuitions are summarized in the following formal definitions of $\mathfrak{b}$ and $\mathfrak{p}$, with Definition 16 being for the simpler case when $\mathcal{T}$ is a singleton, and Definition 17 generalizing it to all DBTs. In concordance with the earlier remark that considering constant games is sufficient when defining propositional connectives, Definition 16 assumes that A is a constant game, and Definition 17 assumes that $\mathcal{T}$ is a **constant DBT**, meaning a DBT whose decoration sends every leaf of its BT-structure to a constant game.

Definition 16. *Assume A is a constant game.*

1. **Branching recurrence** $\mathfrak{b}A$*:*

- $\Gamma \in \mathbf{Lr}^{\mathfrak{b}A}$ *iff* Γ *is a prelegal run of* $\mathfrak{b}A$*, and* $\Gamma^{\preceq u} \in \mathbf{Lr}^{A}$ *for every complete branch* u *of* $\mathit{Tree}^{\mathfrak{b}A}\langle\Gamma\rangle$*.*
- *Whenever* $\Gamma \in \mathbf{Lr}^{\mathfrak{b}A}$*,* $\mathbf{Wn}^{\mathfrak{b}A}\langle\Gamma\rangle = \top$ *iff* $\mathbf{Wn}^{A}\langle\Gamma^{\preceq u}\rangle = \top$ *for every complete branch* u *of* $\mathit{Tree}^{\mathfrak{b}A}\langle\Gamma\rangle$*.*

2. **Branching corecurrence** $\mathbin{⫯}A$*:*

- $\Gamma \in \mathbf{Lr}^{\mathbin{⫯}A}$ *iff* Γ *is a prelegal run of* $\mathbin{⫯}A$*, and* $\Gamma^{\leq u} \in \mathbf{Lr}^{A}$ *for every complete branch* u *of* $Tree^{\mathbin{⫯}A}\langle\Gamma\rangle$*.*
- *Whenever* $\Gamma \in \mathbf{Lr}^{\mathbin{⫯}A}$*,* $\mathbf{Wn}^{\mathbin{⫯}A}\langle\Gamma\rangle = \bot$ *iff* $\mathbf{Wn}^{A}\langle\Gamma^{\leq u}\rangle = \bot$ *for every complete branch* u *of* $Tree^{\mathbin{⫯}A}\langle\Gamma\rangle$*.*

Definition 17. *Assume* $\mathcal{T} = (T, d)$ *is a constant DBT.*

1. ***Branching recurrence*** $\mathbin{⫰}\mathcal{T}$*:*

- $\Gamma \in \mathbf{Lr}^{\mathbin{⫰}\mathcal{T}}$ *iff* Γ *is a prelegal run of* $\mathbin{⫰}\mathcal{T}$*, and* $\Gamma^{\leq wu} \in \mathbf{Lr}^{d(w)}$ *for every complete branch* wu *of* $Tree^{\mathbin{⫰}\mathcal{T}}\langle\Gamma\rangle$ *where* w *is a leaf of* T*.*
- *Whenever* $\Gamma \in \mathbf{Lr}^{\mathbin{⫰}\mathcal{T}}$*,* $\mathbf{Wn}^{\mathbin{⫰}\mathcal{T}}\langle\Gamma\rangle = \top$ *iff* $\mathbf{Wn}^{d(w)}\langle\Gamma^{\leq wu}\rangle = \top$ *for every complete branch* wu *of* $Tree^{\mathbin{⫰}\mathcal{T}}\langle\Gamma\rangle$ *where* w *is a leaf of* T*.*

2. ***Branching corecurrence*** $\mathbin{⫯}\mathcal{T}$*:*

- $\Gamma \in \mathbf{Lr}^{\mathbin{⫯}\mathcal{T}}$ *iff* Γ *is a prelegal run of* $\mathbin{⫯}\mathcal{T}$*, and* $\Gamma^{\leq wu} \in \mathbf{Lr}^{d(w)}$ *for every complete branch* wu *of* $Tree^{\mathbin{⫯}\mathcal{T}}\langle\Gamma\rangle$ *where* w *is a leaf of* T*.*
- *Whenever* $\Gamma \in \mathbf{Lr}^{\mathbin{⫯}\mathcal{T}}$*,* $\mathbf{Wn}^{\mathbin{⫯}\mathcal{T}}\langle\Gamma\rangle = \bot$ *iff* $\mathbf{Wn}^{d(w)}\langle\Gamma^{\leq wu}\rangle = \bot$ *for every complete branch* wu *of* $Tree^{\mathbin{⫯}\mathcal{T}}\langle\Gamma\rangle$ *where* w *is a leaf of* T*.*

Let us not forget to make our already routine observation that the definition of either operation $\mathbin{⫰}, \mathbin{⫯}$ can be obtained from the definition of its dual by just interchanging $\top$ with $\bot$.

Now it would be interesting to see how the moves of unilegal runs affect $\mathbin{⫰}$- and $\mathbin{⫯}$-games. In fact, being able to describe the effect of such moves was our main motivation for defining $\mathbin{⫰}$ and $\mathbin{⫯}$ in the general form as operations on DBTs. We need two preliminary definitions here.

Definition 18. *Suppose* $\mathcal{T} = (T, d)$ *is a DBT and* w *is a leaf of* T*. We define* $Rep_w[\mathcal{T}]$ *as the following DBT* (T', d')*:*

1. $T' = T \cup \{w0, w1\}$.
2. d' *is the decoration for* T' *such that:*
(a) $d'(w0) = d'(w1) = d(w)$*;*
(b) *For every other* $(\neq w0, w1)$ *leaf* u *of* T'*,* $d'(u) = d(u)$.

Examples:

1. $Rep_0[A \circ (B \circ C)] = (A \circ A) \circ (B \circ C)$;
2. $Rep_{10}[A \circ (B \circ C)] = A \circ ((B \circ B) \circ C)$;
3. $Rep_{11}[A \circ (B \circ C)] = A \circ (B \circ (C \circ C))$.

Definition 19. *Suppose $\mathcal{T} = (T, d)$ is a DBT, w is a node of T and, for every leaf u of T with $w \preceq u$, $\langle\lambda\rangle$ is a unilegal position of $d(u)$. We define $Nonrep^{\lambda}_{w}[\mathcal{T}]$ as the DBT (T, d'), where d' is the decoration for T such that:*

(a) *For every leaf u of T with $w \preceq u$, $d'(u) = \langle\lambda\rangle d(u)$;*
(b) *For every other leaf u of T, $d'(u) = d(u)$.*

Examples (assuming the appropriate unilegality conditions on $\langle\lambda\rangle$):

1. $Nonrep^{\lambda}_{10}[A \circ (B \circ C)] = A \circ (\langle\lambda\rangle B \circ C)$;
2. $Nonrep^{\lambda}_{1}[A \circ (B \circ C)] = A \circ (\langle\lambda\rangle B \circ \langle\lambda\rangle C)$;
3. $Nonrep^{\lambda}_{\epsilon}[A \circ (B \circ C)] = \langle\lambda\rangle A \circ (\langle\lambda\rangle B \circ \langle\lambda\rangle C)$.

The following theorem is a combination of Propositions 13.5 and 13.8 proven in Japaridze (2003):

Theorem 20. *Suppose $\mathcal{T} = (T, d)$ is a DBT, and λ any labmove.*

1. $\langle\lambda\rangle \in \mathbf{LR}^{\natural\mathcal{T}}$ *iff one of the following conditions holds:*

(a) *λ is (the replicative labmove) $\bot w:$, where w is a leaf of T. In this case $\langle\bot w:\rangle\natural\mathcal{T} = \natural Rep_{w}[\mathcal{T}]$.*

(b) *λ is (the nonreplicative labmove) $\wp w.\alpha$, where $\wp$ is either player, w is a node of T and, for every leaf u of T with $w \preceq u$, $\langle\wp\alpha\rangle$ is a unilegal position of $d(u)$. In this case $\langle\wp w.\alpha\rangle\natural\mathcal{T} = \natural Nonrep^{\wp\alpha}_{w}[\mathcal{T}]$.*

2. $\langle\lambda\rangle \in \mathbf{LR}^{\wr\mathcal{T}}$ *iff one of the following conditions holds:*

(a) *λ is (the replicative labmove) $\top w:$, where w is a leaf of T. In this case $\langle\top w:\rangle\wr\mathcal{T} = \wr Rep_{w}[\mathcal{T}]$.*

(b) *λ is (the nonreplicative labmove) $\wp w.\alpha$, where $\wp$ is either player, w is a node of T and, for every leaf u of T with $w \preceq u$, $\langle\wp\alpha\rangle$ is a unilegal position of $d(u)$. In this case $\langle\wp w.\alpha\rangle\wr\mathcal{T} = \wr Nonrep^{\wp\alpha}_{w}[\mathcal{T}]$.*

This theorem allows us to easily test whether a given run is a (uni)legal run of a given $\natural$- or $\wr$-game, and if it is, to write out the corresponding sequence of games.

Example 21. Let $\Gamma = \langle\bot\epsilon:, \bot 0.3, \top 0.9, \bot 1:, \bot 10.1, \top 10.1\rangle$, and $A_0 = \natural\sqcap x\sqcup y\,(y = x^2)$. In view of Theorem 20, Γ is legal for A_0, and it brings the latter down to game A_6 as shown below:

A_0: $\natural\Big(\sqcap x\sqcup y(y = x^2)\Big)$;
A_1: $\natural\Big(\sqcap x\sqcup y(y = x^2) \circ \sqcap x\sqcup y(y = x^2)\Big)$, i.e. $\langle\bot\epsilon:\rangle A_0$;
A_2: $\natural\Big(\sqcup y(y = 3^2) \circ \sqcap x\sqcup y(y = x^2)\Big)$, i.e. $\langle\bot 0.3\rangle A_1$;
A_3: $\natural\Big((9 = 3^2) \circ \sqcap x\sqcup y(y = x^2)\Big)$, i.e. $\langle\top 0.9\rangle A_2$;
A_4: $\natural\Big((9 = 3^2) \circ (\sqcap x\sqcup y(y = x^2) \circ \sqcap x\sqcup y(y = x^2))\Big)$, i.e. $\langle\bot 1:\rangle A_3$;
A_5: $\natural\Big((9 = 3^2) \circ (\sqcup y(y = 1^2) \circ \sqcap x\sqcup y(y = x^2))\Big)$, i.e. $\langle\bot 10.1\rangle A_4$;
A_6: $\natural\Big((9 = 3^2) \circ ((1 = 1^2) \circ \sqcap x\sqcup y(y = x^2))\Big)$, i.e. $\langle\top 10.1\rangle A_5$.

The empty run $\langle\rangle$ is a $\top$-won run of each of the three $\circ$-components of A_6. It can be easily seen that then (and only then) $\langle\rangle$ is a $\top$-won run of A_6, for $\mathbin{⅋}(\ldots\circ\ldots)$ essentially acts as parallel conjunction. Hence Γ is a $\top$-won run of A_0.

The run that we see in the above example, though technically different, is still "essentially the same" as the one from Example 13. Indeed, as noted earlier, $⅋$ and $\curlywedge$ are equivalent when applied to traditional, low-interactivity problems such as $\sqcap x\sqcup y(y = x^2)$. What makes the resource $⅋A$ stronger than $\curlywedge A$ is $\bot$'s ability to try several different responses to a same move by $\top$. In $⅋\sqcap x\sqcup y(y = x^2)$, however, $\bot$ cannot take advantage of this flexibility because there are no legal runs of $\sqcap x\sqcup y(y = x^2)$ where $\bot$'s moves follow $\top$'s moves.

To get a feel of the substantial difference between $⅋$ and $\curlywedge$, let us consider, for simplicity, the **bounded versions** $⅋^b$, $⫯^b$, $\curlywedge^b$, $\curlyvee^b$ of our recurrence operations. Here b is a positive integer, setting the bound on the number of parallel plays of game A that can be generated in a legal run of $⅋A$ ($⫯A$, $\curlywedge A$, $\curlyvee A$). That is, $\curlywedge^b A$ and $\curlyvee^b A$ are nothing but the parallel conjunction and parallel disjunction of b copies of A, respectively. And $⅋^b A$ and $⫯^b A$ are defined as $⅋A$ and $⫯A$, with the only difference that, in a legal run Γ, a replicative move can be made at most $b-1$ times, so that $Tree^{⅋^b A}\langle\Gamma\rangle$ or $Tree^{⫯^b A}\langle\Gamma\rangle$ will have at most b complete branches.

We want to compare $⫯^2 D$ with $\curlyvee^2 D$, i.e. with $D \vee D$, where

$$D \;= (Chess \sqcup \neg Chess) \sqcap (Checkers \sqcup \neg Checkers).$$

Winning $D \vee D$ is not easy for $\top$ unless $\top$ is a champion in either chess or checkers. Indeed, a smart environment may choose the left $\sqcap$-conjunct in the left occurrence of D in $D \vee D$ while choose the right $\sqcap$-conjunct in the right occurrence. This will bring the game down to

$$(Chess \sqcup \neg Chess) \vee (Checkers \sqcup \neg Checkers).$$

$\top$ in trouble now. It can, say, make the moves '1.1' and '2.2', bringing the game down to $Chess \vee \neg Checkers$. This will not help much though, as winning $Chess \vee \neg Checkers$, unlike $Chess \vee \neg Chess$, is not any easier than winning either disjunct in isolation.

On the other hand, $\top$ does have a nice winning strategy for $⫯^2 D$. At the beginning, $\top$ waits till $\bot$ chooses one of the two $\sqcap$-conjuncts of D. This brings the game down to, say, $⫯^2(Chess \sqcup \neg Chess)$. Then and only then, $\top$ makes a replicative move, thus creating two copies of $Chess \sqcup \neg Chess$. In one copy $\top$ chooses the left $\sqcup$-disjunct, and in the other copy chooses the right $\sqcup$-disjunct. Now the game will have evolved to $⫯^1(Chess \circ \neg Chess)$. With $⫯^1(A \circ \neg A)$ essentially being nothing but $A \vee \neg A$, mimicking in $Chess$ the moves made by $\bot$ in

$\neg$*Chess* and vice versa guarantees a success for $\top$. Among the runs consistent with this strategy is

$$\langle \bot.1, \top\epsilon{:}, \top 0.1, \top 1.2, \bot 1.\alpha_1, \top 0.\alpha_1, \bot 0.\alpha_2, \top 1.\alpha_2, \bot 1.\alpha_3, \top 0.\alpha_3, \ldots\rangle,$$

to which corresponds the following sequence of games:

$⫯^2((Chess \sqcup \neg Chess) \sqcap (Checkers \sqcup \neg Checkers));$
$⫯^2(Chess \sqcup \neg Chess);$
$⫯^1((Chess \sqcup \neg Chess) \circ (Chess \sqcup \neg Chess));$
$⫯^1(Chess \circ (Chess \sqcup \neg Chess));$
$⫯^1(Chess \circ \neg Chess);$
$⫯^1(Chess \circ \langle \bot\alpha_1\rangle \neg Chess);$
$⫯^1(\langle \top\alpha_1\rangle Chess \circ \langle \bot\alpha_1\rangle \neg Chess);$
$⫯^1(\langle \top\alpha_1, \bot\alpha_2\rangle Chess \circ \langle \bot\alpha_1\rangle \neg Chess);$
$⫯^1(\langle \top\alpha_1, \bot\alpha_2\rangle Chess \circ \langle \bot\alpha_1, \top\alpha_2\rangle \neg Chess);$
$⫯^1(\langle \top\alpha_1, \bot\alpha_2\rangle Chess \circ \langle \bot\alpha_1, \top\alpha_2, \bot\alpha_3\rangle \neg Chess);$
$⫯^1(\langle \top\alpha_1, \bot\alpha_2, \top\alpha_3\rangle Chess \circ \langle \bot\alpha_1, \top\alpha_2, \bot\alpha_3\rangle \neg Chess);$
. . .

As we are going to see later, affine logic is sound with respect to the semantics of computability logic no matter whether the exponential operators !, ? of the former are understood as $\curlywedge, \curlyvee$ or $⫰, ⫯$. Thus, affine logic cannot distinguish between the two groups of recurrence operations. But computability logic certainly sees a difference. As noted earlier, it validates

$$(P \mathbin{\circ\!\!-} R) \sqcap (Q \mathbin{\circ\!\!-} R) \mathbin{\circ\!\!-} (P \sqcup Q \mathbin{\circ\!\!-} R)$$

while makes

$$(P \mathbin{\succ\!\!-} R) \sqcap (Q \mathbin{\succ\!\!-} R) \mathbin{\succ\!\!-} (P \sqcup Q \mathbin{\succ\!\!-} R)$$

fail. Here are two other examples of principles that can be shown to be valid with one sort of recurrence while invalid with the other sort:

$$⫰(P \sqcup Q) \rightarrow ⫰P \sqcup ⫰Q \qquad \textit{is valid};$$
$$\curlywedge(P \sqcup Q) \rightarrow \curlywedge P \sqcup \curlywedge Q \qquad \textit{is not}.$$

$$P \wedge \curlywedge(P \rightarrow Q \wedge P) \rightarrow \curlywedge Q \qquad \textit{is valid};$$
$$P \wedge ⫰(P \rightarrow Q \wedge P) \rightarrow ⫰Q \qquad \textit{is not}.$$

As for how the strengths of $⫰$ and $\curlywedge$ relate, as we may guess, the situation is:

$$⫰P \rightarrow \curlywedge P \qquad \textit{is valid};$$
$$\curlywedge P \rightarrow ⫰P \qquad \textit{is not}.$$

11.5 Static games

Our games are obviously general enough to model anything that one would call a (two-agent, two-outcome) interactive problem. However, they are a bit too general. There are games where the chances of a player to succeed essentially depend on the relative speed at which its adversary acts. A simple example would be the following game:

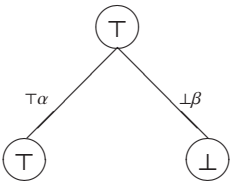


Figure 11.12: A non-static game

One cannot ask which player has a winning strategy here, for this game is a contest of speed rather than intellect: the winner will be whoever is fast enough to move first. CL does not want to consider this sort of games meaningful computational problems, and restricts its attention to the subclass of games that it calls *static*. Intuitively, static games are ones where speed is irrelevant: in order to win, for either player only matters *what* to do (strategy) rather than *how fast* to do (speed). 'These are games where, roughly speaking, it never hurts a player to postpone making moves'.[9]

Static games are defined in terms of the auxiliary concept of $\wp$*-delay*. The notation $\Gamma^{\top}$ used below means the result of deleting from Γ all $\bot$-labeled (lab)moves. Symmetrically for $\Gamma^{\bot}$.

Definition 22. *Let $\wp$ be either player, and Γ, Δ arbitrary runs. We say that Δ is a* **$\wp$-delay** *of Γ iff the following two conditions are satisfied:*

1. $\Delta^{\top} = \Gamma^{\top}$ *and* $\Delta^{\bot} = \Gamma^{\bot}$*;*

2. *For any $k, n \geq 1$, if the k*th *$\neg\wp$-labeled move is made earlier than (is to the left of) the n*th *$\wp$-labeled move in Γ, then so is it in Δ.*

Intuitively, "Δ is a $\wp$-delay of Γ" means that in Δ both players have played the same way as in Γ (condition 1), only, in Δ, $\wp$ might have been acting with some delay, i.e. slower than in Γ (condition 2). In more technical terms, Δ is the result of shifting in Γ some (maybe all, maybe none) $\wp$-labeled moves to the right; in the process of shifting, $\wp$-labeled moves can jump over $\neg\wp$-labeled moves, but a $\wp$-labeled move can never jump over another $\wp$-labeled

[9] From the American Mathematical Society review of Japaridze (2003) by Andreas Blass.

move. For example, the run $\Gamma = \langle \top\alpha, \bot\beta, \top\gamma, \bot\delta \rangle$ has exactly the following five $\top$-delays:

$$\Delta_1 = \langle \top\alpha, \bot\beta, \top\gamma, \bot\delta \rangle \ (= \Gamma);$$
$$\Delta_2 = \langle \bot\beta, \top\alpha, \top\gamma, \bot\delta \rangle;$$
$$\Delta_3 = \langle \top\alpha, \bot\beta, \bot\delta, \top\gamma \rangle;$$
$$\Delta_4 = \langle \bot\beta, \top\alpha, \bot\delta, \top\gamma \rangle;$$
$$\Delta_5 = \langle \bot\beta, \bot\delta, \top\alpha, \top\gamma \rangle.$$

Definition 23. *A constant game A is said to be* **static** *iff, for any player $\wp$ and any runs Γ, Δ such that Δ is a $\wp$-delay of Γ, we have:*

$$\textit{if } \mathbf{Wn}^A\langle\Gamma\rangle = \wp, \textit{ then } \mathbf{Wn}^A\langle\Delta\rangle = \wp.$$

This definition generalizes to all games by stipulating that a not-necessarily-constant game is **static** *iff every instance of it is so.*

Looking at the game of Figure 11.12, $\langle \bot\beta, \top\alpha \rangle$ is a $\top$-delay of $\langle \top\alpha, \bot\beta \rangle$. The latter is $\top$-won while the former is not. So, that game is not static. On the other hand, all of the other examples of games we have seen or will see in this paper are static. This is no surprise. In view of the following theorem, the closure of the set of all strict games—including all predicates—under all of our game operations forms a natural family of static games:

Theorem 24.

1. *Every strict game (and hence every elementary game) is static.*
2. *Every game operation defined in this paper preserves the static property of games.*

Proof. That all strict games are static has been proven in Japaridze (2003) (Proposition 4.8); and, of course, every elementary game is trivially strict. This takes care of clause 1. As for clause 2, it is a part of Theorem 14.1 of Japaridze (2003). Even though the operations $\curlywedge, \curlyvee, \bigwedge, \bigvee$ were not officially introduced in Japaridze (2003), they can be handled in exactly the same way as $\wedge, \vee$. □

See Section 4 of Japaridze (2003) for arguments in favor of the belief that static games are adequate formal counterparts of our intuition of "pure", speed-independent interactive computational problems. Based on that belief, CL uses the terms "static game" and (interactive) "**computational problem**" as synonyms. We have been informally using the concept of validity, which in intuitive terms was characterized as being a scheme of "always computable" problems. As will be seen from the formal definition of validity given in Section 11.7, the exact meaning of a "problem" is a static—rather than any—game.

All of the examples of winning strategies that we have seen so far shared one feature: for every position, the strategy had a strict prescription for a player regarding whether it should move or wait till the adversary makes a move. This

might have given us the wrong illusion of being a common case, somehow meaning that static games, even when properly free, still can always be "adequately" modeled as strict games. Not really. Below is an example, borrowed from Japaridze (2003), of an inherently free static game. The winning strategy for it substantially takes advantage of the flexibility offered by the free-game approach: the fact that it is not necessary for a player to precisely know whether in a given position it needs to move or wait. Any attempt to model such a game as a strict game would signify counterfeiting the true essence of the interactive computational problem that it represents.

Example 25. Let $A(x, z)$ be a decidable arithmetical predicate such that the predicate $\forall z A(x, z)$ is undecidable, and let $B(x, y)$ be an undecidable arithmetical predicate. Consider the following computational problem:

$$\sqcap x\Big(\sqcap y(\forall z A(x, z) \wedge B(x, y)) \sqcap \sqcap z A(x, z) \;\rightarrow\; \forall z A(x, z) \wedge \sqcap y B(x, y)\Big).$$

After $\bot$ specifies a value m for x, $\top$ will seemingly have to decide what to do: to watch or to think. The 'watch' choice is to wait till $\bot$ specifies a value k for y in the consequent, after which $\top$ can select the $\sqcap$-conjunct $\sqcap y(\forall z A(m, z) \wedge B(m, y))$ in the antecedent and specify y as k in it, thus bringing the play down to the always-won elementary game $\forall z A(m, z) \wedge B(m, k) \rightarrow \forall z A(m, z) \wedge B(m, k)$. While being successful if $\forall z A(m, z)$ is true, the watch strategy is a bad choice when $\forall z A(m, z)$ is false, for there is no guarantee that $\bot$ will indeed make a move in $\sqcap y B(m, y)$, and if not, the game will be lost. When $\forall z A(m, z)$ is false, the following 'think' strategy is successful: Start looking for a number n for which $A(m, n)$ is false. This can be done by testing $A(m, z)$, in turn, for $z = 1$, $z = 2$, ... After you find n, select the $\sqcap$-conjunct $\sqcap z A(m, z)$ in the antecedent, specify z as n in it, and you are the winner. The trouble is that if $\forall z A(m, z)$ is true, such a number n will never be found. Thus, which of the above two choices (watch or think) would be successful depends on whether $\forall z A(m, z)$ is true or false, and since $\forall z A(x, z)$ is undecidable, $\top$ has no effective way to make the right choice. Fortunately, there is no need to choose. Rather, these two strategies can be pursued simultaneously: $\top$ starts looking for a number n which makes $A(m, n)$ false and, at the same time, periodically checks if $\bot$ has made a move in $\sqcap y B(m, y)$. If the number n is found before $\bot$ makes such a move, $\top$ continues as prescribed by the think strategy; if vice versa, $\top$ continues as prescribed by the watch strategy; finally, if none of these two events ever occur, which, note, is only possible when $\forall z A(m, z)$ is true (for otherwise a number n falsifying $A(m, n)$ would have been found), again $\top$ will be the winner. This is so because, just as in the corresponding scenario of the watch strategy, $\top$ will have won both of the conjuncts of the consequent.

11.6 Winnability

Now that we know what computational problems are, it is time to explain what *computability*, i.e. *algorithmic solvability*, i.e. existence of an *algorithmic winning strategy* exactly means. The definitions given in this section are semi-formal. The omitted technical details are rather standard or irrelevant, and can be easily restored by anyone familiar with Turing machines. If necessary, the corresponding detailed definitions can be found in Part II of Japaridze (2003).

As we remember, the central point of our philosophy is to require that player $\top$ (here identified with its *strategy*) be implementable as a computer program, with effective and fully determined behavior. On the other hand, the behavior of $\bot$, including its speed, can be arbitrary. This intuition is captured by the model of interactive computation where $\top$ is formalized as what we call **HPM**.[10]

An HPM $\mathcal{H}$ is a Turing machine which, together with an ordinary read/write *work tape*, has two additional, read-only tapes: the *valuation tape* and the *run tape*. The presence of these two tapes is related to the fact that the outcome of a play over a given game depends on two parameters: (1) the valuation that tells us which instance of the game is played, and (2) the run that is generated in the play. $\mathcal{H}$ should have full access to information about these two parameters, and this information is provided by the valuation and run tapes: the former spells a (the "actual") valuation e by listing constants in the lexicographic order of the corresponding variables, and the latter spells, at any given time, the current position, i.e. the sequence of the (labeled) moves made by the two players so far. Thus, both of these two tapes can be considered input tapes. The reason for our choice to keep them separate is the difference in the nature of the input that they provide. Valuation is a *static* input, known at the very beginning of a computation/play and remaining unchanged throughout the subsequent process. On the other hand, the input provided by the run tape is *dynamic*: every time one of the players makes a move, the move (with the corresponding label) is appended to the content of this tape, with such content being unknown and hence blank at the beginning of interaction. Technically the run tape is read-only: the machine has unlimited read access to this (as well as to the valuation) tape, but it cannot write directly on it. Rather, $\mathcal{H}$ makes a move α by constructing it at the beginning of its work tape, delimiting its end with a blank symbol, and entering one of the specially designated states called *move states*. Once this happens, $\top\alpha$ is automatically appended to the current position spelled on the run tape. While the frequency at which the machine can make moves is naturally limited by its clock cycle time (the time each computation step takes), there are no limitations to how often the environment can

[10] HPM stands for 'Hard-Play Machine'. See Japaridze (2003) for a (little long) story about why "hard". The name EPM for the other model defined shortly stands for "Easy-Play Machine".

make a move, so, during one computation step of the machine, any finite number of any moves by the environment can be appended to the content of the run tape. This corresponds to the earlier-pointed-out intuition that not only the strategy, but also the relative speed of the environment can be arbitrary. For technical clarity, we assume that the run tape remains stable during a clock cycle, and is updated only on a transition from one cycle to another. Specifically, where $\langle\Phi\rangle$ is the position spelled on the run tape during a given cycle and $\alpha_1, \ldots, \alpha_n$ (possibly $n = 0$) is the sequence of the moves made by the environment during the cycle, the content of the run tape throughout the next cycle will be either $\langle\Phi, \bot\alpha_1, \ldots, \bot\alpha_n, \top\beta\rangle$ or $\langle\Phi, \bot\alpha_1, \ldots, \bot\alpha_n\rangle$, depending on whether the machine did or did not make a move β during the previous cycle. Such a transition is thus nondeterministic, with nondeterminism being related to the different possibilities for the above sequence $\alpha_1, \ldots, \alpha_n$.

A *configuration* of an HPM $\mathcal{H}$ is defined in the standard way: this is a full description of the ("current") state of the machine, the locations of its three scanning heads and the contents of its tapes, with the exception that, in order to make finite descriptions of configurations possible, we do not formally include a description of the unchanging (and possibly essentially infinite) content of the valuation tape as a part of configuration, but rather account for it in our definition of computation branch as this will be seen below. The *initial configuration* is the configuration where $\mathcal{H}$ is in its start state and the work and run tapes are empty. A configuration C' of $\mathcal{H}$ is said to be an ***e*-successor** of configuration C if, when valuation e is spelled on the valuation tape, C' can legally follow C in the standard sense, based on the transition function (which we assume to be deterministic) of the machine and accounting for the possibility of the above-described nondeterministic updates of the content of the run tape. An ***e*-computation branch** of $\mathcal{H}$ is a sequence of configurations of $\mathcal{H}$ where the first configuration is the initial configuration and each other configuration is an e-successor of the previous one. Thus, the set of all e-computation branches captures all possible scenarios (on valuation e) corresponding to different behaviors by $\bot$. Each e-computation branch B of $\mathcal{H}$ incrementally spells—in the sense that must be clear—a run Γ on the run tape, which we call the **run spelled by** B.

Definition 26. *For games A and B we say that:*

1. *An HPM $\mathcal{H}$* **wins** *A* **on a valuation** *e iff, whenever Γ is the run spelled by some e-computation branch of $\mathcal{H}$, $\mathbf{Wn}^{e[A]}\langle\Gamma\rangle = \top$.*

2. *An HPM $\mathcal{H}$ (simply)* **wins** *A iff it wins A on every valuation.*

3. *A is* **winnable** *iff there is an HPM that wins A. Such an HPM is said to be a* **solution** *for A.*

4. *A is* **reducible** *to B iff $B \rightarrow A$ is winnable. An HPM that wins $B \rightarrow A$ is said to be a* **reduction** *of A to B.*

5. *A and B are* **equivalent** *iff A is reducible to B and vice versa.*

The HPM model of interactive computation seemingly strongly favors the environment in that the latter may be arbitrarily faster than the machine. What happens if we start limiting the speed of the environment? The answer is *nothing* as far as computational problems, i.e. static games, are concerned. The alternative model of computation called EPM takes the idea if limiting the speed of the environment to the extreme by always letting the machine to decide when the environment can move and when it should wait; yet, as it turns out, the EPM model yields the same class of winnable static games as the HPM model does.

An **EPM** is a machine defined in the same way as an HPM, with the only difference that now the environment can (but is not obligated to) make a move only when the machine explicitly allows it to do so, the event called **granting permission**. Technically permission is granted by entering one of the specially designated states called **permission states**. The only requirement that the machine is expected to satisfy is that, as long as the adversary is playing legal, the machine should grant permission every once in a while; how long that "while" lasts, however, is totally up to the machine. This amounts to having full control over the speed of the adversary.

The above intuition is formalized as follows. After correspondingly redefining the '*e*-successor' relation—in particular, accounting for the condition that now a (one single) $\bot$-labeled move may be appended to the contents of the run tape only when the machine is in a permission state—the concepts of an *e-computation branch* of an EPM, the *run spelled* by such a branch, etc. are defined in exactly the same way as for HPMs. We say that a computation branch B of an EPM is **fair** iff permission is granted infinitely many times in B.

Definition 27. *For a game A and an EPM $\mathcal{E}$, we say that:*

1. $\mathcal{E}$ **wins** *A* **on a valuation** *e iff, whenever* Γ *is the run spelled by some e-computation branch B of* $\mathcal{E}$*, unless* Γ *is a* $\bot$*-illegal run of e*[*A*]*, B is fair and* $\mathbf{Wn}^{e[A]}\langle\Gamma\rangle = \top$.

2. $\mathcal{E}$ *(simply)* **wins** *A iff it wins A on every valuation.*

We will be using the expressions {HPMs} and {EPMs} for the sets of all HPMs and all EPMs, respectively.

The following fact, proven in Japaridze (2003) (Theorem 17.2), establishes equivalence between the two models of computation for static games:

Theorem 28. *There is an effective function* f : {*EPMs*}$\longrightarrow${*HPMs*} *such that, for every EPM* $\mathcal{E}$ *and static game A (and valuation e), whenever* $\mathcal{E}$ *wins A (on e), so does* $f(\mathcal{E})$*. And vice versa: there is an effective function* f : {*HPMs*}$\longrightarrow${*EPMs*} *such that, for every HPM* $\mathcal{H}$ *and static game A (and valuation e), whenever* $\mathcal{H}$ *wins A (on e), so does* $f(\mathcal{H})$.

The philosophical significance of this theorem is that it reinforces the belief according to which static games are games that allow us to make full

abstraction from speed. Its technical importance is related to the fact that the EPM-model is much more convenient when it comes to describing interactive algorithms/strategies, as we will have a chance to see later. The two models also act as natural complements to each other: as shown in Section 20 of Japaridze (2003), we can meaningfully talk about the (uniquely determined) play between a given HPM and a given EPM, while this is impossible if both players are EPMs or both are HPMs. This fact has been essentially exploited in the completeness theorems for logic **CL4** and its fragments proven in Japaridze, (2006a, c, d, 2007a), where environment's strategies for the games represented by unprovable formulas were described as EPMs and then it was shown that no HPM can win against such EPMs.

In view of Theorem 28, winnability of a static game A can be equivalently defined as existence of an EPM (rather than HPM) that wins A. Since we are only concerned with static games, from now on we will treat either definition as an equally official definition of winnability. And we extend the usage of the terms **solution** and **reduction** (Definition 26) from HPMs to EPMs. For a static game A, valuation e and HPM or EPM $\mathcal{M}$, we write

$$\mathcal{M} \models_e A, \quad \mathcal{M} \models A \text{ and } \models A$$

to mean that $\mathcal{M}$ wins A on valuation e, that $\mathcal{M}$ (simply) wins A and that A is winnable, respectively. Also, we will be using the terms "**computable**" or "**algorithmically solvable**" as synonyms of "winnable".

One might guess that, just as the ordinary Turing machine model, our HPM and EPM models are highly rigid with respect to reasonable technical variations. For example, the models where only environment's moves are visible to the machine yield the same class of winnable static games. Similarly, there is no difference between whether we allow the scanning heads on the valuation and run tapes to move in either or only one (left to right) direction. Another variation is the one where an attempt by either player to make an illegal move has no effect: such moves are automatically rejected and/or filtered out by some interface hardware or software and thus illegal runs are never generated. Obviously in such a case a minimum requirement would be that the question of legality of moves be decidable (which is indeed "very easily decidable" for naturally emerging games, including all games from the closure of the set of predicates under all of our game operations). This again yields models equivalent to HPM and/or EPM.

11.7 Validity

While winnability is a property of games, validity is a property of logical formulas, meaning that the formula is a scheme of winnable static games. To define this concept formally, we need to agree on a formal language first. It is going to be an extension of the language of classical predicate calculus without

identity or functional symbols. Our language is more expressive than the latter not only because it has non-classical operators in it such as $\sqcap, \bigsqcap, \wp$ etc., but also due to the fact that we now have two sorts of atoms: *elementary* and *general*. Elementary atoms represent elementary games, i.e. predicates, while general atoms represent any computational problems, i.e. any (not-necessarily-elementary) static games. The point is that elementary problems are interesting and meaningful in their own right, and validate principles that may not be valid in the general case. We want to be able to analyze games at a reasonably fine level, which is among the main reasons for our choice to have the two sorts of atoms in the language.

More formally, for each integer $n \geq 0$, our language has infinitely many n-ary **elementary letters** and n-ary **general letters**. Elementary letters are what is called *predicate letters* in ordinary logic. We will consistently use the lowercase p, q, r, s as metavariables for elementary letters, and the uppercase P, Q, R, S as metavariables for general letters. A **nonlogical atom** is $L(t_1, \ldots, t_n)$, where L is an n-ary elementary or general letter, and each t_i is a **term**, i.e. one of the **variables** $v_1, v_2, v_3, \ldots$ or one of the **constants** $1, 2, 3, \ldots$. Such an atom $L(t_1, \ldots, t_n)$ is said to be L**-based**. When L is 0-ary, the only L-based atom will be usually written as L rather than $L()$. An L-based atom is said to be elementary, general, n-ary, etc. if L is so. We also have two **logical atoms** $\top$ and $\bot$. **Formulas** are constructed from atoms and variables in the standard way applying to them the unary connectives $\neg, \curlywedge, \curlyvee, \wp, ?$, the binary connectives $\rightarrow, \succ, \multimap$, the variable- ($\geq$2) arity connectives $\wedge, \vee, \sqcap, \sqcup$, and the quantifiers $\forall, \exists, \bigsqcap, \bigsqcup, \bigwedge, \bigvee$. Throughout the rest of this paper, unless otherwise specified, "formula" will always mean a formula of this language, and letters E, F, G, H will be used as metavariables for formulas. We also continue using x, y, z as metavariables for variables, c for constants and t for terms.

The definitions of a *bound occurrence* and a *free occurrence* of a variable are standard. They extend from variables to all terms by stipulating that an occurrence of a constant is always free. When an occurrence of a variable x is within the scope of Qx for several quantifiers Q, then x is considered bound by the quantifier "nearest" to it. For instance, the occurrence of x within $Q(x, y)$ in $\forall x(P(x) \vee \bigsqcap x \bigwedge y Q(x, y))$ is bound by $\bigsqcap x$ rather than $\forall x$, for the latter is overridden by the former. An occurrence of a variable that is bound by $\forall x$ or $\exists x$ is said to be **blindly bound**.

In concordance with a similar notational practice established earlier for games, sometimes we represent a formula F as $F(x_1, \ldots, x_n)$, where the x_i are pairwise distinct variables. In the context set by such a representation, $F(t_1, \ldots, t_n)$ will mean the result of simultaneously replacing in F all free occurrences of each variable x_i ($1 \leq i \leq n$) by term t_i. In case each t_i is a variable y_i, it may be not clear whether $F(x_1, \ldots, x_n)$ or $F(y_1, \ldots, y_n)$ was originally meant to represent F in a given context. Our disambiguating convention is that

the context is set by the expression that was used earlier. That is, when we first mention $F(x_1, \ldots, x_n)$ and only after that use the expression $F(y_1, \ldots, y_n)$, the latter should be understood as the result of replacing variables in the former rather than vice versa. It should be noted that, when representing F as $F(x_1, \ldots, x_n)$, we do not necessarily mean that $x_1, \ldots, x_n$ are exactly the variables that have free occurrences in F.

An **interpretation** is a function * that sends each n-ary elementary (resp. general) letter L to an elementary (resp. static) game with a fixed attached n-tuple $x_1, \ldots, x_n$ of variables. We denote such a game by $L^*(x_1, \ldots, x_n)$, and call the tuple $(x_1, \ldots, x_n)$ the **canonical tuple of** L^*. When we do not care about the canonical tuple, simply L^* can be written instead of $L^*(x_1, \ldots, x_n)$. According to our earlier conventions, $x_1, \ldots, x_n$ have to be neither *all* nor the *only* variables on which the game $L^* = L^*(x_1, \ldots, x_n)$ depends; in fact, L^* does not even have to be finitary here. The canonical tuple is only used for setting a context, in which $L^*(t_1, \ldots, t_n)$ can be conveniently written later for $L^*(x_1/t_1, \ldots, x_n/t_n)$. This eliminates the need to have a special syntactic construct in the language for the operation of substitution of variables.

Interpretations are meant to turn formulas into games. Not every interpretation is equally good for every formula though, and some precaution is necessary to avoid confusing collisions of variables, as well as to guarantee that $\forall x, \exists x$ are only applied to games for which they are defined, i.e. games unistructural in x. For this reason, we restrict interpretations to "admissible" ones. We say that an interpretation * is **admissible for** a formula F, or simply is **F-admissible** iff, for every n-ary (general or elementary) letter L occurring in F, the following two conditions are satisfied:

(i) L^* does not depend on any variables that are not among its canonical tuple but occur in F.

(ii) If the ith ($1 \leq i \leq n$) term of an occurrence of an L-based atom in F is blindly bound, then L^* is unistructural in the ith variable of its canonical tuple.

Every interpretation * extends from letters to formulas for which * is admissible in the obvious way:

- Where L is an n-ary letter with $L^* = L^*(x_1, \ldots, x_n)$ and $t_1, \ldots, t_n$ are any terms, $(L(t_1, \ldots, t_n))^* = L^*(t_1, \ldots, t_n)$.
- * respects the meanings of logical operators (including logical atoms as 0-ary operators) as the corresponding game operations; that is: $\top^* = \top$; $(\neg G)^* = \neg(G^*)$; $(G \sqcap H)^* = (G^*) \sqcap (H^*)$; $(\forall x G)^* = \forall x (G^*)$; etc.

When $F^* = A$, we say that * **interprets** F as A, and that F^* is **an interpretation** of F.

Notice that condition (ii) of admissibility is automatically satisfied when L is an elementary letter, because an elementary problem (i.e. L^*) is always unistructural and hence unistructural in all variables. In most typical cases we will be interested in interpretations * where L^* is unistructural and does not depend on any variables other than those of its canonical tuple, so that both conditions (i) and (ii) will be automatically satisfied. With this remark in mind and in order to relax terminology, henceforth we may sometimes omit "F-admissible" and simply say "interpretation"; every time an expression F^* is used in a context, it should be understood that the range of * is restricted to F-admissible interpretations.

Definition 29. *We say that a formula F is* **valid***—and write $\Vdash F$—iff, for every F-admissible interpretation *, the game F^* is winnable.*

The main technical goal of CL at this stage of its development is to find axiomatizations for the set of valid formulas or various nontrivial fragments of that set. A considerable progress has already been achieved in this direction; more is probably yet to come in the future.

11.8 Uniform validity

If we disabbreviate "F^* is winnable" as $\exists\mathcal{M}(\mathcal{M} \models F^*)$ where $\mathcal{M}$ ranges over HPMs or EPMs, validity in the sense of Definition 29 can be written as $\forall^*\exists\mathcal{M}(\mathcal{M} \models F^*)$. Reversing the order of quantification yields the following stronger property of uniform validity:

Definition 30. *We say that a formula F is* **uniformly valid***—and write $\Vdash\!\!\Vdash F$—iff there is an HPM or (equivalently) EPM $\mathcal{M}$ such that, for every F-admissible interpretation *, $\mathcal{M} \models F^*$.*

Such an HPM or EPM $\mathcal{M}$ is said to be a **uniform solution** *for F, and $\mathcal{M} \Vdash\!\!\Vdash F$ is written to express that $\mathcal{M}$ is a uniform solution for F.*

Intuitively, a uniform solution $\mathcal{M}$ for a formula F is an interpretation-independent winning strategy: since, unlike valuation, the "intended" or "actual" interpretation * is not visible to the machine, $\mathcal{M}$ has to play in some standard, uniform way that would be successful for any possible interpretation of F.

The term "uniform" is borrowed from Abramsky and Jagadeesan (1994) as this understanding of validity in its spirit is close to that in Abramsky and Jagadeesan's tradition. The concepts of validity in Lorenzen (1959) tradition, or in the sense of Japaridze (2000, 2002), also belong to this category. Common to those uniform-validity-style notions is that validity there is not defined as being "always true" (true = winnable) as this is the case with the classical understanding of this concept; in those approaches the concept of truth is often simply absent, and validity is treated as a basic concept in its own rights.

As for simply validity, it is closer to validities in the sense of Blass (1992) or Japaridze (1997), and presents a direct generalization of the corresponding classical concept in that it indeed means being "true" in every particular setting.

Which of our two versions of validity is more interesting depends on the motivational standpoint. It is validity rather than uniform validity that tells us what can be computed in principle. So, a computability-theoretician would focus on validity. Mathematically, non-validity is generally by an order of magnitude more informative—and correspondingly harder to prove—than non-uniform-validity. Say, the non-validity of $p \sqcup \neg p$ means existence of solvable-in-principle yet algorithmically unsolvable problems[11]—the fact that became known to mankind only as late as in the twentieth century. As for the non-uniform-validity of $p \sqcup \neg p$, it is trivial: of course there is no way to choose one of the two disjuncts that would be true for all possible values of p because, as the Stone Age intellectuals were probably aware, some p are true and some are false.

On the other hand, it is uniform validity rather than validity that is of interest in more applied areas of computer science such as knowledge base systems or systems for planning and action (see Section 11.10). In this sort of applications we want a logic on which a universal problem-solving machine can be based. Such a machine would or should be able to solve problems represented by logical formulas without any specific knowledge of the meanings of their atoms, i.e. without knowledge of the actual interpretation. Remembering what was said about the intuitive meaning of uniform validity, this concept is exactly what fits the bill.

Anyway, the good news is that the two concepts of validity appear to yield the same logic. This has been conjectured for the full language of CL in Japaridze (2003) (Conjecture 26.2), and by now, as will be seen from our Theorem 35, successfully verified for the rather expressive fragment of that language—the language of logic **CL4**.

11.9 Logic CL4

The language of logic **CL4** is the fragment of the language of Section 11.7 obtained by forbidding the parallel group of quantifiers and the recurrence group of propositional connectives. This leaves us with the operators $\neg$, $\wedge$, $\vee$,

[11] Well, saying so is only accurate with the Strong Completeness clause of Theorem 34 (which, as conjectured in Japaridze (2003), extends from **CL4** to any other complete fragments of CL) in mind, according to which the non-validity of $p \sqcup \neg p$ implies the existence of a *finitary* predicate A for which $A \sqcup \neg A$ has no algorithmic solution. As will be pointed out in a comment following Theorem 34, without the finitarity restriction, a machine may fail to win $A \sqcup \neg A$ not (only) due to the fundamental limitations of algorithmic methods, but rather due to the fact that it can never finish reading all necessary information from the valuation tape to determine the truth status of A.

$\rightarrow$, $\sqcap$, $\sqcup$, $\forall$, $\exists$, $\bigsqcap$, $\bigsqcup$, along with the logical atoms $\top$, $\bot$ and the two sorts (elementary and general) of nonlogical atoms. Furthermore, for safety and without loss of expressive power, we agree that a formula cannot contain both bound and free occurrences of the same variable. We refer to the formulas of this language as **CL4-formulas**.

Our axiomatization of **CL4** employs the following terminology. Understanding $F \rightarrow G$ as an abbreviation for $\neg F \vee G$, a **positive** (resp. **negative**) **occurrence** of a subformula is one that is in the scope of an even (resp. odd) number of occurrences of $\neg$. A **surface occurrence** of a subformula is an occurrence that is not in the scope of any choice operators. A **CL4**-formula not containing general atoms and choice operators—i.e. a formula of the language of classical first-order logic—is said to be **elementary**. The **elementarization** of a **CL4**-formula F is the result of replacing in F all surface occurrences of each subformula of the form $G_1 \sqcup \ldots \sqcup G_n$ or $\bigsqcup xG$ by $\bot$, all surface occurrences of each subformula of the form $G_1 \sqcap \ldots \sqcap G_n$ or $\bigsqcap xG$ by $\top$, all positive surface occurrences of each general atom by $\bot$, and all negative surface occurrences of each general atom by $\top$. A **CL4**-formula is said to be **stable** iff its elementarization is classically valid, i.e. provable in classical predicate calculus. Otherwise it is **instable**.

With $\mathcal{P} \mapsto C$ meaning "from premise(s) $\mathcal{P}$ conclude C", logic **CL4** is given by the following four rules where, as can be understood, both the premises and the conclusions range over **CL4**-formulas:

A $\vec{H} \mapsto E$, where E is stable and $\vec{H}$ is a set of formulas satisfying the following conditions:

(i) Whenever E has a positive (resp. negative) surface occurrence of a subformula $G_1 \sqcap \ldots \sqcap G_n$ (resp. $G_1 \sqcup \ldots \sqcup G_n$), for each $i \in \{1, \ldots, n\}$, $\vec{H}$ contains the result of replacing that occurrence in E by G_i.

(ii) Whenever E has a positive (resp. negative) surface occurrence of a subformula $\bigsqcap xG(x)$ (resp. $\bigsqcup xG(x)$), $\vec{H}$ contains the result of replacing that occurrence in E by $G(y)$ for some variable y not occurring in E.

B1 $H \mapsto E$, where H is the result of replacing in E a negative (resp. positive) surface occurrence of a subformula $G_1 \sqcap \ldots \sqcap G_n$ (resp. $G_1 \sqcup \ldots \sqcup G_n$) by G_i for some $i \in \{1, \ldots, n\}$.

B2 $H \mapsto E$, where H is the result of replacing in E a negative (resp. positive) surface occurrence of a subformula $\bigsqcap xG(x)$ (resp. $\bigsqcup xG(x)$) by $G(t)$ for some term t such that (if t is a variable) neither the above occurrence of $\bigsqcap xG(x)$ (resp. $\bigsqcup xG(x)$) in E nor any of the free occurrences of x in $G(x)$ are in the scope of $\forall t$, $\exists t$, $\bigsqcap t$ or $\bigsqcup t$.

C $H \mapsto E$, where H is the result of replacing in E two—one positive and one negative—surface occurrences of some n-ary general letter by an n-ary elementary letter that does not occur in E.

Axioms are not explicitly stated, but note that the set of premises of Rule **A** sometimes can be empty, in which case the conclusion acts as an axiom. Looking at a few examples should help us get a syntactic feel of this most unusual deductive system.

The following is a **CL4**-proof of $\sqcap x \sqcup y(P(x) \rightarrow P(y))$:

1. $p(z) \rightarrow p(z)$ (from {} by Rule **A**);
2. $P(z) \rightarrow P(z)$ (from 1 by Rule **C**);
3. $\sqcup y(P(z) \rightarrow P(y))$ (from 2 by Rule **B2**);
4. $\sqcap x \sqcup y(P(x) \rightarrow P(y))$ (from {3} by Rule **A**).

On the other hand, $\mathbf{CL4} \not\vdash \sqcup y \sqcap x(P(x) \rightarrow P(y))$. Indeed, obviously this instable formula cannot be the conclusion of any rule but **B2**. If it is derived by this rule, the premise should be $\sqcap x(P(x) \rightarrow P(t))$ for some term t different from x. $\sqcap x(P(x) \rightarrow P(t))$, in turn, could only be derived by Rule **A** where, for some variable z different from t, $P(z) \rightarrow P(t)$ is a (the) premise. The latter is an instable formula and does not contain choice operators, so the only rule by which it can be derived is **C**, where the premise is $p(z) \rightarrow p(t)$ for some elementary letter p. Now we deal with a classically non-valid and hence instable elementary formula, and it cannot be derived by any of the four rules of **CL4**.

Note that, in contrast, the "blind version" $\exists y \forall x(P(x) \rightarrow P(y))$ of $\sqcup y \sqcap x(P(x) \rightarrow P(y))$ is provable:

1. $\exists y \forall x(p(x) \rightarrow p(y))$ (from {} by Rule **A**);
2. $\exists y \forall x(P(x) \rightarrow P(y))$ (from 1 by Rule **C**).

'*There is y such that, for all x,* $P(x) \rightarrow P(y)$' is true yet not in a constructive sense, thus belonging to the kind of principles that have been fueling controversies between the classically- and constructivistically-minded. As noted in Section 11.1, computability logic is offering a peaceful settlement, telling the arguing parties: "There is no need to fight at all. It appears that you simply have two different concepts of '*there is*'/'*for all*'. So, why not also use two different names: $\exists/\forall$ and $\sqcup/\sqcap$. Yes, $\exists y \forall x(P(x) \rightarrow P(y))$ is indeed right; and yes, $\sqcup y \sqcap x(P(x) \rightarrow P(y))$ is indeed wrong." Clauses 1 and 2 of Exercise 31 illustrate a similar solution for the law of the excluded middle, the most controversial principle of classical logic.

The above-said remains true with p instead of P, for what is relevant there is the difference between the constructive and non-constructive versions of logical operators rather than how atoms are understood. Then how about the difference between the elementary and non-elementary versions of atoms? This

distinction allows computability logic to again act in its noble role of a reconciliator/integrator, but this time between classical and linear logics, telling them: "It appears that you have two different concepts of the objects that logic is meant to study. So, why not also use two different sorts of atoms to represent such objects: elementary atoms $p, q, \ldots$, and general atoms $P, Q, \ldots$. Yes, $p \to p \wedge p$ is indeed right; and yes, $P \to P \wedge P$ (Exercise 31(4)) is indeed wrong". However, as pointed out in Section 11.1, the term "linear logic" in this context should be understood in a very generous sense, referring not to the particular deductive system proposed by Girard (1987) but rather to the general philosophy and intuitions traditionally associated with it. The formula of clause 3 of the following exercise separates **CL4** from linear logic. That formula is provable in affine logic though. Switching to affine logic, i.e. restoring the deleted (from classical logic) rule of weakening, does not however save the case: the **CL4**-provable formulas of clauses 10, 11 and 18 of the exercise are provable in neither linear nor affine logics.

Exercise 31. In clauses 14 and 15 below, "**CL4** $\vdash E \Leftrightarrow F$" stands for "**CL4** $\vdash E \to F$ *and* **CL4** $\vdash F \to E$". Verify that:

1. **CL4** $\vdash P \vee \neg P$.
2. **CL4** $\nvdash P \sqcup \neg P$. Compare with 1.
3. **CL4** $\vdash P \wedge P \to P$.
4. **CL4** $\nvdash P \to P \wedge P$. Compare with 3,5.
5. **CL4** $\vdash P \to P \sqcap P$.
6. **CL4** $\vdash (P \sqcup Q) \wedge (P \sqcup R) \to P \sqcup (Q \wedge R)$.
7. **CL4** $\nvdash P \sqcup (Q \wedge R) \to (P \sqcup Q) \wedge (P \sqcup R)$. Compare with 6,8.
8. **CL4** $\vdash p \sqcup (Q \wedge R) \to (p \sqcup Q) \wedge (p \sqcup R)$.
9. **CL4** $\nvdash p \sqcap (Q \wedge R) \to (p \sqcap Q) \wedge (p \sqcap R)$. Compare with 8.
10. **CL4** $\vdash (P \wedge P) \vee (P \wedge P) \to (P \vee P) \wedge (P \vee P)$.
11. **CL4** $\vdash (P \wedge (R \sqcap S)) \sqcap (Q \wedge (R \sqcap S)) \sqcap ((P \sqcap Q) \wedge R) \sqcap ((P \sqcap Q) \wedge S) \to (P \sqcap Q) \wedge (R \sqcap S)$.
12. **CL4** $\vdash \forall x P(x) \to \sqcap x P(x)$.
13. **CL4** $\nvdash \sqcap x P(x) \to \forall x P(x)$. Compare with 12.
14. **CL4** $\vdash \exists x P(x) \sqcap \exists x Q(x) \Leftrightarrow \exists x (P(x) \sqcap Q(x))$.

Similarly for $\sqcup$ instead of $\sqcap$, and/or $\forall$ instead of $\exists$.

15. **CL4** $\vdash \sqcap x \exists y P(x, y) \Leftrightarrow \exists y \sqcap x P(x, y)$.

Similarly for $\bigsqcup$ instead of $\bigsqcap$, and/or $\forall$ instead of $\exists$.

16. **CL4** $\vdash \forall x (P(x) \wedge Q(x)) \to \forall x P(x) \wedge \forall x Q(x)$.
17. **CL4** $\nvdash \sqcap x (P(x) \wedge Q(x)) \to \sqcap x P(x) \wedge \sqcap x Q(x)$. Compare with 16.
18. **CL4** $\vdash \sqcap x \Big((P(x) \wedge \sqcap x Q(x)) \sqcap (\sqcap x P(x) \wedge Q(x)) \Big) \to \sqcap x P(x) \wedge \sqcap x Q(x)$.
19. **CL4** $\vdash$ formula (3) of Section 11.4.4.
20. **CL4** $\nvdash$ formula (4) of Section 11.4.4. Compare with 19.

Taking into account that classical validity and hence stability is recursively enumerable, from the way **CL4** is axiomatized it is obvious that the set of theorems of **CL4** is recursively enumerable. Not so obvious, however, may be the following theorem proven in Japaridze (2007a). As it turns out, the choice/constructive quantifiers $\sqcap, \sqcup$ are much better behaved than their blind/classical counterparts $\forall, \exists$, yielding a decidable first-order logic:

Theorem 32. *The* $\forall, \exists$*-free fragment of* (*the set of theorems of*) **CL4** *is decidable.*

Next, based on the straightforward observation that elementary formulas are derivable in **CL4** (in particular, from the empty set of premises by Rule **A**) exactly when they are classically valid, we have:

Theorem 33. CL4 *is a conservative extension of classical predicate logic: the latter is nothing but the elementary fragment (i.e. the set of all elementary theorems) of the former.*

Remember that a predicate A is said to be of arithmetical complexity Δ_2 iff $A = \exists x \forall y B_1$ and $\neg A = \exists x \forall y B_2$ for some decidable predicates B_1 and B_2.

The following Theorem 34 is the strongest soundness and completeness result known so far in computability logic. Its proof has taken about half of the volume of Japaridze (2006a) and almost entire Japaridze (2007a). A similar theorem for the propositional version **CL2** of **CL4** was proven in Japaridze (2006c, d).

Theorem 34. CL4 $\vdash F$ *iff* F *is valid (any* **CL4***-formula* F*). Furthermore:*

Uniform-Constructive Soundness: *There is an effective procedure that takes a* **CL4***-proof of an arbitrary* **CL4***-formula* F *and constructs a uniform solution for* F.

Strong Completeness: *If a* **CL4***-formula* F *is not provable in* **CL4***, then* F^* *is not computable for some* F*-admissible interpretation* * *that interprets all elementary atoms as finitary predicates of arithmetical complexity* Δ_2*, and interprets all general atoms as* $\sqcap, \sqcup$*-combinations of finitary predicates of arithmetical complexity* Δ_2.

A non-finitary game generally depends on infinitely many variables, and appealing to this sort of games in a completeness proof could seriously weaken such a result: the reason for incomputability of a non-finitary game could be just the fact that the machine can never finish reading all the relevant information from its valuation tape. Fortunately, in view of the Strong Completeness clause, it turns out that the question whether non-finitary games are allowed or not has no effect on the (soundness and) completeness of **CL4**; moreover, finitary games can be further restricted to the sort as simple as $\sqcap, \sqcup$-combinations of finitary predicates.

Similarly, the Uniform-Constructive Soundness clause dramatically strengthens the soundness result for **CL4** and, as will be argued in the following section, opens application areas far beyond logic or the pure theory of computation. First of all, notice that it immediately implies a positive verification of the earlier-mentioned Conjecture 26.2 of Japaridze (2003) restricted to the language of **CL4**, according to which validity and uniform validity are extensionally equivalent. Indeed, if a **CL4**-formula F is uniformly valid, then it is automatically also valid, as uniform validity is stronger than validity. Suppose now F is valid. Then, by the completeness part of Theorem 34, **CL4** $\vdash F$. But then, by the Uniform-Constructive Soundness clause, F is uniformly valid. Thus, we have:

Theorem 35. *A* **CL4***-formula is valid if and only if it is uniformly valid.*

But **CL4** is sound in an even stronger sense. Knowing that a solution for a given problem exists might be of little practical importance without being able to actually find such a solution. No problem: according to the Uniform-Constructive Soundness clause, a uniform solution for a **CL4**-provable formula F automatically comes with a **CL4**-proof of F. The earlier-mentioned soundness theorem for Heyting's intuitionistic calculus proven in Japaridze (2006b) comes in the same uniform-constructive form, and so does the soundness theorem for affine logic (Theorem 37) proven later in this paper.

11.10 Applied systems based on CL

The original motivations behind CL were computability-theoretic: the approach provides a systematic answer to the question 'what can be computed?', which is a fundamental question of computer science. Yet, the above discussion of the uniform-constructive nature of the known soundness theorems for various fragments of CL reveals that the CL paradigm is not only about *what* can be computed. It is equally about *how* problems can be computed/solved, suggesting that CL should have potential utility, with its application areas not limited to the theory of computation. In the present section we will briefly examine why and how CL is of interest in some other fields of study, specifically, knowledge base systems and constructive applied theories.

The reason for the failure of $p \sqcup \neg p$ as a computability-theoretic principle is that the problem represented by this formula may have no effective solution—that is, the predicate p^* may be undecidable. The reason why this principle would fail in the context of knowledge base systems, however, is much simpler. A knowledge base system may fail to solve the problem $Female(Dana) \sqcup \neg Female(Dana)$ not because the latter has no effective solution (of course it has one), but because the system simply lacks sufficient knowledge to determine Dana's gender. On the other hand, any system would be able to "solve" the problem $Female(Dana) \vee \neg Female(Dana)$ as this is an automatically won

elementary game so that there is nothing to solve at all. Similarly, while $\forall y \exists x \mathit{Father}(x, y)$ is an automatically solved elementary problem expressing the almost tautological knowledge that every person has a father, ability to solve the problem $\sqcap y \sqcup x \mathit{Father}(x, y)$ implies the nontrivial knowledge of everyone's actual father. Obviously the knowledge expressed by $A \sqcup B$ or $\sqcup x A(x)$ is generally stronger than the knowledge expressed by $A \vee B$ or $\exists x A(x)$, yet the formalism of classical logic fails to capture this difference—the difference whose relevance hardly requires any explanation. The traditional approaches to knowledge base systems (Konolige, 1988; Levesque and Lakemeyer, 2000; Moore, 1985 etc.) try to mend this gap by augmenting the language of classical logic with special epistemic constructs, such as the modal "know that" operator $\Box$, after which probably $\Box A \vee \Box B$ would be suggested as a translation for $A \sqcup B$ and $\forall y \exists x \Box A(x, y)$ for $\sqcap y \sqcup x A(x, y)$. Leaving it for the philosophers to argue whether, say, $\forall y \exists x \Box A(x, y)$ really expresses the constructive meaning of $\sqcap y \sqcup x A(x, y)$, and forgetting that epistemic constructs often yield unnecessary and very unpleasant complications such as messiness and non-semidecidability of the resulting logics, some of the major issues still do not seem to be taken care of. Most of the actual knowledge base and information systems are interactive, and what we really need is a logic of *interaction* rather than just a logic of knowledge. Furthermore, a knowledge base logic needs to be *resource-conscious*. The informational resource expressed by $\sqcap x(\mathit{Female}(x) \sqcup \neg \mathit{Female}(x))$ is not as strong as the one expressed by $\sqcap x(\mathit{Female}(x) \sqcup \neg \mathit{Female}(x)) \wedge \sqcap x(\mathit{Female}(x) \sqcup \neg \mathit{Female}(x))$: the former implies the resource provider's commitment to tell only one (even though an arbitrary one) person's gender, while the latter is about telling any two people's genders. A reader having difficulty in understanding why this difference is relevant, may try to replace $\mathit{Female}(x)$ with $\mathit{Acid}(x)$, and then think of a (single) piece of litmus paper. Neither classical logic nor its standard epistemic extensions have the ability to account for such differences. But CL promises to be adequate. It *is* a logic of interaction, it *is* resource-conscious, and it *does* capture the relevant differences between truth and actual ability to find/compute/know truth.

When CL is used as a logic of knowledge bases, its formulas represent interactive queries. A formula whose main operator is $\sqcup$ or $\sqcup$ can be understood as a question asked by the user, and a formula whose main operator is $\sqcap$ or $\sqcap$ as a question asked by the system. Consider the problem $\sqcap x \sqcup y \mathit{Has}(x, y)$, where $\mathit{Has}(x, y)$ means "patient x has disease y" (with *Healthy* counting as one of the possible "diseases"). This formula is the following question asked by the system: "Who do you want me to diagnose?" The user's response can be "Dana". This move brings the game down to $\sqcup y \mathit{Has}(\mathit{Dana}, y)$. This is now a question asked by the user: "What does Dana have?". The system's response can be "flu", taking us to the terminal position $\mathit{Has}(\mathit{Dana}, \mathit{Flu})$. The system has been successful iff Dana really has a flu.

Successfully solving the above problem $\sqcap x \sqcup y Has(x, y)$ requires having all relevant medical information for each possible patient, which in a real diagnostic system would hardly be the case. Most likely, such a system, after receiving a request to diagnose x, would make counterqueries regarding x's symptoms, blood pressure, test results, age, gender, etc., so that the query that the system will be solving would have a higher degree of interactivity than the two-step query $\sqcap x \sqcup y Has(x, y)$ does, with questions and counterquestions interspersed in some complex fashion. Here is when other computability-logic operations come into play. $\neg$ turns queries into counterqueries; parallel operations generate combined queries, with $\rightarrow$ acting as a query reduction operation; $\mathbin{\diamond}, \curlywedge$ allow repeated queries, etc. Here we are expanding our example. Let $Sympt(x, s)$ mean "patient x has (set of) symptoms s", and $Pos(x, t)$ mean "patient x tests positive for test t". Imagine a diagnostic system that can diagnose any particular patient x, but needs some additional information. Specifically, it needs to know x's symptoms; plus, the system may require to have x taken a test t that it selects dynamically in the course of a dialogue with the user depending on what responses it received. The interactive task/query that such a system is performing/solving can then be expressed by the formula

$$\sqcap x\Big(\sqcup s Sympt(x, s) \wedge \sqcap t(Pos(x, t) \sqcup \neg Pos(x, t)) \rightarrow \sqcup y Has(x, y)\Big). \qquad (6)$$

A possible scenario of playing the above game is the following. At the beginning, the system waits until the user specifies a patient x to be diagnosed. We can think of this stage as systems's requesting the user to select a particular (value of) x, remembering that the presence of $\sqcap x$ automatically implies such a request. After a patient x—say $x = X$—is selected, the system requests to specify X's symptoms. Notice that our game rules make the system successful if the user fails to provide this information, i.e. specify a (the true) value for s in $\sqcup s Sympt(X, s)$. Once a response—say, $s = S$—is received, the system selects a test $t = T$ and asks the user to perform it on X, i.e. to choose the true disjunct of $Pos(X, T) \sqcup \neg Pos(X, T)$. Finally, provided that the user gave correct answers to all counterqueries (and if not, the user has lost), the system makes a diagnostic decision, i.e. specifies a value Y for y in $\sqcup y Has(X, y)$ for which $Has(X, Y)$ is true.

The presence of a single "copy" of $\sqcap t(Pos(x, t) \sqcup \neg Pos(x, t))$ in the antecedent of (6) means that the system may request testing a given patient only once. If n tests were potentially needed instead, this would be expressed by taking the $\wedge$-conjunction of n identical conjuncts $\sqcap t(Pos(x, t) \sqcup \neg Pos(x, t))$. And if the system potentially needed an unbounded number of tests, then we would write $\curlywedge \sqcap t(Pos(x, t) \sqcup \neg Pos(x, t))$, thus further weakening (6): a system that performs this weakened task is not as good as the one performing (6) as it requires stronger external (user-provided) informational resources. Replacing the main

quantifier $\sqcap x$ by $\forall x$, on the other hand, would strengthen (6), signifying the system's ability to diagnose a patent purely on the basis of his/her symptoms and test result without knowing who the patient really is. However, if in its diagnostic decisions the system uses some additional information on patients such their medical histories stored in its knowledge base and hence needs to know the patient's identity, $\sqcap x$ cannot be upgraded to $\forall x$. Replacing $\sqcap x$ by $\bigwedge x$ would be a yet another way to strengthen (6), signifying the system's ability to diagnose all patients rather than any particular one; obviously effects of at least the same strength would be achieved by just prefixing (6) with $⅄$ or $⅂$.

As we just mentioned system's **knowledge base**, let us make clear what it means. Formally, this is a finite $\wedge$-conjunction *KB* of formulas, which can also be thought of as the (multi)set of its conjuncts. We call the elements of this set the **internal informational resources** of the system. Intuitively, *KB* represents all of the nonlogical knowledge available to the system, so that (with a fixed built-in logic in mind) the strength of the former determines the query-solving power of the latter. Conceptually, however, we do not think of *KB* as a part of the system properly. The latter is just "pure", logic-based problem-solving software of universal utility that initially comes to the user without any nonlogical knowledge whatsoever. Indeed, built-in nonlogical knowledge would make it no longer universally applicable: Dana can be a female in the world of one potential user while a male in the world of another user, and $\forall x \forall y(x \times y = y \times x)$ can be false to a user who understands $\times$ as Cartesian rather than number-theoretic product. It is the user who selects and maintains *KB* for the system, putting into it all informational resources that (s)he believes are relevant, correct and maintainable. Think of the formalism of CL as a highly declarative programming language, and the process of creating *KB* as programming in it.

The knowledge base *KB* of the system may include atomic elementary formulas expressing factual knowledge, such as $Female(Dana)$, or non-atomic elementary formulas expressing general knowledge, such as $\forall x(\exists y\, Father(x, y) \rightarrow Male(x))$ or $\forall x \forall y(x \times (y + 1) = (x \times y) + x)$; it can also include nonclassical formulas such as $⅂\sqcap x(Female(x) \sqcup Male(x))$, expressing potential knowledge of everyone's gender, or $⅂\sqcap x \sqcup y(x^2 = y)$, expressing ability to repeatedly compute the square function, or something more complex and more interactive, such as formula (6). With each resource $R \in KB$ is associated (if not physically, at least conceptually) its **provider**—an agent that solves the query *R* for the system, i.e. plays the game *R* against the system. Physically the provider could be a computer program allocated to the system, or a network server having the system as a client, or another knowledge base system to which the system has querying access, or even human personnel servicing the system. For example, the provider for $⅂\sqcap x \sqcup y Bloodpressure(x, y)$ would probably be a team of nurses repeatedly performing the task of measuring the blood pressure of a patient specified by the system and reporting the outcome

back to the system. Again, we do not think of providers as a part of the system itself. The latter only sees *what* resources are available to it, without knowing or caring about *how* the corresponding providers do their job; furthermore, the system does not even care *whether* the providers really do their job right. The system's responsibility is only to correctly solve queries for the user *as long as* none of the providers fail to do their job. Indeed, if the system misdiagnoses a patient because a nurse-provider gave it wrong information about that patient's blood pressure, the hospital (ultimate user) is unlikely to fire the system and demand refund from its vendor; more likely, it would fire the nurse. Of course, when R is elementary, the provider has nothing to do, and its successfully playing R against the system simply means that R is true. Note that in the picture that we have just presented, the system plays each game $R \in KB$ in the role of $\bot$, so that, from the system's perspective, the game that it plays against the provider of R is $\neg R$ rather than R.

The most typical internal informational resources, such as factual knowledge or queries solved by computer programs, can be reused an arbitrary number of times and with unlimited branching capabilities, i.e. in the strong sense captured by $!$, and thus they would be prefixed with a $!$ as we did with $\sqcap x(Female(x) \sqcup Male(x))$ and $\sqcap x \sqcup y(x^2 = y)$. There was no point in $!$-prefixing $Female(Dana)$, $\forall x(\exists y Father(x, y) \rightarrow Male(x))$ or $\forall x \forall y(x \times (y+1) = (x \times y) + x)$ because every elementary game A is equivalent to $!A$ and hence remains "recyclable" even without recurrence operators. As noted in Section 11.4.6, there is no difference between $!$ and $?$ as long as "simple" resources such as $\sqcap x \sqcup y(x^2 = y)$ are concerned. However, in some cases—say, when a resource with a high degree of interactivity is supported by an unlimited number of independent providers each of which however allows to run only one single "session"—the weaker operator $?$ will have to be used instead of $!$. Yet, some of the internal informational resources could be essentially non-reusable. A single provider possessing a single item of disposable pregnancy test device would apparently be able to support the resource $\sqcap x(Pregnant(x) \sqcup \neg Pregnant(x))$ but not $!\sqcap x(Pregnant(x) \sqcup \neg Pregnant(x))$ and not even $\sqcap x(Pregnant(x) \sqcup \neg Pregnant(x)) \wedge \sqcap x(Pregnant(x) \sqcup \neg Pregnant(x))$. Most users, however, would try to refrain from including this sort of a resource into KB, and rather make it a part (antecedent) of possible queries. Indeed, knowledge bases with non-recyclable resources would tend to weaken from query to query and require more careful maintenance and updates. Whether recyclable or not, all of the resources of KB can be used independently and in parallel. This is exactly what allows us to identify KB with the $\wedge$-conjunction of its elements.

Assume $KB = R_1 \wedge \ldots \wedge R_n$, and let us now try to visualize a system solving a query F for the user. The designer would probably select an interface where the user only sees the moves made by the system in F, and hence gets the

illusion that the system is just playing F. But in fact the game that the system is really playing is $KB \rightarrow F$, i.e. $\neg R_1 \vee \ldots \vee \neg R_n \vee F$. Indeed, the system is not only interacting with the user in F, but—in parallel—also with its resource providers against whom, as we already know, it plays $\neg R_1, \ldots, \neg R_n$. As long as those providers do not fail to do their job, the system loses each of the games $\neg R_1, \ldots, \neg R_n$. Then our semantics for $\vee$ implies that the system wins its play over the "big game" $\neg R_1 \vee \ldots \vee \neg R_n \vee F$ if and only if it wins it in the F component, i.e. successfully solves the query F.

Thus, the system's ability to solve a query F reduces to its ability to generate a solution for $KB \rightarrow F$, i.e. a reduction of F to KB. What would give the system such an ability is built-in knowledge of CL—in particular, a **uniform-constructively sound axiomatization** of it, by which we mean a deductive system S (with effective proofs of its theorems) that satisfies the Uniform-Constructive Soundness clause of Theorem 34 with "S" in the role of **CL4**. According to the uniform-constructive soundness property, it would be sufficient for the system to find a proof of $KB \rightarrow F$, which would allow it to (effectively) construct a machine $\mathcal{M}$ and then run it on $KB \rightarrow F$ with a guaranteed success.

Notice that it is uniform-constructive soundness rather than simple soundness of the built-in (axiomatization of the) logic that allows the knowledge base system to function. Simple soundness just means that every provable formula is valid. This is not sufficient for two reasons.

One reason is that validity of a formula E only implies that, for every interpretation *, a solution for the problem E^* exists. It may be the case, however, that different interpretations require different solutions, so that choosing the right solution requires knowledge of the actual interpretation, i.e. the *meaning*, of the atoms of E. Our assumption is that the system has no nonlogical knowledge, which, in more precise terms, means nothing but that it has no knowledge of the interpretation *. Thus, a solution that the system generates for E^* should be successful for any possible interpretation *. In other words, it should be a uniform solution for E. This is where uniform-constructive soundness of the underlying logic becomes critical, by which every provable formula is not only valid, but also uniformly valid. Going back to the example with which this section started, the reason why $p \sqcup \neg p$ fails in the context of computability theory is that it is not valid. On the other hand, the reason for the failure of this principle in the context of knowledge base systems is that it is not uniformly valid: a solution for it, even if such existed for each interpretation *, would depend on whether p^* is true or false, and the system would be unable to figure out the truth status of p^* unless this information was explicitly or implicitly contained in KB. Thus, for knowledge base systems the primary semantical concept of interest is uniform validity rather than validity.

The other reason why simple soundness of the built-in logic would not be sufficient for a knowledge base system to function—even if every provable formula was known to be uniformly valid—is the following. With simple soundness, after finding a proof of E, even though the system would know that a solution for E^* exists, it might have no way to actually find such a solution. On the other hand, uniform-constructive soundness guarantees that a (uniform) solution for every provable formula not only exists, but can be effectively extracted from a proof.

As for completeness of the built-in logic, unlike uniform-constructive soundness, it is a desirable but not necessary condition. So far a complete axiomatization has been found only for the fragment of CL limited to the language of **CL4**. We hope that the future will bring completeness results for more expressive fragments as well. But even if not, we can still certainly succeed in finding ever stronger axiomatizations that are uniform-constructively sound even if not necessarily complete. Extending **CL4** with some straightforward rules such as the ones that allow to replace $⫰F$ by $F \wedge ⫰F$ and $⅄F$ by $F \wedge ⅄F$, the rules $F \mapsto ⫰F$, $F \mapsto ⅄F$, etc. would already immensely strengthen the logic. Our soundness proof for the incomplete affine logic given later is another result in a similar direction. It should be remembered that, when it comes to practical applications in the proper sense, the logic that will be used is likely to be far from complete anyway. For example, the popular classical-logic-based systems and programming languages are incomplete, and the reason is not that a complete axiomatization for classical logic is not known, but rather the unfortunate fact of life that often efficiency only comes at the expense of completeness.

But even **CL4**, despite the absence of recurrence operators in it, is already very powerful. Why don't we see a simple example to get the taste of it as a query-solving logic. Let $Acid(x)$ mean "solution x contains acid", and $Red(x)$ mean "litmus paper turns red in solution x". Assume that the knowledge base KB of a **CL4**-based system contains $\forall x(Red(x) \leftrightarrow Acid(x))$ and $\sqcap x(Red(x) \sqcup \neg Red(x))$, accounting for knowledge of the fact that a solution contains acid iff the litmus paper turns red in it, and for availability of a provider who possesses a piece of litmus paper that it can dip into any solution and report the paper's color to the system. Then the system can solve the acidity query $\sqcap x(Acid(x) \sqcup \neg Acid(x))$. This follows from the fact, left as an exercise for the reader to verify, that $\mathbf{CL4} \vdash KB \rightarrow \sqcap x(Acid(x) \sqcup \neg Acid(x))$.

Section 26 of Japaridze (2003) outlines how the context of knowledge base systems can be further extended to systems for planning and action. Roughly, the formal semantics in such applications remains the same, and what changes is only the underlying philosophical assumption that the truth values of predicates and propositions are fixed or predetermined. Rather, those values in CL-based planning systems are viewed as something that interacting agents may be able to manage. That is, predicates or propositions there

stand for *tasks* rather than *facts*. For example *Pregnant*(*Dana*)—or, perhaps, *Impregnate*(*Dana*) instead—can be seen as having no predetermined truth value, with Dana or her mate being in control of whether to make it true or not. And the nonelementary formula $\sqcap x\mathit{Hit}(x)$ describes the task of hitting any one target x selected by the environment/commander/user. Note how naturally resource-consciousness arises here: while $\sqcap x\mathit{Hit}(x)$ is a task accomplishable with one ballistic missile, the stronger task $\sqcap x\mathit{Hit}(x) \wedge \sqcap x\mathit{Hit}(x)$ would require two missiles instead. All of the other operators of CL, too, have natural interpretations as operations on physical (as opposed to just informational) tasks, with $\rightarrow$ acting as a task reduction operation. To get a feel of this, let us look at the task

$$\begin{array}{l} \mathit{Give\ me\ a\ wooden\ stake} \sqcap \mathit{Give\ me\ a\ silver\ bullet} \\ \rightarrow\ \ \mathit{Destroy\ the\ vampire} \sqcap \mathit{Kill\ the\ werewolf}. \end{array}$$

This is a task accomplishable by an agent who has a mallet and a gun as well as sufficient time, energy and bravery to chase and eliminate any one (but not both) of the two monsters, and only needs a wooden stake and/or a silver bullet to complete his noble mission. Then the story told by the legal run $\langle \bot 2.2, \top 1.2\rangle$ of the above game is that the environment asked the agent to kill the werewolf, to which the agent replied by the counterrequest to give him a silver bullet. The task will be considered eventually accomplished by the agent iff he indeed killed the werewolf as long as a silver bullet was indeed given to him.

The fact that CL is a conservative extension of classical logic also makes the former a reasonable and appealing alternative to the latter in its most traditional and unchallenged application areas. In particular, it makes perfect sense to base applied theories—such as, say, Peano arithmetic (axiomatic number theory)—on CL instead of classical logic. Due to conservativity, no old information would be lost or weakened this way. On the other hand, we would get by an order of magnitude more expressive, constructive and computationally meaningful theories than their classical-logic-based versions. Let us see a little more precisely what we mean by a CL-based applied theory. For simplicity, here we restrict our considerations to the cases when the set AX of nonlogical **axioms** of the applied theory is finite. As we did with KB, we identify AX with the $\wedge$-conjunction of its elements. From (the problem represented by) AX—or, equivalently, each conjunct of it—we require to be computable in our sense, i.e. come with an HPM or EPM that solves it. So, notice, all axioms of the old, classical-logic-based version of the theory could be automatically included into the new set AX because they represent true and hence computable elementary problems. Many of those old axioms can be constructivized by, say, replacing blind or parallel operators with their choice equivalents. For example, we would want to rewrite the axiom $\forall x \exists y(y = x + 1)$ of arithmetic as the more informative $\sqcap x \sqcup y(y = x + 1)$. And, of course, to the old axioms or their constructivized versions could be added some essentially new axioms

expressing basic computability principles specific to (the particular interpretation underlying) the theory. Provability (theoremhood) of a formula F in such a theory we understand as provability of the formula $AX \to F$ in the underlying axiomatization of CL which, as in the case of knowledge base systems, is assumed to be uniform-constructively sound. The rule of modus ponens has been shown in Japaridze (2003) (Proposition 21.3)[12] to preserve computability in the following constructive sense:

Theorem 36. *There is an effective function h: $\{EPMs\}\times\{EPMs\} \longrightarrow \{EPMs\}$ such that, for any EPMs $\mathcal{E}$,$\mathcal{C}$, static games A,B and valuation e, if $\mathcal{E} \models_e A$ and $\mathcal{C} \models_e A \to B$, then $h(\mathcal{E},\mathcal{C}) \models_e B$.*

This theorem, together with our assumptions that AX is computable and that the underlying logic is uniform-constructively sound, immediately implies that the problem represented by any theorem F of the applied theory is computable and that, furthermore, a solution for such a problem can be effectively constructed from a proof of F. So, for example, once a formula $\sqcap x \sqcup y\, p(x,y)$ has been proven, we would know that, for every x, a y with $p(x,y)$ not only exists, but can be algorithmically found; furthermore, we would be able to actually construct such an algorithm. Similarly, a reduction—in the sense of Definition 26(4)—of the acceptance problem to the halting problem would automatically come with a proof of $\sqcap x \sqcap y(H(x,y) \sqcup \neg H(x,y)) \to \sqcap x \sqcap y(A(x,y) \sqcup \neg A(x,y))$ in such a theory. Is not this exactly what the constructivists have been calling for?

11.11 Affine logic

Linear logic and its variations such as affine logic have only one group $!, ?$ of exponential operators. The semantics of CL induces at least two equally natural "counterparts" of $!, ?$: the parallel group $⅄, ⅂$ and the branching group $⅋, ⅁$ of recurrence operators. Hence, when rewritten in terms of computability logic, each $(!, ?)$-involving rule of linear logic produces two identical versions: one with $(⅄, ⅂)$ and one with $(⅋, ⅁)$.

Precisely, the language of what we here call **affine logic AL** is obtained from the more expressive language of Section 11.7 by forbidding nonlogical elementary atoms (but not the logical elementary atoms $\top$ and $\bot$), and restricting the operators of the language to $\neg$, $\wedge$, $\vee$, $\sqcap$, $\sqcup$, $⅄$, $⅂$, $⅋$, $⅁$, $\bigsqcap$, $\bigsqcup$. For simplicity, this list does not officially include $\to$ or other definable operators such as $\succ$ and $\circ\!\!-$. If we write $F \to G$, it should be understood as an abbreviation of $\neg F \vee G$. Furthermore, without loss of expressive power, we allow $\neg$ to be applied only to nonlogical atoms, in all other cases understanding $\neg F$ as an abbreviation defined by: $\neg\top = \bot$; $\neg\bot = \top$; $\neg\neg F = F$; $\neg(F_1 \wedge \ldots \wedge F_n) = \neg F_1 \vee \ldots \vee \neg F_n$;

[12] In the official formulation of Proposition 21.3 in Japaridze (2003), the first argument of h was an HPM. In view of Theorem 28, however, replacing "HPM" with "EPM" is perfectly legitimate.

$\neg(F_1 \vee \ldots \vee F_n) = \neg F_1 \wedge \ldots \wedge \neg F_n$; $\neg(F_1 \sqcap \ldots \sqcap F_n) = \neg F_1 \sqcup \ldots \sqcup \neg F_n$; $\neg(F_1 \sqcup \ldots \sqcup F_n) = \neg F_1 \sqcap \ldots \sqcap \neg F_n$; $\neg\curlywedge F = \curlyvee\neg F$; $\neg\curlyvee F = \curlywedge\neg F$; $\neg\multimap F = \circ\!\!-\!\!\!\!\circ\neg F$; $\neg\circ\!\!-\!\!\!\!\circ F = \multimap\neg F$; $\neg\bigsqcap xF = \bigsqcup x\neg F$; $\neg\bigsqcup xF = \bigsqcap x\neg F$. The formulas of this language will be referred to as **AL-formulas**.

Let x be a variable, t a term and $F(x)$ a formula. We say that t is **free for** x **in** $F(x)$ iff none of the free occurrences of x in $F(x)$ is in the scope of Qt for some quantifier Q. Of course, when t is a constant, this condition is always satisfied.

A **sequent** is a nonempty finite sequence of **AL**-formulas. We think of each sequent $F_1, \ldots, F_n$ as the formula $F_1 \vee \ldots \vee F_n$. This allows us to automatically extend the concepts of validity, uniform validity, free occurrence, etc. from formulas to sequents. A formula F is considered **provable** in **AL** iff F, understood as a one-element sequent, is provable.

Deductively logic **AL** is given by the following 16 rules, where: $\underline{G}, \underline{H}$ are arbitrary (possibly empty) sequences of **AL**-formulas; $\curlyvee\underline{G}$ is an arbitrary (possibly empty) sequence of $\curlyvee$-prefixed **AL**-formulas; $\circ\!\!-\!\!\!\!\circ\underline{G}$ is an arbitrary (possibly empty) sequence of $\circ\!\!-\!\!\!\!\circ$-prefixed **AL**-formulas; $n \geq 2$; $1 \leq i \leq n$; x is any variable; E, F, E_1, …, E_n, $E(x)$ are any **AL**-formulas; y is any variable not occurring (whether free or within $\bigsqcap y$ or $\bigsqcup y$) in the conclusion of the rule; and t is any term free for x in $E(x)$.

Identity Axiom:
$$\frac{}{\neg E, E}$$

$\top$-Axiom:
$$\frac{}{\top}$$

Exchange:
$$\frac{\underline{G}, E, F, \underline{H}}{\underline{G}, F, E, \underline{H}}$$

Weakening:
$$\frac{\underline{G}}{\underline{G}, E}$$

$\curlyvee$-Contraction:
$$\frac{\underline{G}, \curlyvee E, \curlyvee E}{\underline{G}, \curlyvee E}$$

$\circ\!\!-\!\!\!\!\circ$-Contraction:
$$\frac{\underline{G}, \circ\!\!-\!\!\!\!\circ E, \circ\!\!-\!\!\!\!\circ E}{\underline{G}, \circ\!\!-\!\!\!\!\circ E}$$

$\sqcup$-Introduction: $$\frac{\underline{G}, E_i}{\underline{G}, E_1 \sqcup \ldots \sqcup E_n}$$

$\sqcap$-Introduction: $$\frac{\underline{G}, E_1 \quad \ldots \quad \underline{G}, E_n}{\underline{G}, E_1 \sqcap \ldots \sqcap E_n}$$

$\vee$-Introduction: $$\frac{\underline{G}, E_1, \ldots, E_n}{\underline{G}, E_1 \vee \ldots \vee E_n}$$

$\wedge$-Introduction: $$\frac{\underline{G_1}, E_1 \quad \ldots \quad \underline{G_n}, E_n}{\underline{G_1}, \ldots, \underline{G_n}, E_1 \wedge \ldots \wedge E_n}$$

$\curlyvee$-Introduction: $$\frac{\underline{G}, E}{\underline{G}, \curlyvee E}$$

$\wr$-Introduction: $$\frac{\underline{G}, E}{\underline{G}, \wr E}$$

$\curlywedge$-Introduction: $$\frac{\curlyvee \underline{G}, E}{\curlyvee \underline{G}, \curlywedge E}$$

$\backsim$-Introduction: $$\frac{\wr \underline{G}, E}{\wr \underline{G}, \backsim E}$$

$\bigsqcup$-Introduction: $$\frac{\underline{G}, E(t)}{\underline{G}, \bigsqcup x E(x)}$$

$\bigsqcap$-Introduction: $$\frac{\underline{G}, E(y)}{\underline{G}, \bigsqcap x E(x)}$$

Unlike any other results that we have surveyed so far, the soundness and completeness of affine logic, while claimed already in Japaridze (2003), has never been officially proven. For this reason, the following theorem comes with a full proof, to which most of the remaining part of this paper is devoted.

Theorem 37. *If* **AL** $\vdash S$*, then* $\Vdash S$ (*any sequent* S). *Furthermore:*

Uniform-Constructive Soundness: *There is an effective procedure that takes any* **AL***-proof of any sequent* S *and constructs a uniform solution for* S.

As mentioned earlier, a similar (uniform-constructive) soundness theorem for Heyting's intuitionistic calculus has been proven in Japaridze (2006b), with intuitionistic implication understood as $\circ\!\!-$, and intuitionistic conjunction, disjunction and quantifiers as $\sqcap, \sqcup, \bigsqcap, \bigsqcup$.

11.12 Soundness proof for affine logic

This technical section is devoted to a proof of Theorem 37. It also contains a number of useful lemmas that could be employed in other proofs.

11.12.1 CL4-derived validity lemmas

In our proof of Theorem 37 we will need a number of lemmas concerning uniform validity of certain formulas. Some of such validity proofs will be given directly in Sections 11.12.3 and 11.12.4. But some proofs come "for free", based on the already known soundness of **CL4**. In fact, here we will only exploit the propositional fragment **CL2** of **CL4**. The former is obtained from the latter by mechanically restricting its language to 0-ary letters, and disallowing the (now meaningless) usage of quantifiers. Let us call the formulas of such a language **CL2-formulas**. Restricting the language to **CL2**-formulas simplifies the formulation of **CL4**: Rule **B2** disappears, and so does clause (ii) of Rule **A**. **CL4** is a conservative extension of **CL2**, so, for a **CL2**-formula F, it does not matter whether we say **CL4** $\vdash F$ or **CL2** $\vdash F$.

In Section 11.11 we started using the notation $\underline{G}$ for sequences of formulas. We also agreed to identify sequences of formulas with $\vee$-disjunctions of those formulas. So, from now on, an underlined expression such as $\underline{G}$ will mean an arbitrary formula $G_1 \vee \ldots \vee G_n$ for some $n \geq 0$. The expression $\underline{G} \vee E$ should be read as $G_1 \vee \ldots \vee G_n \vee E$ rather than as $(G_1 \vee \ldots \vee G_n) \vee E$. The number of disjuncts in $\underline{G}$ may be empty. When this is a possibility, $\underline{G}$ will usually occur as a disjunct within a bigger expression such as $\underline{G} \vee E$ or $\underline{G} \rightarrow E$, both of which simply mean E.

As we agreed that p, q, r, s (with no tuples of terms attached) stand for nonlogical elementary atoms and P, Q, R, S for general atoms, $\underline{p}, \underline{q}, \underline{r}, \underline{s}$ will correspondingly stand for $\vee$-disjunctions of elementary atoms, and $\underline{P}, \underline{Q}, \underline{R}, \underline{S}$ for $\vee$-disjunctions of general atoms.

We will also be underlining complex expressions such as $G \rightarrow H$, $\bigsqcup x G(x)$ or $⫯G$. $\underline{G \rightarrow H}$ should be understood as $(G_1 \rightarrow H_1) \vee \ldots \vee (G_n \rightarrow H_n)$, $\underline{\bigsqcup x G(x)}$ as $\bigsqcup x G_1(x) \vee \ldots \vee \bigsqcup x G_n(x)$ (note that only the G_i vary but not x), $\underline{⫯G}$ as $⫯G_1 \vee \ldots \vee ⫯G_n$, $⫯\underline{⫯G}$ as $⫯(⫯G_1 \vee \ldots \vee ⫯G_n)$, etc.

A **CL2**-formula E is said to be **general-base** iff it does not contain any elementary atoms. A **substitution** is a function σ that sends every general atom P of the language of **CL2** to a (not necessarily **CL2**-) formula $\sigma(P)$. This mapping extends to all general-base **CL2**-formulas by stipulating that σ commutes with each operator: $\sigma(\neg E) = \neg\sigma(E)$, $\sigma(E_1 \sqcap \ldots \sqcap E_k) = \sigma(E_1) \sqcap \ldots \sqcap \sigma(E_k)$, etc. We say that a formula F is a **substitutional instance** of a general-base **CL2**-formula E iff $F = \sigma(E)$ for some substitution σ. Thus, "F is a substitutional instance of E" means that F has the same form as E.

In the following lemma, we assume $n \geq 2$, and $1 \leq i \leq n$. Note that the expressions given in clauses (d)–(k) are schemata of formulas rather than formulas, for the lengths of their underlined expressions, as well as i and n, may vary. Strictly speaking, the expressions of clauses (a)–(c) are so as well, because P, Q, R are metavariables for general atoms rather than particular general atoms.

Lemma 38. *All substitutional instances of all* **CL2***-formulas given by the following schemata are uniformly valid. Furthermore, there is an effective procedure that takes any particular formula matching a given scheme and constructs an EPM that is a uniform solution for all substitutional instances of that formula.*

(a) $\neg P \vee P$;
(b) $P \vee Q \to Q \vee P$;
(c) $(P \to Q) \wedge (Q \to R) \to (P \to R)$;
(d) $(\underline{Q_1} \vee P_1) \wedge \ldots \wedge (\underline{Q_n} \vee P_n) \to \underline{Q_1} \vee \ldots \vee \underline{Q_n} \vee (P_1 \wedge \ldots \wedge P_n)$;
(e) $(\underline{P \to Q}) \to (\underline{R} \vee \underline{P} \vee \underline{S} \to \underline{R} \vee \underline{Q} \vee \underline{S})$;
(f) $\underline{Q} \vee \underline{R} \vee \underline{S} \to \underline{Q} \vee (\underline{R}) \vee \underline{S}$;
(g) $\underline{Q} \vee (\underline{R}) \vee \underline{S} \to \underline{Q} \vee \underline{R} \vee \underline{S}$;
(h) $(P_1 \wedge P_2 \wedge \ldots \wedge P_n \to Q) \to (P_1 \to (P_2 \to \ldots (P_n \to Q)\ldots))$;
(i) $\underline{Q} \to \underline{Q} \vee P$;
(j) $\underline{P_i} \to \underline{P_1 \sqcup \ldots \sqcup P_n}$;
(k) $(\underline{Q} \vee P_1) \wedge \ldots \wedge (\underline{Q} \vee P_n) \to \underline{Q} \vee (P_1 \sqcap \ldots \sqcap P_n)$.

Proof. In order to prove this lemma, it would be sufficient to show that all formulas given by the above schemata are provable in **CL4** (in fact, **CL2**). Indeed, if we succeed in doing so, then an effective procedure whose existence is claimed in the present lemma could be designed to work as follows. First, the procedure finds a **CL4**-proof of a given formula E. Then, based on that proof and using the procedure whose existence is stated in Theorem 34, it finds a uniform solution $\mathcal{E}$ for that formula. It is not hard to see that the same $\mathcal{E}$ will automatically be a uniform solution for every substitutional instance of E as well. So, now it remains to do the simple syntactic exercise of checking **CL4**-provabilities for each clause of the lemma.

Notice that every formula E given by one of the clauses (a)–(h) has—more precisely, we may assume that it has—exactly two, one negative and one positive, occurrences of each (general) atom, with all occurrences being surface ones. For such an E, let E' be the result of rewriting each general atom P of E into a nonlogical elementary atom p in such a way that different general atoms are rewritten as different elementary atoms. Then E follows from E' in **CL4** by a series of applications of Rule **C**, specifically, as many applications as the number of different atoms of E. In turn, observe that for each of the clauses (a)–(h), the formula E' would be a classical tautology. Hence E' follows from the empty set of premises by Rule **A**. Thus, $\mathbf{CL4} \vdash E$.

For clause (i), let $\underline{q}$ be the result of replacing in $\underline{Q}$ all atoms by pairwise distinct nonlogical elementary atoms. The formula $\underline{q} \rightarrow \underline{q} \vee P$ is stable and choice-operator-free, so it follows from {} by Rule **A**. From the latter, applying Rule **C** as many times as the number of disjuncts in $\underline{Q}$, we obtain the desired $\underline{Q} \rightarrow \underline{Q} \vee P$.

For clause (j), the following is a **CL4**-proof of the corresponding formula(s):

1. $p_i \rightarrow p_i$ (from {} by Rule **A**);
2. $P_i \rightarrow P_i$ (from 1 by Rule **C**);
3. $P_i \rightarrow P_1 \sqcup \ldots \sqcup P_n$ (from 2 by Rule **B1**).

For clause (k), note that $(\underline{Q} \vee P_1) \wedge \ldots \wedge (\underline{Q} \vee P_n) \rightarrow \underline{Q} \vee (P_1 \sqcap \ldots \sqcap P_n)$ is stable. Hence it follows by Rule **A** from n premises, where each premise is $(\underline{Q} \vee P_1) \wedge \ldots \wedge (\underline{Q} \vee P_n) \rightarrow \underline{Q} \vee P_i$ for one of the $i \in \{1, \ldots, n\}$. Each such formula, in turn, can be obtained by a series of applications of Rule **C** from

$$(\underline{Q} \vee P_1) \wedge \ldots \wedge (\underline{Q} \vee P_{i-1}) \wedge (\underline{q} \vee p_i) \wedge (\underline{Q} \vee P_{i+1}) \wedge \ldots \wedge (\underline{Q} \vee P_n) \rightarrow \underline{q} \vee p_i,$$

where p_i is an elementary nonlogical atom and $\underline{q}$ is obtained from $\underline{Q}$ by replacing its general atoms by pairwise distinct (and distinct from p_i) elementary nonlogical atoms. In turn, the above formula can be seen to be stable and hence, as it does not contain choice operators, derivable from the empty set of premises by Rule **A**. □

11.12.2 Closure lemmas

In this section we let n range over positive integers, x over any variables, E, F, G (possibly with subscripts) over any **AL**-formulas, and $\mathcal{E}, C, \mathcal{D}$ (possibly with subscripts) over any EPMs. Unless otherwise specified, in each context these metavariables are assumed to be universally quantified.

First two of the following three lemmas have been proven in Section 21 of Japaridze (2003). Here we provide a proof only for the third, never officially proven, one.

Lemma 39. *For any static game A, if $\models A$, then $\models \diamond A$.*

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ such that, for any EPM $\mathcal{E}$, static game A and valuation e, if $\mathcal{E} \models_e A$, then $h(\mathcal{E}) \models_e \diamond A$.

Lemma 40. *For any static game A, if $\models A$, then $\models \sqcap xA$.*

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ such that, for any EPM $\mathcal{E}$ and static game A, if $\mathcal{E} \models A$, then $h(\mathcal{E}) \models \sqcap xA$.

Lemma 41. *For any static game A, if $\models A$, then $\models \barwedge A$.*

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ such that, for any EPM $\mathcal{E}$, static game A and valuation e, if $\mathcal{E} \models_e A$, then $h(\mathcal{E}) \models_e \barwedge A$.

Proof. Intuitively the idea here is simple: if we (machine $\mathcal{E}$) know how to win A, then, applying the same strategy to each conjunct separately, we (machine $h(\mathcal{E})$) can win the infinite conjunction $\barwedge A = A \wedge A \wedge A \wedge \ldots$ as well.

To give a detailed description of the machine $h(\mathcal{E})$ that materializes this idea, we need some preliminaries. Remember the e-successor relation between HPM configurations from Section 11.6. In the context of a fixed HPM $\mathcal{H}$, valuation e and configuration C, the transition from C to a successor (e-successor) configuration C' is nondeterministic because it depends on the sequence Ψ of the moves (labeled with $\bot$) made by the environment while the machine was in configuration C. Once such a Ψ is known, however, the value of C' becomes determined and can be calculated from C, (the relevant finite part of) e and (the transition function of) $\mathcal{H}$. We call the e-successor of C uniquely determined by such Ψ the (e, Ψ)**-successor** of C (in $\mathcal{H}$).

On the way of constructing the EPM $h(\mathcal{E})$, we first turn $\mathcal{E}$ into an HPM $\mathcal{H}$ such that, for every static game A and valuation e, $\mathcal{H} \models_e A$ whenever $\mathcal{E} \models_e A$. According to Theorem 28, such an $\mathcal{H}$ can be constructed effectively. Now, using $\mathcal{H}$, we define $h(\mathcal{E})$ to be the EPM which, with a valuation e spelled on its valuation tape, acts as follows. Its work consists in iterating the following procedure ITERATION(k) infinitely many times, starting from $k = 1$ and incrementing k by one at every new step. During each ITERATION(k) step, $h(\mathcal{E})$ maintains $k - 1$ records $C_1, \ldots, C_{k-1}$ and creates one new record C_k, with each such C_i holding a certain configuration of $\mathcal{H}$. Here is how ITERATION(k) proceeds:

Procedure ITERATION(k):
1. Grant permission. Let $\Psi = \langle \bot\alpha \rangle$ if the adversary responds by a move α, and $\Psi = \langle\rangle$ if there is no response.
2. For $i = 1$ to $i = k - 1$, do the following:
   (a) If $\mathcal{H}$ makes a move β in configuration C_i, make the move $i.\beta$;
   (b) Update C_i to the $(e, \Psi^{i.})$-successor[13] of C_i.

[13] For $\Psi^{i.}$, remember the notation Γ^{α} from page 273.

3. Let C be the initial configuration of $\mathcal{H}$, and Φ the position currently spelled on the run tape.

(a) If $\mathcal{H}$ makes a move β in configuration C, make the move $k.\beta$.

(b) Create the record C_k and initialize it to the $(e, \Phi^{k.})$-successor of C.

Obviously (the description of) $h(\mathcal{E})$ can be effectively obtained from $\mathcal{H}$ and hence from $\mathcal{E}$, so that, as promised, h is indeed an effective function. What remains to verify is that, whenever $\mathcal{E}$ wins a static game A on a valuation e, we have $h(\mathcal{E}) \models_e \backslash A$. Consider any such A and e, and suppose $h(\mathcal{E}) \not\models_e \backslash A$. We want to show that then $\mathcal{E} \not\models_e A$. Let B be an arbitrary e-computation branch of $h(\mathcal{E})$, and let Γ be the run spelled by B. Permission is granted at the beginning of each of the infinitely many routines ITERATION(k), so B is fair. Therefore, $h(\mathcal{E}) \not\models_e A$ simply means that $\mathbf{Wn}_e^{\backslash A}\langle\Gamma\rangle = \bot$. The latter, in turn, implies that for some $n \in \{1, 2, 3, \ldots\}$, $\mathbf{Wn}_e^{A}\langle\Gamma^{n.}\rangle = \bot$. This can be easily seen from the fact that every move that $h(\mathcal{E})$ makes starts with an 'n.' for some n. But an analysis of the procedure followed by $h(\mathcal{E})$ can convince us that $\Gamma^{n.}$ is the run spelled by some e-computation branch of $\mathcal{H}$. This means that $\mathcal{H} \not\models_e A$. Remembering that $\mathcal{H} \models_e A$ whenever $\mathcal{E} \models_e A$, we find that $\mathcal{E} \not\models_e A$. □

Lemma 42. *If* $\Vdash E$, *then* $\Vdash \backslash E$.

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}$ *and* E, *if* $\mathcal{E} \Vdash E$, *then* $h(\mathcal{E}) \Vdash \backslash E$.

Proof. As Lemma 41 asserts (or rather implies), there is an effective function $h : \{\text{EPMs}\} \longrightarrow \{\text{EPMs}\}$ such that, for any EPM $\mathcal{E}$ and any static game A, if $\mathcal{E} \models A$, then $h(\mathcal{E}) \models \backslash A$. We now claim for that very function h that, if $\mathcal{E} \Vdash E$, then $h(\mathcal{E}) \Vdash \backslash E$. Indeed, assume $\mathcal{E} \Vdash E$. Consider any $\backslash E$-admissible interpretation *. Of course, the same interpretation is also E-admissible. Hence, $\mathcal{E} \Vdash E$ implies $\mathcal{E} \models E^*$. But then, by the known behavior of h, we have $h(\mathcal{E}) \models \backslash E^*$. Since * was arbitrary, we conclude that $h(\mathcal{E}) \Vdash \backslash E$. □

Lemma 43. *If* $\Vdash E$, *then* $\Vdash \diamond E$.

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}$ *and* E, *if* $\mathcal{E} \Vdash E$, *then* $h(\mathcal{E}) \Vdash \diamond E$.

Proof. Similar to Lemma 42, only use Lemma 39 instead of Lemma 41. □

Lemma 44. *If* $\Vdash E$, *then* $\Vdash \sqcap xE$.

Moreover, there is an effective function $h : \{EPMs\} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}$, x *and* E, *if* $\mathcal{E} \Vdash E$, *then* $h(\mathcal{E}) \Vdash \sqcap xE$.

Proof. Similar to Lemma 42, only use Lemma 40 instead of Lemma 41. □

Lemma 45. (Modus ponens) *If* $\Vdash F$ *and* $\Vdash F \rightarrow E$, *then* $\Vdash E$.

Moreover, there is an effective function $h : \{EPMs\} \times \{EPMs\} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}$, $\mathcal{C}$, F *and* E, *if* $\mathcal{E} \Vdash F$ *and* $\mathcal{C} \Vdash F \rightarrow E$, *then* $h(\mathcal{E}, \mathcal{C}) \Vdash E$.

Proof. According to Theorem 36, there is an effective function h: {EPMs}× {EPMs} → {EPMs} such that, for any static games A,B, valuation e and EPMs $\mathcal{E}$ and C,

$$\text{if } \mathcal{E} \models_e A \text{ and } C \models_e A \to B, \text{ then } h(\mathcal{E}, C) \models_e B. \tag{7}$$

We claim that such a function h also behaves as our lemma states. To see this, assume $\mathcal{E} \Vdash F$ and $C \Vdash F \to E$, and consider an arbitrary valuation e and an arbitrary E-admissible interpretation *. Our goals is to show that $h(\mathcal{E}, C) \models_e E^*$, which obviously means the same as

$$h(\mathcal{E}, C) \models_e e[E^*]. \tag{8}$$

We define the new interpretation † by stipulating that, for every k-ary letter L with $L^* = L^*(x_1, \ldots, x_k)$, $L^\dagger$ is the game $L^\dagger(x_1, \ldots, x_k)$ such that, for any tuple $c_1, \ldots, c_k$ of constants,

$$L^\dagger(c_1, \ldots, c_k) = e[L^*(c_1, \ldots, c_k)].$$

Unlike $L^*(x_1, \ldots, x_k)$ that may depend on some "hidden" variables (those that are not among $x_1, \ldots, x_k$), obviously $L^\dagger(x_1, \ldots, x_k)$ does not depend on any variables other that $x_1, \ldots, x_k$. This makes † admissible for any **AL**-formula, including F and $F \to E$. Then our assumptions $\mathcal{E} \Vdash F$ and $C \Vdash F \to E$ imply $\mathcal{E} \models_e F^\dagger$ and $C \models_e F^\dagger \to E^\dagger$. Consequently, by (7), $h(\mathcal{E}, C) \models_e E^\dagger$, i.e. $h(\mathcal{E}, C) \models_e e[E^\dagger]$. Now, with some thought, we can see that $e[E^\dagger] = e[E^*]$. Hence (8) is true. □

Lemma 46. (Generalized modus ponens) *If* $\Vdash F_1, \ldots, \Vdash F_n$ *and* $\Vdash F_1 \wedge \ldots \wedge F_n \to E$, *then* $\Vdash E$.

Moreover, there is an effective function $h : \{EPMs\}^{n+1} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}_1, \ldots, \mathcal{E}_n, C, F_1, \ldots, F_n, E$, *if* $\mathcal{E}_1 \Vdash F_1, \ldots, \mathcal{E}_n \Vdash F_n$ *and* $C \Vdash F_1 \wedge \ldots \wedge F_n \to E$, *then* $h(\mathcal{E}_1, \ldots, \mathcal{E}_n, C) \Vdash E$. *Such a function, in turn, can be effectively constructed for each particular n.*

Proof. In case $n = 1$, h is the function whose existence is stated in Lemma 45. Below we will construct h for the case $n = 2$. It should be clear how to generalize that construction to any greater n.

Assume $\mathcal{E}_1 \Vdash F_1$, $\mathcal{E}_2 \Vdash F_2$ and $C \Vdash F_1 \wedge F_2 \to E$. By Lemma 38(h), the formula $(F_1 \wedge F_2 \to E) \to (F_1 \to (F_2 \to E))$ has a uniform solution. Lemma 45 allows us to combine that solution with C and find a uniform solution $\mathcal{D}_1$ for $F_1 \to (F_2 \to E)$. Now applying Lemma 45 to $\mathcal{E}_1$ and $\mathcal{D}_1$, we find a uniform solution $\mathcal{D}_2$ for $F_2 \to E$. Finally, applying the same lemma to $\mathcal{E}_2$ and $\mathcal{D}_2$, we find a uniform solution $\mathcal{D}$ for E. Note that $\mathcal{D}$ does not depend on F_1, F_2, E, and that we constructed $\mathcal{D}$ in an effective way from the arbitrary $\mathcal{E}_1, \mathcal{E}_2$ and C. Formalizing this construction yields the function h whose existence is claimed by our present lemma. □

Lemma 47. (Transitivity) *If* $\Vdash F \to E$ *and* $\Vdash E \to G$, *then* $\Vdash F \to G$.

Moreover, there is an effective function $h : \{EPMs\} \times \{EPMs\} \longrightarrow \{EPMs\}$ *such that, for any* $\mathcal{E}_1$, $\mathcal{E}_2$, F, E *and* G, *if* $\mathcal{E}_1 \Vdash F \to E$ *and* $\mathcal{E}_2 \Vdash E \to G$, *then* $h(\mathcal{E}_1, \mathcal{E}_2) \Vdash F \to G$.

Proof. Assume $\mathcal{E}_1 \Vdash F \to E$ and $\mathcal{E}_2 \Vdash E \to G$. By Lemma 38(c), we also have $C \Vdash (F \to E) \wedge (E \to G) \to (F \to G)$ for some C. Lemma 46 allows us to combine the three uniform solutions and construct a uniform solution $\mathcal{D}$ for $F \to G$. Formalizing this construction yields the function h whose existence is claimed by our lemma. □

11.12.3 More validity lemmas

In this section we will prove a number of winnability facts by describing winning strategies in terms of EPMs. When trying to show that a given EPM wins a given game, it is always safe to assume that the runs that the machine generates are never $\bot$-illegal, i.e. that the environment never makes a (first) illegal move, for if it does, the machine automatically wins. This assumption, that we call the **clean environment assumption**, will always be explicitly or implicitly present in our winnability proofs.

We will often employ a uniform solution for $P \to P$ called the **copy-cat strategy** ($CC\mathcal{S}$). This strategy, sketched for $\neg Chess \vee Chess$ in Section 11.4.3, consists in mimicking, in the antecedent, the moves made by the environment in the consequent, and vice versa. More formally, the algorithm that $CC\mathcal{S}$ follows is an infinite loop, on every iteration of which $CC\mathcal{S}$ keeps granting permission until the environment makes a move $1.\alpha$ (resp. $2.\alpha$), to which the machine responds by the move $2.\alpha$ (resp. $1.\alpha$). As shown in the proof of Proposition 22.1 of Japaridze (2003), this strategy guarantees success in every game of the form $A \vee \neg A$ and hence $A \to A$. A perhaps important detail is that $CC\mathcal{S}$ never looks at the past history of the game, i.e. the movement of its scanning head on the run tape is exclusively left-to-right. This guarantees that, even if the original game was something else and it only evolved to $A \to A$ later as a result of making a series of moves, switching to $CC\mathcal{S}$ after the game has been brought down to $A \to A$ ensures success no matter what happened in the past.

Throughout this section, E and F (possibly with indices and attached tuples of variables) range over **AL**-formulas, x and y over variables, t over terms, n, k, i, j over nonnegative integers, w over bitstrings, and α, γ over moves. These metavariables are assumed to be universally quantified in each context unless otherwise specified. In accordance with our earlier convention, ϵ stands for the empty bitstring.

Next, * always means an arbitrary but fixed interpretation admissible for the formula whose uniform validity we are trying to prove. For readability, we

will sometimes omit this parameter and write, say, E instead of E^*. From the context it will be usually clear whether "E" stands for the formula E or the game E^*. Similarly, in our winnability proofs e will always stand for an arbitrary but fixed valuation—specifically, the valuation spelled on the valuation tape of the machine under question. Again, for readability, we will typically omit e when it is irrelevant, and write E^* (or just E) instead of $e[E^*]$.

Lemma 48. $\Vdash E \to \curlyvee E$.

Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} \Vdash E \to \curlyvee E$.

Proof. The idea of a uniform solution $\mathcal{E}$ for $E \to \curlyvee E$ is simple: seeing the consequent as the infinite disjunction $E \vee E \vee E \vee \ldots\ldots$, ignore all of its disjuncts but the first one, and play $E \to \curlyvee E$ as if it was just $E \to E$.

In more precise terms, $\mathcal{E}$ follows the following procedure LOOP which, notice, only differs from *CCS* in that the move prefix '2.' is replaced by '2.1.':

Procedure LOOP: Keep granting permission until the environment makes a move '$1.\alpha$' or '$2.1.\alpha$'; in the former case respond by the move '$2.1.\alpha$', and in the latter case respond by the move '$1.\alpha$'; then repeat LOOP.

Consider an arbitrary e-computation branch B of $\mathcal{E}$, and let Θ be the run spelled by B. Obviously permission is granted infinitely many times in B, so B is fair. Hence, in order to show that $\mathcal{E}$ wins $E^* \to \curlyvee E^*$ (on the irrelevant valuation e which we, according to our conventions, are omitting and pretend that E^* is a constant game so that $e[E^*] = E$), it would suffice to show that $\mathbf{Wn}^{E^* \to \curlyvee E^*}\langle\Theta\rangle = \top$.

Let Θ_i denote the initial segment of Θ consisting of the (lab)moves made by the players by the beginning of the ith iteration of LOOP in B (if such an iteration exists). By induction on i, based on the clean environment assumption and applying a routine analysis of the behavior of LOOP, one can easily find that

$$\begin{array}{l} \text{(a) } \Theta_i \in \mathbf{Lr}^{E^* \to \curlyvee E^*}; \\ \text{(b) } \neg\Theta_i^{1.} = \Theta_i^{2.1.}. \end{array} \tag{9}$$

If LOOP is iterated infinitely many times, then the above obviously extends from Θ_i to Θ, because every initial segment of Θ is an initial segment of some Θ_i, and similarly for $\Theta^{1.}$ and $\Theta^{2.1.}$. Suppose now LOOP is iterated only a finite number m of times. Then $\Theta = \langle\Theta_m, \Gamma\rangle$, where Γ entirely consists of $\bot$-labeled moves none of which has the prefix '1.' or '2.1'. This is so because the environment cannot make a move $1.\alpha$ or $2.1.\alpha$ during the mth iteration (otherwise there would be a next iteration) and, since $\mathcal{E}$'s moves are only triggered by the above two sorts of moves, $\mathcal{E}$ does not move during the mth iteration of LOOP. But then, in view of the clean environment assumption, Θ inherits condition (a) of (9) from Θ_m, because there are no $\top$-labeled moves in Γ; and the same is the case with condition (b), because $\langle\Theta_m, \Gamma\rangle^{1.} = \Theta_m^{1.}$ and $\langle\Theta_m, \Gamma\rangle^{2.1.} = \Theta_m^{2.1.}$.

Thus, no matter whether LOOP is iterated a finite or infinite number of times, we have:

$$\begin{array}{l}\text{(a) } \Theta \in \mathbf{Lr}^{E^*\rightarrow \curlyvee E^*};\\ \text{(b) } \neg\Theta^{1\cdot} = \Theta^{2.1\cdot}.\end{array} \tag{10}$$

Since $\Theta \in \mathbf{Lr}^{E^*\rightarrow \curlyvee E^*}$, in order to show that $\mathbf{Wn}^{E^*\rightarrow \curlyvee E^*}\langle\Theta\rangle = \top$, i.e. show that $\mathbf{Wn}^{\neg E^*\vee \curlyvee E^*}\langle\Theta\rangle = \top$, by the definition of $\vee$, it would suffice to verify that either $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1\cdot}\rangle = \top$ or $\mathbf{Wn}^{\curlyvee E^*}\langle\Theta^{2\cdot}\rangle = \top$. So, assume $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1\cdot}\rangle \neq \top$, i.e. $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1\cdot}\rangle = \bot$, i.e. $\mathbf{Wn}^{E^*}\langle\neg\Theta^{1\cdot}\rangle = \top$. Then, by clause (b) of (10), $\mathbf{Wn}^{E^*}\langle\Theta^{2.1\cdot}\rangle = \top$. But then, by the definition of $\curlyvee$, $\mathbf{Wn}^{\curlyvee E^*}\langle\Theta^{2\cdot}\rangle = \top$.

Thus, $\mathcal{E} \models E^* \rightarrow \curlyvee E^*$. Since * was arbitrary and the work of $\mathcal{E}$ did not depend on it, we conclude that $\mathcal{E} \Vdash E \rightarrow \curlyvee E$. □

Lemma 49. $\Vdash E \rightarrow \curlywedge E$.

Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} \Vdash E \rightarrow \curlywedge E$.

Proof. Again, the idea of a uniform solution $\mathcal{E}$ for $E \rightarrow \curlywedge E$ is simple: just act as $CC\mathcal{S}$, never making any replicative moves in the consequent and pretending that the latter is E rather than (the easier-to-win) $\curlywedge E$. The following formal description of the interactive algorithm that $\mathcal{E}$ follows is virtually the same as that of $CC\mathcal{S}$, with the only difference that the move prefix '2.' is replaced by '$2.\epsilon$.'.

Procedure LOOP: Keep granting permission until the environment makes a move '$1.\alpha$' or '$2.\epsilon.\alpha$'; in the former case respond by the move '$2.\epsilon.\alpha$', and in the latter case respond by the move '$1.\alpha$'; then repeat LOOP.

Consider an arbitrary e-computation branch B of $\mathcal{E}$. Let Θ be the run spelled by B. As in the proof of the previous lemma, clearly permission will be granted infinitely many times in B, so this branch is fair. Hence, in order to show that $\mathcal{E}$ wins the game, it would suffice to show that $\mathbf{Wn}^{E^*\rightarrow \curlywedge E^*}\langle\Theta\rangle = \top$.

Let Θ_i denote the initial segment of Θ consisting of the (lab)moves made by the players by the beginning of the ith iteration of LOOP in B. By induction on i, based on the clean environment assumption and applying a routine analysis of the behavior of LOOP and the definitions of the relevant game operations, one can easily find that

(a) $\Theta_i \in \mathbf{Lr}^{E^*\rightarrow \curlywedge E^*}$;
(b) $\neg\Theta_i^{1\cdot} = \Theta_i^{2.\epsilon\cdot}$;
(c) *All of the moves in $\Theta_i^{2\cdot}$ have the prefix '$\epsilon.$'.*

If LOOP is iterated infinitely many times, then the above obviously extends from Θ_i to Θ, because every initial segment of Θ is an initial segment of some Θ_i, and similarly for $\Theta^{1\cdot}$ and $\Theta^{2.\epsilon\cdot}$. And if LOOP is iterated only a finite number

m of times, then $\Theta = \Theta_m$. This is so because the environment cannot make a move $1.\alpha$ or $2.\epsilon.\alpha$ during the mth iteration (otherwise there would be a next iteration), and any other move would violate the clean environment assumption; and, as long as the environment does not move during a given iteration, neither does the machine. Thus, no matter whether LOOP is iterated a finite or infinite number of times, we have:

$$\begin{array}{l}\text{(a) } \Theta \in \mathbf{Lr}^{E^*\to{}^{\mathsf{o}}E^*};\\ \text{(b) } \neg\Theta^{1.} = \Theta^{2.\epsilon.};\\ \text{(c) } \textit{All of the moves in } \Theta^{2.} \textit{ have the prefix `}\epsilon.\textit{'.}\end{array} \tag{11}$$

Since $\Theta \in \mathbf{Lr}^{E^*\to{}^{\mathsf{o}}E^*}$, in order to show that $\mathbf{Wn}^{E^*\to{}^{\mathsf{o}}E^*}\langle\Theta\rangle = \top$, it would suffice to verify that either $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1.}\rangle = \top$ or $\mathbf{Wn}^{{}^{\mathsf{o}}E^*}\langle\Theta^{2.}\rangle = \top$. So, assume $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1.}\rangle \neq \top$, i.e. $\mathbf{Wn}^{\neg E^*}\langle\Theta^{1.}\rangle = \bot$, i.e. $\mathbf{Wn}^{E^*}\langle\neg\Theta^{1.}\rangle = \top$. Then, by clause (b) of (11), $\mathbf{Wn}^{E^*}\langle\Theta^{2.\epsilon.}\rangle = \top$. Pick any complete branch w of $\mathit{Tree}^{{}^{\mathsf{o}}E^*}\langle\Theta^{2.}\rangle$. In view of clause (c) of (11), we obviously have $\Theta^{2.\epsilon.} = (\Theta^{2.})^{\preceq w}$ (in fact, $w = \epsilon$). Hence $\mathbf{Wn}^{E^*}\langle(\Theta^{2.})^{\preceq w}\rangle = \top$. Then, by the definition of ${}^{\mathsf{o}}$, $\mathbf{Wn}^{{}^{\mathsf{o}}E^*}\langle\Theta^{2.}\rangle = \top$.

Thus, $\mathcal{E} \models E^* \to {}^{\mathsf{o}}E^*$ and, as * was arbitrary and the work of $\mathcal{E}$ did not depend on it, we conclude that $\mathcal{E} \Vdash E \to {}^{\mathsf{o}}E$. □

Having already seen two examples, in the remaining uniform validity proofs we will typically limit ourselves to just constructing interactive algorithms, leaving routine verification of their correctness to the reader. An exception will be the proof of Lemma 57 given separately in Section 11.12.4 where, due to the special complexity of the case, correctness verification will be done even more rigorously than we did this in the proofs of Lemmas 48 and 49.

Lemma 50. $\Vdash {\downarrow}(E \to F) \to ({\downarrow}E \to {\downarrow}F)$.

Moreover, there is an EPM $\mathcal{E}$ such that, for every E and F,

$$\mathcal{E} \Vdash {\downarrow}(E \to F) \to ({\downarrow}E \to {\downarrow}F).$$

Proof. Here is a strategy for $\mathcal{E}$ to follow:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then:

If $\gamma = 1.i.1.\alpha$ (resp. $\gamma = 2.1.i.\alpha$), then make the move $2.1.i.\alpha$ (resp. $1.i.1.\alpha$), and repeat LOOP;

If $\gamma = 1.i.2.\alpha$ (resp. $\gamma = 2.2.i.\alpha$), then make the move $2.2.i.\alpha$ (resp. $1.i.2.\alpha$), and repeat LOOP. □

Lemma 51. $\Vdash {\downarrow\!\circ}(E \to F) \to ({\downarrow\!\circ}E \to {\downarrow\!\circ}F)$.

Moreover, there is an EPM $\mathcal{E}$ such that, for every E and F,

$$\mathcal{E} \Vdash {\downarrow\!\circ}(E \to F) \to ({\downarrow\!\circ}E \to {\downarrow\!\circ}F).$$

Proof. A relaxed description of a uniform solution $\mathcal{E}$ for $⫰(E \to F) \to (⫰E \to ⫰F)$ is as follows. In $⫰(E^* \to F^*)$ and $⫰E^*$ the machine is making exactly the same replicative moves (moves of the form w:) as the environment is making in $⫰F^*$. This ensures that the underlying BT-structures of the three $⫰$-components of the game stay identical, and now all the machine needs for a success is to win the game $(E^* \to F^*) \to (E^* \to F^*)$ within each branch of those trees. This can be easily achieved by applying copy-cat methods to the two occurrences of E and the two occurrences of F.

In precise terms, the strategy that $\mathcal{E}$ follows is described by the following interactive algorithm.

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then:

If $\gamma = 2.2.w$:, then make the moves $1.w$: and $2.1.w$:, and repeat LOOP;

If $\gamma = 2.2.w.\alpha$ (resp. $\gamma = 1.w.2.\alpha$), then make the move $1.w.2.\alpha$ (resp. $2.2.w.\alpha$), and repeat LOOP;

If $\gamma = 2.1.w.\alpha$ (resp. $\gamma = 1.w.1.\alpha$), then make the move $1.w.1.\alpha$ (resp. $2.1.w.\alpha$), and repeat LOOP. □

Lemma 52. $⊩ ⅄(E_1 \vee \ldots \vee E_n) \to ⅄E_1 \vee \ldots \vee ⅄E_n$.

Moreover, there is an effective procedure that takes any particular value of n and constructs an EPM $\mathcal{E}$ such that, for any $E_1, \ldots, E_n$, $\mathcal{E} ⊩ ⅄(E_1 \vee \ldots \vee E_n) \to ⅄E_1 \vee \ldots \vee ⅄E_n$.

Proof. We let $\mathcal{E}$ act as the following strategy prescribes, with i ranging over $\{1, 2, 3, \ldots\}$ and j over $\{1, \ldots, n\}$:

Procedure LOOP: Keep granting permission until the adversary makes a move $1.i.j.\alpha$ (resp. $2.j.i.\alpha$); then make the move $2.j.i.\alpha$ (resp. $1.i.j.\alpha$), and repeat LOOP. □

Lemma 53. $⊩ ⅂(E_1 \vee \ldots \vee E_n) \to ⅂E_1 \vee \ldots \vee ⅂E_n$.

Moreover, there is an effective procedure that takes any particular value of n and constructs an EPM $\mathcal{E}$ such that, for any $E_1, \ldots, E_n$, $\mathcal{E} ⊩ ⅂(E_1 \vee \ldots \vee E_n) \to ⅂E_1 \vee \ldots \vee ⅂E_n$.

Proof. Here is the algorithm for $\mathcal{E}$:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then:

If $\gamma = 1.w$:, then make the n moves $2.1.w$:, …, $2.n.w$:, and repeat LOOP;

If $\gamma = 1.w.j.\alpha$ (resp. $\gamma = 2.j.w.\alpha$) where $1 \le j \le n$, then make the move $2.j.w.\alpha$ (resp. $1.w.j.\alpha$), and repeat LOOP. □

Lemma 54. $⊩ ⅄E \vee ⅄E \to ⅄E$.

Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} ⊩ ⅄E \vee ⅄E \to ⅄E$.

Proof. We let $\mathcal{E}$ work as the following strategy prescribes:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then act depending on which of the following cases applies, and after that repeat LOOP:
If $\gamma = 1.1.i.\alpha$, then make the move $2.j.\alpha$ where $j = 2i - 1$;
If $\gamma = 1.2.i.\alpha$, then make the move $2.j.\alpha$ where $j = 2i$;
If $\gamma = 2.j.\alpha$ where $j = 2i - 1$, then make the move $1.1.i.\alpha$;
If $\gamma = 2.j.\alpha$ where $j = 2i$, then make the move $1.2.i.\alpha$. □

Lemma 55. $\Vdash \wr E \vee \wr E \to \wr E$.
Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} \Vdash \wr E \vee \wr E \to \wr E$.

Proof. The idea of a strategy for $\mathcal{E}$ is to first replicate the consequent turning it into $\wr(E^* \circ E^*)$, which is essentially the same as $\wr E^* \vee \wr E^*$, and then switch to a strategy that is essentially the same as the ordinary copy-cat strategy. Precisely, here is how $\mathcal{E}$ works: it makes the move $2.\epsilon$: (replicating the consequent), after which it follows the following algorithm:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then:
If $\gamma = 2.0\alpha$ (resp. $\gamma = 1.1.\alpha$), then make the move $1.1.\alpha$ (resp. 2.0α), and repeat LOOP;
If $\gamma = 2.1\alpha$ (resp. $\gamma = 1.2.\alpha$), then make the move $1.2.\alpha$ (resp. 2.1α), and repeat LOOP;
If $\gamma = 2.\epsilon.\alpha$, then make the moves $1.1.\epsilon.\alpha$ and $1.2.\epsilon.\alpha$, and repeat LOOP. □

Lemma 56. $\Vdash \curlyvee\curlyvee E \to \curlyvee E$.
Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} \Vdash \curlyvee\curlyvee E \to \curlyvee E$.

Proof. We select any effective one-to-one function f from the set of all pairs of nonnegative integers onto the set of all nonnegative integers. Below is the interactive algorithm that $\mathcal{E}$ follows:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. Then:
If $\gamma = 1.i.j.\alpha$, then make the move $2.k.\alpha$ where $k = f(i, j)$, and repeat LOOP;
If $\gamma = 2.k.\alpha$, then make the move $2.i.j.\alpha$ where $k = f(i, j)$, and repeat LOOP. □

Lemma 57. $\Vdash \wr\wr E \to \wr E$.
Moreover, there is an EPM $\mathcal{E}$ such that, for any E, $\mathcal{E} \Vdash \wr\wr E \to \wr E$.

Proof. Our proof of this lemma, unlike that of the "similar" Lemma 56, is fairly long. For this reason, it is given separately in Section 11.12.4. □

In what follows, we will be using the expressions $E^*(x)$, $E^*(t)$, etc. to mean the same as the more clumsy $(E(x))^*$, $(E(t))^*$, etc. Also, remember from Section 11.3 that, when t is a constant, $e(t) = t$.

Lemma 58. $\Vdash \sqcap x(E(x) \to F(x)) \to (\sqcap xE(x) \to \sqcap xF(x))$.

Moreover, there is an EPM $\mathcal{E}$ such that, for any $E(x)$ and $F(x)$,

$$\mathcal{E} \Vdash \sqcap x(E(x) \to F(x)) \to (\sqcap xE(x) \to \sqcap xF(x)).$$

Proof. Strategy: Wait till the environment makes the move '2.2.c' for some constant c. This brings the $\sqcap xF^*(x)$ component down to $F^*(c)$ and hence the entire game to

$$\sqcap x(E^*(x) \to F^*(x)) \to (\sqcap xE^*(x) \to F^*(c)).$$

Then make the same move c in the antecedent and in $\sqcap xE^*(x)$, i.e. make the two moves '1.c' and '2.1.c'. The game will be brought down to $(E^*(c) \to F^*(c)) \to (E^*(c) \to F^*(c))$. Finally, switch to $\mathcal{CCS}$. □

Lemma 59. *Assume t is free for x in $E(x)$. Then $\Vdash E(t) \to \sqcup xE(x)$.*

Moreover, there is an effective function that takes any t and constructs an EPM $\mathcal{E}$ such that, for any $E(x)$, whenever t is free for x in $E(x)$, $\mathcal{E} \Vdash E(t) \to \sqcup xE(x)$.

Proof. Strategy: Let $c = e(t)$. Read c from the valuation tape if necessary (i.e. if t is a variable). Then make the move '2.c', bringing the game down to $E^*(c) \to E^*(c)$. Then switch to $\mathcal{CCS}$. □

Lemma 60. *Assume E does not contain x. Then $\Vdash E \to \sqcap xE$.*

Moreover, there is an EPM $\mathcal{E}$ such that, for any E and x, as long as E does not contain x, $\mathcal{E} \Vdash E \to \sqcap xE$.

Proof. In this case we prefer to explicitly write the usually suppressed parameter e. Consider an arbitrary E not containing x, and an arbitrary interpretation * admissible for $E \to \sqcap xE$. The formula $E \to \sqcap xE$ contains x yet E does not. Therefore, from the definition of admissibility and with a little thought we can see that E^* does not depend on x. In turn, this means—as can be seen with some additional thought—that the move c by the environment (whatever constant c) in $e[\sqcap xE^*]$ brings this game down to $e[E^*]$. With this observation in mind, the following strategy can be seen to be successful: Wait till the environment makes the move '2.c' for some constant c. Then read the sequence '1.α_1', ..., '1.α_n' of (legal) moves possibly made by the environment before it made the above move '2.c', and make the n moves '2.α_1', ..., '2.α_n'. It can be seen that now the original game $e[E^*] \to e[\sqcap xE^*]$ will have been brought down to

$\langle\Phi\rangle e[E^*] \to \langle\Phi\rangle e[E^*]$, where $\Phi = \langle \top\alpha_1, \ldots, \top\alpha_n\rangle$. So, switching to CCS at this point guarantees success. □

Lemma 61. *Assume $E(x)$ does not contain y. Then $\Vdash \sqcap yE(y) \to \sqcap xE(x)$. In fact, $CCS \Vdash \sqcap yE(y) \to \sqcap xE(x)$.*

Proof. Assuming that $E(x)$ does not contain y and analyzing the relevant definitions, it is not hard to see that, for any interpretation * admissible for $\sqcap yE(y) \to \sqcap xE(x)$, we simply have $(\sqcap yE(y))^* = (\sqcap xE(x))^*$. So, we deal with a game of the form $A \to A$, for which the ordinary copy-cat strategy is successful. □

11.12.4 Iteration principle for branching recurrence

The computability principles expressed by the formulas $\curlyvee\curlyvee E \to \curlyvee E$ and $\wr\wr E \to \wr E$, that can be equivalently rewritten as $\curlywedge E \to \curlywedge\curlywedge E$ and $\diamondsuit E \to \diamondsuit\diamondsuit E$, we call **iteration principles** (for $\curlywedge$ and $\diamondsuit$, respectively). This section is entirely devoted to a proof of Lemma 57, i.e. the iteration principle for $\diamondsuit$. We start with some auxiliary definitions.

A **colored bit** b is a pair (c, d), where c, called the **content** of b, is in $\{0, 1\}$, and d, called the **color** of b, is in $\{blue, yellow\}$. We will be using the notation $\overline{c}$ ("blue c") for the colored bit whose content is c and color is *blue*, and $\underline{c}$ ("yellow c") for the bit whose content is c and color is *yellow*. The four colored bits will be treated as symbols, from which, just as from ordinary bits, we can form strings.

A **colored bitstring** is a finite or infinite string of colored bits. Consider a colored bitstring v. The **content** of v is the result of "ignoring the colors" in v, i.e. replacing every bit of v by the content of that bit. The **blue content** of v is the content of the string that results from deleting in v all but blue bits. **Yellow content** is defined similarly. We use $\underline{v}$, $\overline{v}$ and $\underline{v}$ to denote the content, blue content and yellow content of v, respectively. Example: if $v = \overline{1}\underline{0}\overline{0}\underline{1}$, we have $\underline{v} = 10001$, $\overline{v} = 10$ and $\underline{v} = 001$. As in the case of ordinary bitstrings, ϵ stands for the empty colored bitstring, and $u \preceq w$ means that u is a (not necessarily proper) initial segment of w.

Definition 62. *A* **colored bitstring tree** (**CBT**) *is a set T of colored bitstrings, called its* **branches***, such that the following conditions are satisfied:*

(a) *The set $\{\overline{\underline{v}} \mid v \in T\}$, which we denote by $\overline{\underline{T}}$, is a BT in the sense of Definition 14.*

(b) *For any $w, u \in T$, if $\overline{\underline{w}} = \overline{\underline{u}}$, then $w = u$.*

(c) *For no (finite) $v \in T$ do we have $\{v\overline{0}, v\underline{1}\} \subseteq T$ or $\{v\underline{0}, v\overline{1}\} \subseteq T$.*

A colored bitstring v is said to be a **leaf** *of T iff $\overline{\underline{v}}$ is a leaf of $\overline{\underline{T}}$.*

When represented in the style of (the underlying tree of) Figure 11.11 of Section 11.4.6, a CBT will look like an ordinary BT, with the only difference that now every edge will have one of the colors *blue* or *yellow*. Also, by condition (c), both of the outgoing edges ("sibling" edges) of any non-leaf node will have the same color.

Lemma 63. *Assume T is a CBT, and w, u are branches of T with $\overline{w} \preceq \overline{u}$ and $\underline{w} \preceq \underline{u}$. Then $w \preceq u$.*

Proof. Assume T is a CBT, $w, u \in T$, and $w \not\preceq u$. We want to show that then $\overline{w} \not\preceq \overline{u}$ or $\underline{w} \not\preceq \underline{u}$. Let v be the longest common initial segment of w and u, so we have $w = vw'$ and $u = vu'$ for some w', u' such that w' is nonempty and w' and u' do not have a nonempty common initial segment. Assume the first bit of w' is $\overline{0}$ (the cases when it is $\overline{1}$, $\underline{0}$ or $\underline{1}$, of course, will be similar). If u' is empty, then w obviously contains more blue bits than u does, and we are done. Assume now u' is nonempty, in particular, b is the first bit of u'. Since w' and u' do not have a nonempty common initial segment, b should be different from $\overline{0}$. By condition (b) of Definition 62, the content of b cannot be 0 (for otherwise we would have $v\overline{0} = vb$ and hence $b = \overline{0}$). Consequently, b is either $\overline{1}$ or $\underline{1}$. The case $b = \underline{1}$ is ruled out by condition (c) of Definition 62. Thus, $b = \overline{1}$. But the blue content of $v\overline{0}$ is $\overline{v}0$ while the blue content of $v\overline{1}$ is $\overline{v}1$. Taking into account the obvious fact that the former is an initial segment of $\overline{w}$ and the latter is an initial segment of $\overline{u}$, we find $\overline{w} \not\preceq \overline{u}$. □

The uniform solution $\mathcal{E}$ for $\wr\wr E \to \wr E$ that we are going to construct essentially uses a copy-cat strategy between the antecedent and the consequent. Of course, however, this strategy cannot be applied directly in the form of $\mathcal{CCS}$. The problem is that while a position of $\wr E^*$ is a decorated tree $\mathcal{T}$ in the style of Figure 11.11 of Section 11.4.6, in the case of $\wr\wr E^*$ it is a tree $\mathcal{T}'$ of trees such as, say, the one shown below:

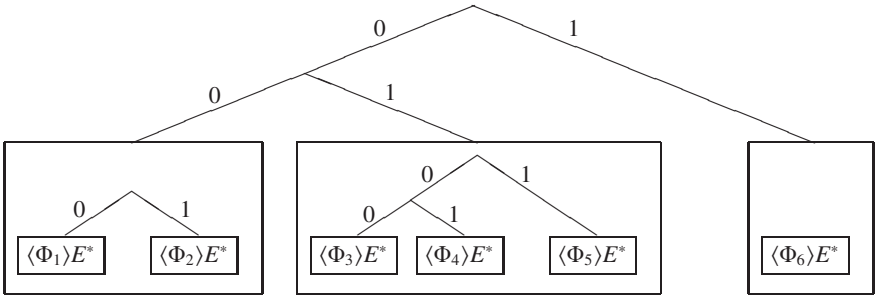


Figure 11.13: A position $\mathcal{T}'$ of $\wr\wr E^*$

The trick that $\mathcal{E}$ uses is that it sees the BT-structure of $\mathcal{T}$ as a colored bit-string tree, letting such a $\mathcal{T}$ "simulate" $\mathcal{T}'$. Specifically, $\mathcal{E}$ tries to maintain a one-to-one correspondence between the leaves (windows) of $\mathcal{T}$ and the leaves of the leaves (small windows) of $\mathcal{T}'$, with the positions of E^* in each pair of corresponding windows being identical. Figure 11.14 shows a possible $\mathcal{T}$, whose six windows, as we see, have the same contents as the six small windows of $\mathcal{T}'$, even if some permutation in the order of windows has taken place.

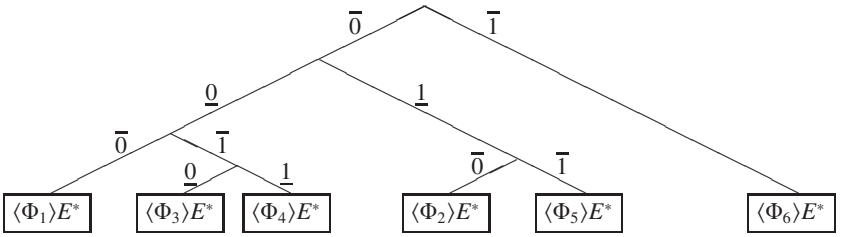


Figure 11.14: A possible corresponding position $\mathcal{T}$ of $\wr E^*$

The way this mapping works is that to a leaf y of a leaf x of $\mathcal{T}'$ corresponds a (the) leaf of $\mathcal{T}$ whose yellow content is y and blue content is x, and vice versa. For example, look at the small window containing $\langle\Phi_3\rangle E^*$ in Figure 11.13. It is leaf 00 of leaf 01 of $\mathcal{T}'$; and the window containing the same $\langle\Phi_3\rangle E^*$ in Figure 11.14 is leaf $\overline{0}\underline{0}\overline{1}\underline{0}$ of $\mathcal{T}$, whose yellow content is indeed 00 and blue content is indeed 01. $\mathcal{E}$ has a way to maintain such a correspondence. Let us see how it acts when the antecedent and the consequent of $\wr\wr E^* \to \wr E^*$ have evolved to the positions shown in Figures 11.13 and 11.14, respectively. If, say, the environment makes a non-replicative move α in the $\langle\Phi_3\rangle E^*$ component of $\mathcal{T}'$ (resp. $\mathcal{T}$), $\mathcal{E}$ responds by the same move α in the $\langle\Phi_3\rangle E^*$ component of $\mathcal{T}$ (resp. $\mathcal{T}'$). This ensures that the positions of E^* in the two components remain identical. Of course, the environment can make a move α in several components at once. For example, the move $1.00.\epsilon.\alpha$ by the environment would amount to making move α in the two components $\langle\Phi_1\rangle E^*$ and $\langle\Phi_2\rangle E^*$ of $\mathcal{T}'$. No problem, $\mathcal{E}$ can respond by the two consecutive moves $2.000.\alpha$ and $2.010.\alpha$. Now let us see how $\mathcal{E}$ handles replicative moves. If the environment replicates a small window of $\mathcal{T}'$, $\mathcal{E}$ replicates the corresponding window of $\mathcal{T}$, and colors the two newly emerged edges into yellow. In fact, the environment can replicate several small windows at once, as, say, would be the case if it makes the move 0.1: in the antecedent, amounting to replicating the two windows containing $\langle\Phi_2\rangle E^*$ and $\langle\Phi_5\rangle E^*$. But, again, this is not a problem, for $\mathcal{E}$ can respond by two (or whatever number of small windows of the antecedent have been replicated) consecutive replicative moves in the consequent. Suppose now the environment

replicates a large window of $\mathcal{T}'$, such as, say, window 00. Then $\mathcal{E}$ responds by replicative moves in all leaves of $\mathcal{T}$ whose blue content is 00, specifically, leaves 000 and 010, and colors the newly emerged edges into blue. With some thought one can see that, with this strategy, it is guaranteed that to any leaf y of any leaf x of the updated $\mathcal{T}'$ again corresponds the leaf of the updated $\mathcal{T}$ whose yellow content is y and blue content is x, and vice versa.

In an attempt to understand what replicative moves (ignoring all non-replicative moves in-between) could have resulted in the tree of Figure 11.13 and the corresponding tree of Figure 11.14, we find the following. First, the environment made the replicative move ϵ: in the antecedent of $\mathbin{⫯}\mathbin{⫯}E^* \rightarrow \mathbin{⫯}E^*$. To this $\mathcal{E}$ responded by the replicative move ϵ: in the consequent. Then the environment made the ("deep") replicative move $0.\epsilon$: in the antecedent. To this $\mathcal{E}$ responded by 0: in the consequent. Next the environment made the replicative move 0: in the antecedent. $\mathcal{E}$ responded by 00: and 01: in the consequent. Finally, the environment made (the "deep") replicative move 01.0: in the antecedent, and $\mathcal{E}$ responded by 001: in the consequent.

Keeping, according to the above scenario, all runs of E^* in the branches of branches of the antecedent identical with runs of E^* in the corresponding branches of the consequent can be eventually seen to guarantee a win for $\mathcal{E}$.

Now we describe $\mathcal{E}$ in precise terms. At the beginning, this EPM creates a record T of the type 'finite CBT', and initializes it to $\{\epsilon\}$. After that, $\mathcal{E}$ follows the following procedure:

Procedure LOOP: Keep granting permission until the adversary makes a move γ. If γ satisfies the conditions of one of the following four cases, act as the corresponding case prescribes. Otherwise go to an infinite loop in a permission state.

Case (i): $\gamma = 1.w$: for some bitstring w. Let $v_1, \ldots, v_k$ be[14] all of the leaves v of T with $w = \overline{v}$. Then make the moves $2.\underline{\overline{v}_1}:, \ldots, 2.\underline{\overline{v}_k}:$, update T to $T \cup \{v_1\overline{0}, v_1\overline{1}, \ldots, v_k\overline{0}, v_k\overline{1}\}$, and repeat LOOP.

Case (ii): $\gamma = 1.w.u$: for some bitstrings w, u. Let $v_1, \ldots, v_k$ be all of the leaves v of T such that $w \preceq \overline{v}$ and $u = \underline{v}$. Then make the moves $2.\underline{\overline{v}_1}:, \ldots, 2.\underline{\overline{v}_k}:$, update T to $T \cup \{v_1\underline{0}, v_1\underline{1}, \ldots, v_k\underline{0}, v_k\underline{1}\}$, and repeat LOOP.

Case (iii): $\gamma = 1.w.u.\alpha$ for some bitstrings w, u and move α. Let $v_1, \ldots, v_k$ be all of the leaves v of T such that $w \preceq \overline{v}$ and $u \preceq \underline{v}$. Then make the moves $2.\underline{\overline{v}_1}.\alpha, \ldots, 2.\underline{\overline{v}_k}.\alpha$, and repeat LOOP.

Case (iv): $\gamma = 2.w.\alpha$ for some bitstring w. Let $v_1, \ldots, v_k$ be all of the leaves v of T with $w \preceq \underline{\overline{v}}$. Then make the moves $1.\overline{v}_1.\underline{v}_1.\alpha, \ldots, 1.\overline{v}_k.\underline{v}_k.\alpha$, and repeat LOOP.

[14] In each of the four cases we assume that the list $v_1, \ldots, v_k$ is arranged lexicographically.

Pick an arbitrary interpretation * admissible for $\wp\wp E \to \wp E$, an arbitrary valuation e and an arbitrary e-computation branch B of $\mathcal{E}$. Let Θ be the run spelled by B. The work of $\mathcal{E}$ does not depend on e. And, as e is going to be fixed, we can safely omit this parameter (as we usually did in the previous section) and just write E^* instead of $e[E^*]$. Of course, $\mathcal{E}$ is interpretation-blind, so, as long as it wins $\wp\wp E^* \to \wp E^*$, it is a uniform solution for $\wp\wp E \to \wp E$.

From the description of LOOP it is immediately clear that B is a fair. Hence, in order to show that $\mathcal{E}$ wins, it would be sufficient to verify that $\mathbf{Wn}^{\wp\wp E^* \to \wp E^*}\langle\Theta\rangle = \top$.

Let $N = \{1, \ldots, m\}$ if LOOP is iterated the finite number m of times in B, and $N = \{1, 2, 3, \ldots\}$ otherwise. For $i \in N$, we let T_i denote the value of record T at the beginning of the ith iteration of LOOP. Next, Θ_i will mean the initial segment of Θ consisting of the (lab)moves made by the beginning of the ith iteration of LOOP. Finally, Ψ_i will stand for $\neg\Theta_i^{1\cdot}$ and Φ_i for $\Theta_i^{2\cdot}$.

From the description of LOOP it is obvious that, for each $i \in N$, T_i is a finite colored tree, and that $T_1 \subseteq T_2 \subseteq \ldots \subseteq T_i$. In our subsequent reasoning we will implicitly rely on this fact.

Lemma 64. *For every i with $i \in N$, we have:*

(a) *Φ_i is a prelegal position of $\wp E^*$, and $Tree^{\wp E^*}\langle\Phi_i\rangle = \overline{\underline{T_i}}$.*

(b) *Ψ_i is a prelegal position of $\wp\wp E^*$.*

(c) *For every leaf x of $Tree^{\wp\wp E^*}\langle\Psi_i\rangle$, $\Psi_i^{\preceq x}$ is a prelegal position of $\wp E^*$.*

(d) *For every leaf z of T_i, $\overline{z}$ is a leaf of $Tree^{\wp\wp E^*}\langle\Psi_i\rangle$ and $\underline{z}$ is a leaf of $Tree^{\wp E^*}\langle\Psi_i^{\preceq\overline{z}}\rangle$.*

(e) *For every leaf x of $Tree^{\wp\wp E^*}\langle\Psi_i\rangle$ and every leaf y of $Tree^{\wp E^*}\langle\Psi_i^{\preceq x}\rangle$, there is a leaf z of T_i such that $x = \overline{z}$ and $y = \underline{z}$. By Lemma 63, such a z is unique.*

(f) *For every leaf z of T_i, $\Phi_i^{\preceq\overline{z}} = (\Psi_i^{\preceq\overline{z}})^{\preceq\underline{z}}$.*

(g) *Θ_i is a legal position of $\wp\wp E^* \to \wp E^*$; hence, $\Phi_i \in \mathbf{Lr}^{\wp E^*}$ and $\Psi_i \in \mathbf{Lr}^{\wp\wp E^*}$.*

Proof. We proceed by induction on i. The basis case with $i = 1$ is rather straightforward for each clause of the lemma and we do not discuss it. For the inductive step, assume $i + 1 \in N$, and the seven clauses of the lemma are true for i.

Clause (a): By the induction hypothesis, Φ_i is a prelegal position of $\wp E^*$ and $Tree^{\wp E^*}\langle\Phi_i\rangle = \overline{\underline{T_i}}$. Assume the ith iteration of LOOP deals with Case (i), so that $\Phi_{i+1} = \langle\Phi_i, \top\overline{\underline{v}}_1{:}, \ldots, \top\overline{\underline{v}}_k{:}\rangle$.[15] Each of $\overline{\underline{v}}_1, \ldots, \overline{\underline{v}}_k$ is a leaf of $\overline{\underline{T_i}}$, i.e. a leaf of $Tree^{\wp E^*}\langle\Phi_i\rangle$. This guarantees that Φ_{i+1} is a prelegal position of $\wp E^*$. Also, by the definition of function *Tree*, we have $Tree^{\wp E^*}\langle\Phi_{i+1}\rangle = Tree^{\wp E^*}\langle\Phi_i\rangle \cup \{\overline{\underline{v}}_1 0, \overline{\underline{v}}_1 1, \ldots, \overline{\underline{v}}_k 0, \overline{\underline{v}}_k 1\}$. But the latter is nothing but $\overline{\underline{T_{i+1}}}$ as can

[15] With $v_1, \ldots, v_k$ here and in later cases being as in the corresponding clause of the description of LOOP.

be seen from the description of how Case (i) updates T_i to T_{i+1}. A similar argument applies when the ith iteration of LOOP deals with Case (ii). Assume now the ith iteration of LOOP deals with Case (iii). Note that the moves made in the consequent of $\mathbin{?}\mathbin{?}E^* \rightarrow \mathbin{?}E^*$ (the moves that bring Φ_i to Φ_{i+1}) are nonreplicative—specifically, look like $\underline{\overline{v}}.\alpha$ where $\underline{\overline{v}} \in \underline{\overline{T}}_i = \mathit{Tree}^{\mathbin{?}E^*}\langle\Phi_i\rangle$. Such moves do not destroy prelegality nor do they change the value of *Tree*, so $\mathit{Tree}^{\mathbin{?}E^*}\langle\Phi_i\rangle = \mathit{Tree}^{\mathbin{?}E^*}\langle\Phi_{i+1}\rangle$. It remains to note that T is not updated in this subcase, so that we also have $\underline{\overline{T}}_{i+1} = \underline{\overline{T}}_i$. Hence $\mathit{Tree}^{\mathbin{?}E^*}\langle\Phi_{i+1}\rangle = \underline{\overline{T}}_{i+1}$. Finally, suppose the ith iteration of LOOP deals with Case (iv). It is the environment who moves in the consequent of $\mathbin{?}\mathbin{?}E^* \rightarrow \mathbin{?}E^*$, and does so before the machine makes any moves (in the antecedent). Then the clean environment assumption, in conjunction with the induction hypothesis for clause (g), implies that such a move by the environment cannot bring Φ_i to an illegal and hence non-prelegal position of $\mathbin{?}E^*$. So, Φ_{i+1} remains a prelegal position of $\mathbin{?}E^*$. As for $\mathit{Tree}^{\mathbin{?}E^*}\langle\Phi_{i+1}\rangle = \underline{\overline{T}}_{i+1}$, it holds for the same reason as in the previous case.

Clause (b): If the ith iteration of LOOP deals with Case (i), (ii) or (iii), it is the environment who moves in the antecedent of $\mathbin{?}\mathbin{?}E^* \rightarrow \mathbin{?}E^*$, and does so before the machine makes any moves. Therefore the clean environment assumption, with the induction hypothesis for clause (g) in mind, guarantees that Ψ_{i+1} is a legal and hence prelegal position of $\mathbin{?}\mathbin{?}E^*$. Assume now that the ith iteration of LOOP deals with Case (iv), so that $\Psi_{i+1} = \langle\Psi_i, \bot\overline{v}_1.\underline{v}_1.\alpha, \ldots, \bot\overline{v}_k.\underline{v}_k.\alpha\rangle$. By the induction hypothesis, Ψ_i is a prelegal position of $\mathbin{?}\mathbin{?}E^*$. And, by the induction hypothesis for clause (d), each $\overline{v}_j$ $(1 \leq j \leq k)$ is a leaf of $\mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_i\rangle$, so adding the (lab)moves $\bot\overline{v}_1.\underline{v}_1.\alpha, \ldots, \bot\overline{v}_k.\underline{v}_k$ does not bring Ψ_i to a non-prelegal position. Ψ_{i+1} thus remains a prelegal position of $\mathbin{?}\mathbin{?}E^*$. As an aside, note also that those moves, being nonreplicative, do not modify $\mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_i\rangle$.

Clause (c): Just as in the previous clause, when the ith iteration of LOOP deals with Case (i), (ii) or (iii), the desired conclusion follows from the clean environment assumption in conjunction with the induction hypothesis for clause (g). Assume now that the ith iteration of LOOP deals with Case (iv). Consider any leaf x of $\mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_{i+1}\rangle$. As noted at the end of our proof of Clause (b), we have $\mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_i\rangle = \mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_{i+1}\rangle$, so x is also a leaf of $\mathit{Tree}^{\mathbin{?}\mathbin{?}E^*}\langle\Psi_i\rangle$. Therefore, if $\Psi_{i+1}^{\leq x} = \Psi_i^{\leq x}$, the conclusion that $\Psi_{i+1}^{\leq x}$ is a prelegal position of $\mathbin{?}E^*$ follows from the induction hypothesis. Suppose now $\Psi_{i+1}^{\leq x} \neq \Psi_i^{\leq x}$. Note that then, in view of the induction hypothesis for clause (d), $\Psi_{i+1}^{\leq x}$ looks like $\langle\Psi_i^{\leq x}, \bot y_1.\alpha, \ldots, \bot y_m.\alpha\rangle$, where for each y_j $(1 \leq j \leq m)$ we have $\overline{z} = x$ and $\underline{z} = y_j$ for some leaf z of T_i, with y_j being a leaf of $\mathit{Tree}^{\mathbin{?}E^*}\langle\Psi_i^{\leq x}\rangle$. By the induction hypothesis for the present clause, $\Psi_i^{\leq x}$ is a prelegal position of $\mathbin{?}E^*$. Adding to such a position the nonreplicative labmoves $\bot y_1.\alpha, \ldots, \bot y_m.\alpha$—where the y_j are leaves of $\mathit{Tree}^{\mathbin{?}E^*}\langle\Psi_i^{\leq x}\rangle$—cannot bring it to a non-prelegal position. Thus, $\Psi_{i+1}^{\leq x}$ remains a prelegal position of $\mathbin{?}E^*$.

Clauses (d) and (e): If the ith iteration of LOOP deals with Cases (iii) or (iv), T_i is not modified, and no moves of the form x: or $x.y$: (where x, y are bitstrings) are made in the antecedent of $\wp\wp E^* \to \wp E^*$, so $Tree^{\wp\wp E^*}\langle\Psi_i\rangle$ and $Tree^{\wp E^*}\langle\Psi_i^{\preceq x}\rangle$ (any leaf x of $Tree^{\wp\wp E^*}\Psi_i$) are not affected, either. Hence Clauses (d) and (e) for $i+1$ are automatically inherited from the induction hypothesis for these clauses. This inheritance also takes place—even if no longer "automatically"—when the ith iteration of LOOP deals with Case (i) or (ii). This can be verified by a routine analysis of how Cases (i) and (ii) modify T_i and the other relevant parameters. Details are left to the reader.

Clause (f): Consider any leaf z of T_{i+1}. When the ith iteration of LOOP deals with Case (i) or (ii), no moves of the form $x.\alpha$ are made in the consequent of $\wp\wp E^* \to \wp E^*$, and no moves of the form $x.y.\alpha$ are made in the antecedent (any bitstrings x, y). Based on this, it is easy to see that for all bitstrings x, y we have $\Phi_{i+1}^{\preceq x} = \Phi_i^{\preceq x}$ and $(\Psi_{i+1}^{\preceq x})^{\preceq y} = (\Psi_i^{\preceq x})^{\preceq y}$. Hence clause (f) for $i+1$ is inherited from the same clause for i. Now suppose the ith iteration of LOOP deals with Case (iii). Then $T_{i+1} = T_i$ and hence z is also a leaf of T_i. From the description of Case (iii) one can easily see that if $w \not\preceq \overline{z}$ or $u \not\preceq \underline{z}$, we have $\Phi_{i+1}^{\preceq\overline{z}} = \Phi_i^{\preceq\overline{z}}$ and $(\Psi_{i+1}^{\preceq\overline{z}})^{\preceq\underline{z}} = (\Psi_i^{\preceq\overline{z}})^{\preceq\underline{z}}$, so the equation $\Phi_{i+1}^{\preceq\overline{z}} = (\Psi_{i+1}^{\preceq\overline{z}})^{\preceq\underline{z}}$ is true by the induction hypothesis; and if $w \preceq \overline{z}$ and $u \preceq \underline{z}$, then $\Phi_{i+1}^{\preceq\overline{z}} = \langle\Phi_i^{\preceq\overline{z}}, \top\alpha\rangle$ and $(\Psi_{i+1}^{\preceq\overline{z}})^{\preceq\underline{z}} = \langle(\Psi_i^{\preceq\overline{z}})^{\preceq\underline{z}}, \top\alpha\rangle$. But, by the induction hypothesis, $\Phi_i^{\preceq\overline{z}} = (\Psi_i^{\preceq\overline{z}})^{\preceq\underline{z}}$. Hence $\Phi_{i+1}^{\preceq\overline{z}} = (\Psi_{i+1}^{\preceq\overline{z}})^{\preceq\underline{z}}$. A similar argument applies when the ith iteration of LOOP deals with Case (iv).

Clause (g): Below we implicitly rely on the induction hypothesis, according to which $\Theta_i \in \mathbf{Lr}^{\wp\wp E^* \to \wp E^*}$ and hence $\Phi_i \in \mathbf{Lr}^{\wp E^*}$ and $\Psi_i \in \mathbf{Lr}^{\wp\wp E^*}$. Note that, with the clean environment assumption in mind, all of the moves made in any of Cases (i)–(iv) of LOOP have the prefix '1.' or '2.', i.e. are made either in the antecedent or the consequent of $\wp\wp E^* \to \wp E^*$. Hence, in order to show that Θ_{i+1} is a legal position of $\wp\wp E^* \to \wp E^*$, it would suffice to verify that $\Phi_{i+1} \in \mathbf{Lr}^{\wp E^*}$ and $\Psi_{i+1} \in \mathbf{Lr}^{\wp\wp E^*}$.

Suppose the ith iteration of LOOP deals with Case (i) or (ii). The clean environment assumption guarantees that $\Psi_{i+1} \in \mathbf{Lr}^{\wp\wp E^*}$. In the consequent of $\wp\wp E^* \to \wp E^*$ only replicative moves are made. Replicative moves can yield an illegal position (Φ_{i+1} in our case) of a $\wp$-game only if they yield a non-prelegal position. But, by clause (a), Φ_{i+1} is a prelegal position of $\wp E^*$. Hence it is also a legal position of $\wp E^*$.

Suppose now the ith iteration of LOOP deals with Case (iii). Again, that $\Psi_{i+1} \in \mathbf{Lr}^{\wp\wp E^*}$ is guaranteed by the clean environment assumption. So, we only need to verify that $\Phi_{i+1} \in \mathbf{Lr}^{\wp E^*}$. By clause (a), this position is a prelegal position of $\wp E^*$. So, it remains to see that, for any leaf y of $Tree^{\wp E^*}\langle\Phi_{i+1}\rangle$, $\Phi_{i+1}^{\preceq y} \in \mathbf{Lr}^{E^*}$. Pick an arbitrary leaf y of $Tree^{\wp E^*}\langle\Phi_{i+1}\rangle$ —i.e., by clause (a), of $\overline{\underline{T}}_{i+1}$. Let z be the leaf of T_{i+1} with $y = \overline{\underline{z}}$. We already know that $\Psi_{i+1} \in \mathbf{Lr}^{\wp\wp E^*}$.

By clause (d), we also know that $\overline{z}$ is a leaf of $Tree^{\mathrm{??}E^*}\langle\Psi_{i+1}\rangle$. Consequently, $\Psi_{i+1}^{\leq\overline{z}} \in \mathbf{Lr}^{\mathrm{?}E^*}$. Again by clause (d), $\underline{z}$ is a leaf of $Tree^{\mathrm{?}E^*}\langle\Psi_{i+1}^{\leq\overline{z}}\rangle$. Hence, $(\Psi_{i+1}^{\leq\overline{z}})^{\leq\underline{z}}$ should be a legal position of E^*. But, by clause (f), $\Phi_{i+1}^{\leq\overline{z}} = (\Psi_{i+1}^{\leq\overline{z}})^{\leq\underline{z}}$. Thus, $\Phi_{i+1}^{\leq\overline{z}} \in \mathbf{Lr}^{E^*}$, i.e. $\Phi_{i+1}^{\leq y} \in \mathbf{Lr}^{E^*}$.

Finally, suppose the ith iteration of LOOP deals with Case (iv). By the clean environment assumption, $\Phi_{i+1} \in \mathbf{Lr}^{\mathrm{?}E^*}$. Now consider Ψ_{i+1}. This position is a prelegal position of $\mathrm{??}E^*$ by clause (b). So, in order for Ψ_{i+1} to be a legal position of $\mathrm{??}E^*$, for every leaf x of $Tree^{\mathrm{??}E^*}\langle\Psi_{i+1}\rangle$ we should have $\Psi_{i+1}^{\leq x} \in \mathbf{Lr}^{\mathrm{?}E^*}$. Consider an arbitrary such leaf x. By clause (c), $\Psi_{i+1}^{\leq x}$ is a prelegal position of $\mathrm{?}E^*$. Hence, a sufficient condition for $\Psi_{i+1}^{\leq x} \in \mathbf{Lr}^{\mathrm{?}E^*}$ is that, for every leaf y of $Tree^{\mathrm{?}E^*}\langle\Psi_{i+1}^{\leq x}\rangle$, $(\Psi_{i+1}^{\leq x})^{\leq y} \in \mathbf{Lr}^{E^*}$. So, let y be an arbitrary such leaf. By clause (e), there is a leaf z of T_{i+1} such that $\overline{z} = x$ and $\underline{z} = y$. Therefore, by clause (f), $\Phi_{i+1}^{\leq\overline{z}} = (\Psi_{i+1}^{\leq x})^{\leq y}$. But we know that $\Phi_{i+1} \in \mathbf{Lr}^{\mathrm{?}E^*}$ and hence (with clause (a) in mind) $\Phi_{i+1}^{\leq\overline{z}} \in \mathbf{Lr}^{E^*}$. Consequently, $(\Psi_{i+1}^{\leq x})^{\leq y} \in \mathbf{Lr}^{E^*}$. □

Lemma 65. *For every finite initial segment* Υ *of* Θ*, there is* $i \in N$ *such that* Υ *is a* (*not necessarily proper*) *initial segment of* Θ_i *and hence of every* Θ_j *with* $i \leq j \in N$.

Proof. The statement of the lemma is straightforward when there are infinitely many iterations of LOOP, for each iteration adds a nonzero number of new moves to the run and hence there are arbitrarily long Θ_is, each of them being an initial segment of Θ. Suppose now LOOP is iterated a finite number m of times. It would be (necessary and) sufficient to verify that in this case $\Theta = \Theta_m$, i.e. no moves are made during the last iteration of LOOP. But this is indeed so. From the description of LOOP we see that the machine does not make any moves during a given iteration unless the environment makes a move γ first. So, assume $\bot$ makes move γ during the mth iteration of LOOP. By the clean environment assumption, we should have $\langle\Theta_m, \bot\gamma\rangle \in \mathbf{Lr}^{\mathrm{??}E^*\rightarrow\mathrm{?}E^*}$. It is easy to see that such a γ would have to satisfy the conditions of one of the Cases (i)–(iv) of LOOP. But then there would be an $(m + 1)$th iteration of LOOP, contradicting out assumption that there are only m iterations. □

Let us use Ψ and Φ to denote $\neg\Theta^{1\cdot}$ and $\Theta^{2\cdot}$, respectively. Of course, the statement of Lemma 65 is true for Φ and Ψ (instead of Θ) as well. Taking into account that, by definition, a given run is legal iff all of its finite initial segments are legal, the following fact is an immediate corollary of Lemmas 65 and 64(g):

$$\Theta \in \mathbf{Lr}^{\mathrm{??}E^*\rightarrow\mathrm{?}E^*}. \textit{ Hence, } \Psi \in \mathbf{Lr}^{\mathrm{??}E^*} \textit{ and } \Phi \in \mathbf{Lr}^{\mathrm{?}E^*}. \tag{12}$$

To complete our proof of Lemma 57, we need to show that

$$\mathbf{Wn}^{⫯⫯E^*\to⫯E^*}\langle\Theta\rangle = \top.$$

With (12) in mind, if $\mathbf{Wn}^{⫯⫯E^*}\langle\Psi\rangle = \bot$, by the definition of $\to$, we are done. Assume now $\mathbf{Wn}^{⫯⫯E^*}\langle\Psi\rangle = \top$. Then, by the definition of ⫯, there is a complete branch x of $\mathit{Tree}^{⫯⫯E^*}\langle\Psi\rangle$ such that $\mathbf{Wn}^{⫯E^*}\langle\Psi^{\preceq x}\rangle = \top$. This, in turn, means that, for some complete branch y of $\mathit{Tree}^{⫰E^*}\langle\Psi^{\preceq x}\rangle$,

$$\mathbf{Wn}^{E^*}\langle(\Psi^{\preceq x})^{\preceq y}\rangle = \top. \tag{13}$$

Fix these x and y. For each $i \in N$, let x_i denote the (obviously unique) leaf of $\mathit{Tree}^{⫯⫯E^*}\langle\Psi_i\rangle$ such that $x_i \preceq x$; and let y_i denote the (again unique) leaf of $\mathit{Tree}^{⫯E^*}\langle\Psi_i^{\preceq x_i}\rangle$ such that $y_i \preceq y$. Next, let z_i denote the leaf of T_i with $\overline{z}_i = x_i$ and $\underline{z}_i = y_i$. According to Lemma 64(e), such a z_i exists and is unique.

Consider any i with $i+1 \in N$. Clearly $x_i \preceq x_{i+1}$ and $y_i \preceq y_{i+1}$. By our choice of the z_j, we then have $\overline{z}_i \preceq \overline{z}_{i+1}$ and $\underline{z}_i \preceq \underline{z}_{i+1}$. Hence, by Lemma 63, $z_i \preceq z_{i+1}$. Let us fix z as the shortest (perhaps infinite if N is infinite) colored bitstring such that for every $i \in N$, $z_i \preceq z$. Based on the just-made observation that we always have $z_i \preceq z_{i+1}$, such a z exists. And, in view of Lemma 64(a), it is not hard to see that $\underline{\overline{z}}$ is a complete branch of $\mathit{Tree}^{⫯E^*}\langle\Phi\rangle$.

With Lemma 65 in mind, Lemma 64(f) easily allows us to find that $\Phi^{\preceq\underline{\overline{z}}} = (\Psi^{\preceq x})^{\preceq y}$. Therefore, by (13), $\mathbf{Wn}^{E^*}\langle\Phi^{\preceq\underline{\overline{z}}}\rangle = \top$. By the definition of ⫯, this means that $\mathbf{Wn}^{⫯E^*}\langle\Phi\rangle = \top$. Hence, by the definition of $\to$ and with (12) in mind, $\mathbf{Wn}^{⫯⫯E^*\to⫯E^*}\langle\Theta\rangle = \top$. Done.

11.12.5 Finishing the soundness proof for affine logic

Now we are ready to prove Theorem 37. Consider an arbitrary sequent S provable in **AL**. By induction on the **AL**-derivation of S, we are going to show that S has a uniform solution $\mathcal{E}$. This is sufficient to conclude that **AL** is 'uniformly sound'. The theorem also claims 'constructive soundness', i.e. that such an $\mathcal{E}$ can be effectively built from a given **AL**-derivation of S. This claim of the theorem will be automatically taken care of by the fact that our proof of the existence of $\mathcal{E}$ is constructive: all of the uniform-validity and closure lemmas on which we rely provide a way for actually constructing a corresponding uniform solution. With this remark in mind and for the considerations of readability, in what follows we only talk about uniform validity without explicitly mentioning uniform solutions for the corresponding formulas/sequents and without explicitly showing how to construct such solutions.

There are 16 cases to consider, corresponding to the 16 possible rules that might have been used at the last step of an **AL**-derivation of S, with S being the conclusion of the rule. In each non-axiom case below, "induction hypothesis"

means the assumption that the premise(s) of the corresponding rule is (are) uniformly valid. The goal in each case is to show that the conclusion of the rule is also uniformly valid. "Modus ponens" should be understood as Lemma 45, "generalized modus ponens" as Lemma 46, and "transitivity" as Lemma 47. Also, clauses (f) and (g) of Lemma 38, in combination with modus ponens, always allow us to rewrite a statement $\Vdash \underline{G_1} \vee \underline{H} \vee \underline{G_2}$ as $\Vdash \underline{G_1} \vee (\underline{H}) \vee \underline{G_2}$, and vice versa. We will often explicitly or implicitly rely on this fact, which we call **associativity** (of $\vee$).

Identity Axiom: By Lemma 38(a).

$\top$**-Axiom:** Of course, $\Vdash \top$.

Exchange: By Lemma 38(b), $\Vdash E \vee F \rightarrow F \vee E$. And, by Lemma 38(e),

$$\Vdash (E \vee F \rightarrow F \vee E) \rightarrow (\underline{G} \vee E \vee F \vee \underline{H} \rightarrow \underline{G} \vee F \vee E \vee \underline{H}).$$

Hence, by modus ponens,

$$\Vdash \underline{G} \vee E \vee F \vee \underline{H} \rightarrow \underline{G} \vee F \vee E \vee \underline{H}.$$

But, by the induction hypothesis, $\Vdash \underline{G} \vee E \vee F \vee \underline{H}$. Hence, by modus ponens, $\Vdash \underline{G} \vee F \vee E \vee \underline{H}$.

Weakening: Similar to the previous case, using Lemma 38(i) instead of Lemma 38(b).

$\curlyvee$**-Contraction**: By Lemma 38(e) (with empty $\underline{S}$),

$$\Vdash (\curlyvee E \vee \curlyvee E \rightarrow \curlyvee E) \rightarrow (\underline{G} \vee \curlyvee E \vee \curlyvee E \rightarrow \underline{G} \vee \curlyvee E).$$

And, by Lemma 54, $\Vdash \curlyvee E \vee \curlyvee E \rightarrow \curlyvee E$. Hence, by modus ponens, $\Vdash \underline{G} \vee \curlyvee E \vee \curlyvee E \rightarrow \underline{G} \vee \curlyvee E$. But, by the induction hypothesis, $\Vdash \underline{G} \vee \curlyvee E \vee \curlyvee E$. Hence, by modus ponens, $\Vdash \underline{G} \vee \curlyvee E$.

$\curlywedge$**-Contraction**: Similar to $\curlyvee$-contraction, using Lemma 55 instead of Lemma 54.

$\sqcup$**-Introduction:** By Lemma 38(j), $\Vdash E_i \rightarrow E_1 \sqcup \ldots \sqcup E_n$; and, by Lemma 38(e),

$$\Vdash (E_i \rightarrow E_1 \sqcup \ldots \sqcup E_n) \rightarrow (\underline{G} \vee E_i \rightarrow \underline{G} \vee (E_1 \sqcup \ldots \sqcup E_n)).$$

Modus ponens yields $\Vdash \underline{G} \vee E_i \rightarrow \underline{G} \vee (E_1 \sqcup \ldots \sqcup E_n)$. But, by the induction hypothesis, $\Vdash \underline{G} \vee E_i$. So, by modus ponens, $\Vdash \underline{G} \vee (E_1 \sqcup \ldots \sqcup E_n)$.

$\sqcap$-Introduction: By the induction hypothesis,

$$\Vdash \underline{G} \vee E_1, \ \ldots, \ \Vdash \underline{G} \vee E_n.$$

And, from Lemma 38(k),

$$\Vdash (\underline{G} \vee E_1) \wedge \ldots \wedge (\underline{G} \vee E_n) \rightarrow \underline{G} \vee (E_1 \sqcap \ldots \sqcap E_n).$$

Generalized modus ponens yields $\Vdash \underline{G} \vee (E_1 \sqcap \ldots \sqcap E_n)$.

$\vee$-Introduction: In view of associativity, this rule is trivial.

$\wedge$-Introduction: By the induction hypothesis,

$$\Vdash \underline{G_1} \vee E_1, \ \ldots, \ \Vdash \underline{G_n} \vee E_n.$$

And, from Lemma 38(d),

$$\Vdash (\underline{G_1} \vee E_1) \wedge \ldots \wedge (\underline{G_n} \vee E_n) \rightarrow \underline{G_1} \vee \ldots \vee \underline{G_n} \vee (E_1 \wedge \ldots \wedge E_n).$$

Generalized modus ponens yields $\Vdash \underline{G_1} \vee \ldots \vee \underline{G_n} \vee (E_1 \wedge \ldots \wedge E_n)$.

$?$-Introduction: By the induction hypothesis, $\Vdash \underline{G} \vee E$. And, by Lemma 48, $\Vdash E \rightarrow ?E$. So, by Lemma 38(e) and modus ponens applied twice, $\Vdash \underline{G} \vee ?E$.

$!$-Introduction: Similar to $?$-introduction, using Lemma 49 instead of Lemma 48.

λ-Introduction: By the induction hypothesis, $\Vdash \underline{?G} \vee E$. If $\underline{?G}$ is empty, then $\underline{?G} \vee E = E$ and thus $\Vdash E$. Hence, by Lemma 42, $\Vdash \lambda E$, i.e. $\Vdash \underline{?G} \vee \lambda E$. Suppose now $\underline{?G}$ is not empty. Associativity allows us to rewrite $\Vdash \underline{?G} \vee E$ as $(\Vdash (\underline{?G}) \vee E$ and thus) $\Vdash \neg \underline{?G} \rightarrow E$. Then, by Lemma 42, $\Vdash \lambda(\neg \underline{?G} \rightarrow E)$. From here, by Lemma 50 and modus ponens, we get $\Vdash \lambda \neg \underline{?G} \rightarrow \lambda E$, which can be rewritten as $\Vdash ?? \underline{G} \vee \lambda E$. But, by Lemma 52, $\Vdash ?? \underline{G} \rightarrow \underline{??G}$ and, by Lemma 38(e),

$$\Vdash (?? \underline{G} \rightarrow \underline{??G}) \rightarrow (?? \underline{G} \vee \lambda E \rightarrow \underline{??G} \vee \lambda E).$$

Applying modus ponens twice yields $\Vdash \underline{??G} \vee \lambda E$. From here, using Lemmas 56, 38(e) and modus ponens as many times as the number of disjuncts in $\underline{??G}$, we get $\Vdash \underline{?G} \vee \lambda E$.

$!$-Introduction: Similar to λ-introduction, using Lemmas 43, 51, 53 and 57 instead of Lemmas 42, 50, 52 and 56, respectively.

$\sqcup$-Introduction: By Lemma 59, $\Vdash E(t) \rightarrow \sqcup x E(x)$. And, by Lemma 38(e),

$$\Vdash (E(t) \rightarrow \sqcup x E(x)) \rightarrow (\underline{G} \vee E(t) \rightarrow \underline{G} \vee \sqcup x E(x)).$$

Modus ponens yields $\Vdash \underline{G} \vee E(t) \rightarrow \underline{G} \vee \bigsqcup x E(x)$. But, by the induction hypothesis, $\Vdash \underline{G} \vee E(t)$. Hence, by modus ponens, $\Vdash \underline{G} \vee \bigsqcup x E(x)$.

$\bigsqcap$-Introduction: First, consider the case when $\underline{G}$ is nonempty. By the induction hypothesis, we have $\Vdash \underline{G} \vee E(y)$, which can be rewritten as $\Vdash \neg\underline{G} \rightarrow E(y)$. Therefore, by Lemma 44, $\Vdash \bigsqcap y(\neg\underline{G} \rightarrow E(y))$ and, by Lemma 58 and modus ponens, $\Vdash \bigsqcap y \neg\underline{G} \rightarrow \bigsqcap y E(y)$. At the same time, by Lemma 60, $\Vdash \neg\underline{G} \rightarrow \bigsqcap y \neg\underline{G}$. By transitivity, we then get $\Vdash \neg\underline{G} \rightarrow \bigsqcap y E(y)$. But, by Lemma 61, $\Vdash \bigsqcap y E(y) \rightarrow \bigsqcap x E(x)$. Transitivity yields $\Vdash \neg\underline{G} \rightarrow \bigsqcap x E(x)$, which can be rewritten as $\Vdash \underline{G} \vee \bigsqcap x E(x)$. The case when $\underline{G}$ is empty is simpler, for then $\Vdash \underline{G} \vee \bigsqcap x E(x)$, i.e. $\Vdash \bigsqcap x E(x)$, can be obtained directly from the induction hypothesis by Lemmas 44, 61 and modus ponens.

11.13 What could be next?

As a conclusive remark, the author wants to point out that the story told in this chapter was, in fact, only about the tip of the iceberg. Even though the phrase "*the* language of CL" was used in some semiformal contexts, such a language has no official boundaries and, depending on particular needs or taste, remains open to various interesting extensions. In a broad sense, CL is not a particular syntactic system or a particular semantics for a particular collection of operators, but rather a platform and ambitious program for redeveloping logic as a formal theory of computability, as opposed to the formal theory of truth which it has more traditionally been.

The general framework of CL is also ready to accommodate any reasonable weakening modifications of its absolute-strength computation model HPM, thus keeping a way open for studying logics of sub-Turing computability and developing a systematic theory of interactive complexity. Among modifications of this sort, for example, might be depriving the HPM of its infinite work tape, leaving in its place just a write-only buffer where the machine constructs its moves. In such a modification the exact type of read access to the run and valuation tapes becomes relevant, and a reasonable restriction would apparently be to allow—perhaps now multiple—read heads to move only in one direction. An approach favoring this sort of machines would try to model Turing (unlimited) or sub-Turing (limited) computational resources such as memory, time, etc. as games, and then understand computing a problem A with resources represented by R as computing $R \rightarrow A$, thus making explicit not only trans-Turing (incomputable) resources as we have been doing in this paper, but also all of the Turing/sub-Turing resources needed or allowed for computing A—the resources that the ordinary HPM or Turing machine models take for granted. So, with T representing the infinite read/write tape as a computational resource, computability of A in the old sense would mean nothing but computability of $T \rightarrow A$ in the new sense: having T in the antecedent would

amount to having infinite memory, only this time provided externally (by the environment) via the run tape rather than internally via the work tape.

Complexity and sub-Turing computability aside, there are also good philosophical reasons for questioning the legitimacy of the presence of an infinite work tape, whether it be in our HPM model or in the ordinary Turing machine (TM) model. The point is that neither HPMs nor TMs can be implemented—even in principle—as actual physical beings. This is so for the simple reason that no real mechanical device will ever have an infinite (even if only potentially so) internal memory. The reason why this fact does not cause much frustration and usually remains unnoticed is that the tape can be easily thought of as an *external resource*, and thus TMs or HPMs can be identified only with their finite control parts; then and only then, they indeed become implementable devices. Yet, the standard formal treatment of TMs or our treatment of HPMs does not account for this implicit intuition, and views the infinite work tape as a part of the machine. Computability logic, with its flexibility and ability to keep an accurate and explicit count of all resources, makes it possible to painlessly switch from TMs or HPMs to truly finite devices, and make things really what they were meant to be.

An alternative or parallel direction for CL to evolve with a focus shift from computability to complexity, could be extending its vocabulary with *complexity-conscious* operators. For example, winning a complexity-conscious version $\sqcap^p x \sqcup^p y A(x, y)$ of $\sqcap x \sqcup y A(x, y)$ could mean existence of a polynomial-time function f such that, to any move m by the environment, the machine responds with a move n within time $f(m)$, and then wins the game $A(m, n)$.

Time has not yet matured for seriously addressing complexity or sub-Turing computability issues though, and in the nearest future CL will probably remain focused on just computability: as it happens, there are still too many unanswered questions here. The most important and immediate task is finding axiomatizations for incrementally expressive fragments of CL—first of all, fragments that involve recurrence operators, for which practically no progress has been made so far (with the intuitionistic fragment of CL being one modest exception). It is also highly desirable to fully translate **CL4**-style ad hoc axiomatizations into some systematic and nice proof theory, such as cirquent calculus. So far this has only been done in Japaridze (2007b, 2008a) for the $\neg, \wedge, \vee$-fragment.

References

Abramsky, S. and Jagadeesan, R. (1994). Games and full completeness for multiplicative linear logic. *Journal of Symbolic Logic*, 59(2):543–574.

Blass, A. (1972). Degrees of indeterminacy of games. *Fundamenta Mathematicæ*, 77:151–166.

Blass, A. (1992). A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56: 183–220.

Felscher, W. (1985). Dialogues, strategies, and intuitionistic provability. *Annals of Pure and Applied Logic*, 28:217–254.

Girard, J.Y. (1987). Linear logic. *Theoretical Computer Science*, 50(1):1–102.

Gödel, K. (1958). Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287.

Goldin, D. Q., Smolka, S. A., Attie, P. C., and Sonderegger, E. L. (2004). Turing machines, transition systems, and interaction. *Information and Computation*, 194(2):101–128.

Japaridze, G. (1997). A constructive game semantics for the language of linear logic. *Annals of Pure and Applied Logic*, 85:87–156.

Japaridze, G. (2000). The propositional logic of elementary tasks. *Notre Dame Journal of Formal Logic*, 41(2):171–183.

Japaridze, G. (2002). The logic of tasks. *Annals of Pure and Applied Logic*, 117:263–295.

Japaridze, G. (2003). Introduction to computability logic. *Annals of Pure and Applied Logic*, 123:1–99.

Japaridze, G. (2006a). From truth to computability I. *Theoretical Computer Science*, 357:100–135.

Japaridze, G. (2006b). Introduction to cirquent calculus and abstract resource semantics. *Journal of Logic and Computation*, 16(4):489–532.

Japaridze, G. (2006c). Propositional computability logic I. *ACM Transactions on Computational Logic*, 7(2):302–330.

Japaridze, G. (2006d). Propositional computability logic II. *ACM Transactions on Computational Logic*, 7(2):331–362.

Japaridze, G. (2006e). Computability logic: a formal theory of interaction. In Goldin, D., Smolka, S., and Wegner, P., editors, *Interactive Computation: The New Paradigm*. pages 183–223. Springer, Berlin.

Japaridze, G. (2007a). From truth to computability II. *Theoretical Computer Science*, 379:20–52.

Japaridze, G. (2007b). Intuitionistic computability logic. *Acta Cybernetica*, 18(1):77–113.

Japaridze, G. (2007c). The logic of interactive Turing reduction. *Journal of Symbolic Logic*, 72(1):243–276.

Japaridze, G. (2007d). The intuitionistic fragment of computability logic at the propositional level. *Annals of Pure and Applied Logic*, 147(3):187–227.

Japaridze, G. (2008a). Cirquent calculus deepened. *Journal of Logic and Computation*, 18(6):983–1028.

Japaridze, G. (2008b). Sequential operators in computability logic. *Information and Computation*, 206(12):1443–1475.

Japaridze, G. (2009). Many concepts and two logics of algorithmic reduction. *Studia Logica*, 91 (to appear).

Kleene, S.C. (1952). *Introduction to Metamathematics*. D. van Nostrand Company, New York/Toronto.

Kolmogorov, A.N. (1932). Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift*, 35:58–65.

Konolige, K. (1988). On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35(3):343–382.

Levesque, H. and Lakemeyer, G. (2000). *The Logic of Knowledge Bases*. MIT Press, Cambridge, MA.

Lorenzen, P. (1959). Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods*, pages 193–200. PWN, Warsaw.

Milner, R. (1993). Elements of interaction. *Communications of the ACM*, 36(1):79–89.

Moore, R. (1985). A formal theory of knowledge and action. In Hobbs, J. and Moore, R., editors, *Formal Theories of Commonsense Worlds*. Ablex, Norwood, NJ.

Pietarinen, A. (2002). *Semantic Games in Logic and Language*. Academic dissertation, University of Helsinki, Helsinki.

Sipser, M. (2006). *Introduction to the Theory of Computation*. Thomson Course Technology, USA, 2nd edition.

Turing, A. (1936). On computable numbers with an application to the entsheidungsproblem. *Proceedings of the London Mathematical Society*, 42(2):230–265.

van Benthem, J. (2001). Logic in games. ILLC preprint, University of Amsterdam, Amsterdam.

Vereshchagin, N. (2006) Japaridze's computability logic and intuitionistic propositional calculus. Moscow State University Preprint, 2006.
`http://lpcs.math.msu.su/~ver/papers/japaridze.ps`.

Wegner, P. (1998). Interactive foundations of computing. *Theoretical Computer Science*, 192:315–351.

Chapter 12

THE PROBLEM OF DETERMINACY OF INFINITE GAMES FROM AN INTUITIONISTIC POINT OF VIEW

Wim Veldman
Institute for Mathematics, Astrophysics and Particle Physics,
Radboud University Nijmegen
W.Veldman@math.ru.nl

Abstract Taking Brouwer's intuitionistic standpoint, we examine finite and infinite games of perfect information for players *I* and *II*. If one understands the disjunction occurring in the classical notion of determinacy constructively, even finite games are not always determinate. We therefore suggest an intuitionistically different notion of determinacy and prove that every subset of Cantor space is determinate in the proposed sense. Our notion is biased and considers games from the viewpoint of player *I*. In Cantor space, both player *I* and player *II* have two alternative possibilities for each move. It turns out that every subset of a space, where player *II* has, for each one of his moves, no more than a finite number of alternative possibilities while player *I* perhaps has infinitely many choices, is determinate in the proposed sense from the viewpoint of player *I*.

'We must have a bit of a fight, but I don't care about going on long,'
said Tweedledum. 'What's the time now?'
Tweedledee looked at his watch, and said, 'Half-past four.'
'Let's fight till six, and then have dinner,' said Tweedledum.

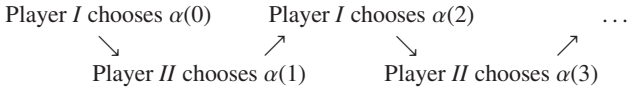
—Lewis Carroll, [1865, page 190]

12.1 Intuitionistic determinacy: the problem, and the case of two-move-games

12.1.1 $\mathbb{N}$ is the set of natural numbers, and *Baire space* $\mathcal{N}$ is the set of all infinite sequences of natural numbers. We use $m, n, p, q, \ldots$ as variables over the set $\mathbb{N}$ and $\alpha, \beta, \ldots$ as variables over the set $\mathcal{N}$.

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

Let A be a subset of $\mathcal{N}$. We describe *the game for* A, sometimes called $\mathcal{G}(A)$. There are two players, I and II, who, each time they play the game, together build an infinite sequence α in $\mathcal{N}$, as follows:



The sequence α is called a *play* in the game for A. Player I is the winner if and only if α belongs to A. The set A is sometimes called the *payoff set* of the game.

Following the classical definition, we say that the set A is *determinate* if and only if *either* player I has a method to secure that he wins every play in the game for A, *or* player II has a method to prevent that player I wins any play in the game for A.

We take the intuitionistic point of view advocated by L. E. J. Brouwer. He insisted that every mathematical statement should be considered as a report on what we have been able to prove and that connectives and quantifiers and the corresponding set-theoretic operations should be interpreted constructively. In particular, a disjunctive statement $P \vee Q$ is considered proven if and only if we either have a proof of P or a proof of Q. We not only follow the rules of intuitionistic logic but also make use of some of the new axioms Brouwer proposed as a result of his reflection on the problem how to handle the concept of the continuum. Some of these axioms, the so-called *continuity principles*, are classically unacceptable but our main result, Theorem 3.5, does not depend on any axiom that does not stand a classical reading.

12.1.2 We not only want to study games in Baire space $\mathcal{N}$ but also games that are played in certain subspaces of Baire space $\mathcal{N}$, traditionally called *spreads* in intuitionistic mathematics. To this end we introduce some notations and some terminology.

$\mathbb{N}^*$ is the set of finite sequences of natural numbers. We suppose that a bijective mapping $(a_0, a_1, \ldots, a_{n-1}) \mapsto \langle a_0, a_1, \ldots, a_{n-1}\rangle$ from $\mathbb{N}^*$ to $\mathbb{N}$ is given, a function *coding* the finite sequences of natural numbers by means of natural numbers. There is a function, *length*, from $\mathbb{N}$ to $\mathbb{N}$ such that, for every natural number a, $m := \mathit{length}(a)$ is the length of the finite sequence coded by a.

Let a be a number of length m. We consider a as a function from the set $\{0, 1, \ldots, m-1\}$ to $\mathbb{N}$, and, for each n, if $n < m$, we define $a(n)$ to be the value of this function at n.

$*$ is the binary function on $\mathbb{N}$ which, via the coding, corresponds to the operation of concatenating finite sequences. We assume that for each a, n, $a \leq a*\langle n\rangle$.

For each infinite sequence of natural numbers α, and each natural number n, we define $\overline{\alpha}(n)$ to be (the code number of) the finite sequence $\langle \alpha(0), \ldots, \alpha(n-1)\rangle$. If confusion seems unlikely, we write $\overline{\alpha}n$ rather than $\overline{\alpha}(n)$.

For each infinite sequence of natural numbers α, for each natural number s, we define: α *passes through* s if and only if there exists n such that $\overline{\alpha}n = s$.

Let σ belong to $\mathcal{N}$. σ is called a *spread-law* if and only if $\sigma(\langle\ \rangle) = 0$ and, for each a, $\sigma(a) = 0$ if and only if, for some n, $\sigma(a * \langle n\rangle) = 0$. If $\sigma(a) = 0$, we will say that a is *admitted by* σ.

Let σ be a spread-law and let α belong to $\mathcal{N}$. We say: σ *admits* α, if and only if, for each n, $\sigma(\overline{\alpha}n) = 0$. The set of all infinite sequences of natural numbers α admitted by the spread-law σ is called a *spread* and this set is also named σ. The statements "σ *admits* α" and "α *belongs to* σ" are equivalent.

Observe that a subset X of $\mathcal{N}$ coincides with a spread if and only if (i) X is *(sequentially) closed*, that is, every α such that, for each n, some element of X passes through $\overline{\alpha}n$, belongs itself to X, and (ii) X is *located*, that is, there exists σ in $\mathcal{N}$ such that for every s, s contains an element of X if and only if $\sigma(s) = 0$.

Let σ be a spread and let A be a subset of σ. We describe *the game for A in* σ. There are again two players, *I* and *II*, who, each time they play the game, join up to build an infinite sequence α in $\mathcal{N}$, but they have to take care that the infinite sequence α will belong to the spread σ:

Player *I* chooses $\alpha(0)$ such that $\sigma(\overline{\alpha}(1)) = 0$ $\searrow$ … $\nearrow$

Player *II* chooses $\alpha(1)$ such that $\sigma(\overline{\alpha}(2)) = 0$

Player *I* is the winner if and only if the infinite sequence α belongs to the set A.

Given some spread σ, we want to call a subset A of σ *determinate* if and only if *either* player *I* has a sure method to win the game for A in σ, *or* player *II* has a sure method to prevent player *I* from winning the game for A in σ.

12.1.3 In order to make the notion of determinacy more precise, we introduce the concept of a strategy.

Let σ be a spread. We let $\mathrm{Strat}_I(\sigma)$, the set of *strategies in* σ *for player I*, be the set of all functions γ in $\mathcal{N}$ such that for each a, if σ admits a and *length*(a) is even, then σ admits $a * \langle\gamma(a)\rangle$, and, if σ does not admit a or *length*(a) is odd, then $\gamma(a) = 0$. Observe that $\mathrm{Strat}_I(\sigma)$ itself is a spread.

Let α belong to σ and γ to $\mathrm{Strat}_I(\sigma)$. We define: α *I-obeys* γ, if and only if, for each n, $\alpha(2n) = \gamma(\overline{\alpha}(2n))$.

Similarly, we let $\mathrm{Strat}_{II}(\sigma)$, the set of *strategies for player II in* σ, be the set of all functions γ in $\mathcal{N}$ such that for each a, if σ admits a and *length*(a) is odd, then σ admits $a * \langle\gamma(a)\rangle$, and, if σ does not admit a or *length*(a) is even, then $\gamma(a) = 0$.

Let α belong to σ and γ to $\mathrm{Strat}_{II}(\sigma)$. We define: α *II-obeys* γ, if and only if, for each n, $\alpha(2n + 1) = \gamma(\overline{\alpha}(2n + 1))$.

Let A be a subset of σ. We define: A is *strongly determinate in* σ if and only if *either* there is a strategy γ for player I, such that every α in σ I-obeying γ belongs to A, *or* there is a strategy δ for player II in σ such that every α in σ II-obeying δ does not belong to A.

Let A be a subset of σ and let γ be strategy for player I in σ. We say: γ *wins* A *for player* I if and only if every α in σ I-obeying γ belongs to A.

12.1.4 Even games in which players I, II make only finitely many moves, and each move is a choice from finitely many alternatives, need not be strongly determinate.

Consider for instance the game with no moves at all that is won by player I if and only if Riemann's hypothesis holds. To say that this game is strongly determinate is equivalent to deciding Riemann's hypothesis.

Fortunately, the language of intuitionistic mathematics is more refined than the language of classical mathematics, and we may consider formulations of the notion of determinacy that, from a classical point view, would be equivalent to the first formulation, but, from an intuitionistic point of view, are weaker. Here is such a notion.

Let σ be a spread and A a subset of σ. We define: A is *determinate in* σ *from the viewpoint of player* I if and only if: *if* every strategy for player II in σ is II-obeyed by at least one element of A, *then* there is a strategy γ for player I in σ such that every α in σ I-obeying γ belongs to A.

We took the disjunctive formulation of strong determinacy, $P \vee Q$, changed it into $(\neg Q) \rightarrow P$, and then replaced the negative antecedent $\neg Q$ by a stronger, positive statement.

The definition is biased, as it considers the problem of the determinacy of A from the viewpoint of player I. It is easy to guess when we want to call a subset A of σ *determinate from the viewpoint of player* II. We will see, in Section 12.1.8, that there exist a spread σ and a subset A of σ such that A is determinate from the viewpoint of player I, while we are unable to prove that A is determinate from the viewpoint of player II.

12.1.5 We interrupt our discussion of the notion of determinacy and ask attention for one of the axioms of intuitionistic analysis.

Let σ be a spread and let ζ belong to $\mathcal{N}$. We define: ζ *codes a continuous function from* σ *to* $\mathcal{N}$ if and only if, for all n, for all α in σ, there exists m such that $\alpha(\langle n\rangle * \overline{\alpha}m) \neq 0$.

Suppose that σ is a spread, and that ζ codes a continuous function from σ to $\mathcal{N}$. For each α in σ we define $\zeta|\alpha$ to be the sequence β such that, for all n, p in $\mathbb{N}$, if p is the least m such that $\zeta(\langle n\rangle * \overline{\alpha}m) \neq 0$, then $\zeta(\langle n\rangle * \overline{\alpha}p) = \beta(n) + 1$.

The following axiom, occurring under its present name in Veldman (2006a), and called *Brouwer's principle for functions* in Kleene and Vesley (1965),

GAC$_{1,1}$ in Gielen et al. (1981), and **C-C** in Troelstra and van Dalen (1988), is incompatible with a classical reading of the quantifiers. It seems to be the strongest possible formulation of a principle Brouwer is using in his intuitionistic papers.

Second Axiom of Continuous Choice: *Let σ be a spread and let R be a subset of $\sigma \times \mathcal{N}$. If, for all α in σ, there exists β such that $\alpha R \beta$, then there exists ζ coding a continuous function from σ to $\mathcal{N}$ such that, for all α in σ, $\alpha R(\zeta|\alpha)$.*

(We write "$\alpha R \beta$" while intending "(α, β) belongs to R".)

12.1.6 We now continue the discussion of the notion of determinacy.

Let σ be a spread, let A be a subset of σ and suppose that every strategy for player *II* in σ is *II*-obeyed by at least one element of A. Using the Second Axiom of Continuous Choice we find some ζ coding a continuous function from $\mathrm{Strat}_{II}(\sigma)$ to $\mathcal{N}$ such that, for every γ in $\mathrm{Strat}_{II}(\sigma)$, $\zeta|\gamma$ *II*-obeys γ and belongs to A.

An element ζ of $\mathcal{N}$ coding a continuous function from $\mathrm{Strat}_{II}(\sigma)$ to σ such that for every γ in $\mathrm{Strat}_{II}(\sigma)$, $\zeta|\gamma$ belongs to σ and *II*-obeys γ will be called an *anti-strategy* for player *I*. If, in addition, for every γ in $\mathrm{Strat}_{II}(\sigma)$, $\zeta|\gamma$ belongs to A we say that *ζ secures the set A for player I*.

Suppose that there exists an anti-strategy ζ for player *I* that secures the set A for player *I*. What use can player *I* make of it, when actually playing the game? Observe that, when playing the game, player *I* does not know which strategy his opponent is following. In order to win, he should be able, while producing, together with his opponent, a play α, to conjecture a strategy δ for player *II* such that $\zeta|\delta = \alpha$. At first sight, that does not seem to be a very easy task. Observe however that, if $\zeta|\delta = \alpha$, then, for each n there exists m such that for every strategy γ for player *II*, if $\overline{\delta}m = \overline{\gamma}m$, then $(\zeta|\delta)(2n) = (\zeta|\gamma)(2n)$. This means that, for a given n, player *I* may be sure that $\zeta|\delta$ passes through $\overline{\alpha}(2n)$, while he has only a finite piece of information on the strategy player *II* is following. Everyone who has a nephew and once played chess with him, should now imagine this nephew to be player *I*. Player *I*, each time he has to make a move, first asks a number of questions: "What will be your reply if I should make this move? And if I should continue so-and-so and make that move?" Somehow knowing how to make his opponent answer his questions, he collects information and ponders, consulting ζ, and then, at some point, he triumphantly takes his decision. Knowing also how to compel player *II* to act according to the given answers, he is sure that the resulting play α will belong to A.

If you are a grown-up, such questioning is no longer allowed, and you have lost the power of making your opponent do as you like. But might not player *I*, by studying his anti-strategy ζ, find a *strategy* γ, such that every α *I*-obeying γ

belongs to A, that is, might he not develop, by some hard thinking, a successful way of playing the game without asking unlawful questions and intimidating player *II*? That question is the main subject of this paper.

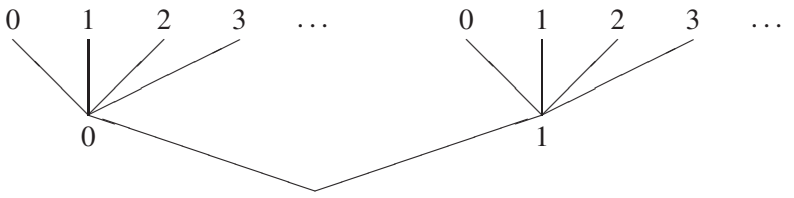
12.1.7 Let σ be a spread, let A be a subset of σ. We define: *A is predeterminate in σ from the viewpoint of player I* if and only if, *if* there is an anti-strategy for player I in σ that secures the set A for player I, *then* there is a strategy for player I in σ that wins the set A for player I.

Observe that every subset of σ that is determinate in σ from the viewpoint of player I, is also predeterminate in σ from the viewpoint of player I.

The Second Axiom of Continuous Choice implies the converse: every subset of σ that is predeterminate in σ from the viewpoint of player I, is also determinate in σ from the viewpoint of player I.

12.1.8 Disappointingly, if we allow player *II* to choose from countably many alternatives, there exist two-move games that are not determinate from the viewpoint of player I in the new weak sense.

We consider games of the following kind: Player I chooses either 0 or 1, player *II* chooses a natural number, and the game is over.



A strategy for player I in a game of this kind consists of a single number, viz. his first and only move which is either 0 or 1.

A strategy for player *II* is a pair (p, q) of natural numbers, p being the answer player *II* will give to a first move 0, and q being the answer player *II* will give to a first move 1.

A subset A of $\{0, 1\} \times \mathbb{N}$ is determinate from the viewpoint of player I in the sense of Section 12.1.4 if and only if: *if*, for all p, for all q, *either* $(0, p)$ belongs to A *or* $(1, q)$ belongs to A, *then: either*, for all p, $(0, p)$ belongs to A, *or*, for all q, $(1, q)$ belongs to A.

A subset A of $\{0, 1\} \times \mathbb{N}$ is predeterminate from the viewpoint of player I in the sense of Section 12.1.7 if and only if: *if* there exists α such that for all p, for all q, either $\alpha(\langle p, q\rangle) = 0$ and $(0, p)$ belongs to A, or $\alpha(\langle p, q\rangle) = 1$ and $(1, q)$ belongs to A, *then: either* for all p, $(0, p)$ belongs to A, *or*, for all q, $(1, q)$ belongs to A.

The following axiom, called $^*2.2$ in Kleene and Vesley (1965) and $\mathbf{AC_{0,0}}$ in Gielen et al. (1981), and a weak consequence of the Second Axiom of Continuous Choice, implies that every subset of $\{0,1\} \times \mathbb{N}$ that is determinate from the viewpoint of player I is also predeterminate from the viewpoint of player I.

First Axiom of Countable Choice: *For each subset R of $\mathbb{N} \times \mathbb{N}$, if for all m there exists n such that mRn, then there exists α such that, for all m, $mR(\alpha(m))$.*

The intuitionistic mathematician will judge this axiom to be true because he allows himself to build an infinite sequence $\alpha = \alpha(0), \alpha(1), \ldots$ step by step, by successive free choices. He does not demand that the future course of the sequence be prescribed by means of an algorithm.

A classical mathematician would say that, if we have a proof that for all m there exists n such that mRn, a suitable α may be *defined*, as follows: let, for each m, $\alpha(m)$ be the least n such that mRn. It may occur, however, that we have a proof of $0R1$ and are uncertain if $0R0$ is true or not. In such a case the given rule is useless for the constructive mathematician.

The unwelcome truth is that not every subset A of $\{0,1\} \times \mathbb{N}$ is predeterminate from the viewpoint of player I, as we may learn from the following *counterexample in Brouwer's style*:

> Let $p : \mathbb{N} \to \{0, 1, \ldots, 9\}$ be the decimal expansion of π. We let A be the subset of $\{0,1\} \times \mathbb{N}$ consisting of all pairs (i, n) such that *either* $i = 0$ and *if* there exists $j < n$ such that, for all $k < 99$, $p(j + k) = 9$, *then* the first such j is odd, *or* $i = 1$ and *if* there exists $j < n$ such that, for all $k < 99$, $p(j + k) = 9$, *then* the first such j is even. For all p, for all q, either $(0, p)$ belongs to A or $(1, q)$ belongs to A.
>
> Assuming that A is predeterminate we obtain the conclusion that *either* for all p, $(0, p)$ belongs to A, *or* for all q, $(1, q)$ belongs to A. In the first case we must have a proof that, *if* there exists j such that, for all $k < 99$, $p(j + k) = 9$, *then* the first such j is odd, and in the second case we must have a proof that, *if* there exists j such that, for all $k < 99$, $p(j + k) = 9$, *then* the first such j is even.
>
> The assumption that A is predeterminate from the viewpoint of player I thus leads to a conclusion for which we have no evidence.

A subset C of $\mathbb{N}$ is a *decidable subset of* $\mathbb{N}$ if and only if there exists α such that, for every n, n belongs to C if and only if $\alpha(n) = 1$. The intuitionistic mathematician does not require that α is given by means of an algorithm.

It will be clear how to extend this notion to subsets of $\{0,1\} \times \mathbb{N}$. Observe that the set A in the counterexample just given is a decidable subset of $\{0,1\} \times \mathbb{N}$.

Also note that an anti-strategy for player *II* in $\{0,1\} \times \mathbb{N}$ is the same as a strategy for player *II* in $\{0,1\} \times \mathbb{N}$. Therefore, every game in $\{0,1\} \times \mathbb{N}$ is determinate from the viewpoint of player *II*. We may conclude that there are subsets of $\{0,1\} \times \mathbb{N}$ that are determinate from the viewpoint of player *II* while we are unable to prove that they are determinate from the viewpoint of player *I*.

12.1.9 In Section 12.1.10 we intend to discuss a second class of two-move-games. We will be led to use the *Fan Theorem*. In the literature, the expression "Fan Theorem" is not used unequivocally, and, for this reason, we introduce two precise versions of the theorem in Section 12.1.9.1. In Section 12.1.9.2 we prove a small combinatorial lemma that will be useful in Section 12.1.10.

12.1.9.1 Let σ be a spread-law. σ is called a *finitary spread-law* or a *fan-law* if and only if for each a, if σ admits a, then there are only finitely many numbers n such that σ admits $a * \langle n \rangle$. The set of all infinite sequences obeying a fan-law is called a *fan*.

Let X be a subset of $\mathcal{N}$ and let B be a subset of $\mathbb{N}$. We say: *B is a bar in X* if and only if every infinite sequence in X has an initial part in B. We say: B is *bounded* if and only if there exists n such that, for each b in B, $length(b) \leq n$. Here are two versions of Brouwer's Fan Theorem:

Unrestricted Fan Theorem: *Let σ be a fan and let B be a subset of $\mathbb{N}$ that is a bar in σ. There exists a bounded subset B' of B that is a bar in σ.*

Strict Fan Theorem: *Let σ be a fan and let B be a decidable subset of $\mathbb{N}$ that is a bar in σ. There exists a bounded subset B' of B that is a bar in σ.*

(The second version occurs as $^*26.6a$ in Kleene and Vesley (1965) and as FAN_D in Troelstra and van Dalen (1988).)

Brouwer's philosophical argument for the bar theorem seems to establish the unrestricted as well as the strict version of the Fan Theorem, see Veldman (2006b). Sometimes, one derives the Unrestricted Fan Theorem from the Strict Fan Theorem by means of the First Axiom of Continuous Choice, a special case of the Second Axiom of Continuous Choice. From a classical point of view, both versions of the Fan Theorem are reformulations of König's lemma. The usual formulation of König's lemma ("*Every infinite finitely-branching tree has an infinite branch*") is not valid intuitionistically.

12.1.9.2 For all natural numbers m, p we let $S(p,m)$ be the set of all numbers a such that $length(a) = m$ and for each $i < m$, $a(i) < p$.

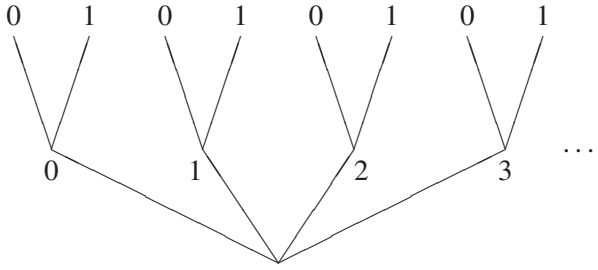
Lemma: *For all m, p, for each subset A of $\mathbb{N} \times \mathbb{N}$, if for all a in $S(p, m)$ there exists $i < m$ such that $(i, a(i))$ belongs to A, then there exists $i < m$ such that, for all $q < p$, (i, q) belongs to A.*

Proof. The proof uses induction on m. The case $m = 1$ is obvious.

Suppose that m is a natural number and that the case m has been established. Let p be a natural number and let A be a subset of $\mathbb{N} \times \mathbb{N}$ such that for all a in $S(p, m + 1)$ there exists $i < m + 1$ such that $(i, a(i))$ belongs to A.

Let a belong to $S(p, m)$. Observe that for each $q < p$, either for some $i < m$, $(i, a(i))$ belongs to A, or (m, q) belongs to A. Therefore, either, for some $i < m$, $(i, a(i))$ belongs to A, or, for all $q < p$, (m, q) belongs to A. Let B be the set of all pairs (i, j) of natural numbers such that either (i, j) belongs to A, or, for all $q < p$, (m, q) belongs to A. Observe that for all a in $S(p, m)$ there exists i such that $(i, a(i))$ belongs to B. Applying the induction hypothesis we find $i < m$ such that for all $q < p$, (i, q) belongs to B, that is, either for all $q < p$, (i, q) belongs to A, or for all $q < p$, (m, q) belongs to A.

12.1.10 We consider two-move-games of the following kind: Player *I* chooses a natural number, player *II* chooses 0 or 1, and the game is over.



A strategy for player *I* in a game like this consists of a single number, player *I*'s first and only move. A strategy for player *II*, on the other hand, is a function from $\mathbb{N}$ to $\{0, 1\}$, assigning to each natural number p the answer player *II* will give if player *I* opens the game with p. The set of strategies for player *II* is the set of all functions from $\mathbb{N}$ to $\{0, 1\}$. This set is a finitary spread, called: the *binary fan*, or: *(intuitionistic) Cantor space C*.

It turns out that every subset of $\mathbb{N} \times \{0, 1\}$ is determinate from the viewpoint of player *I* in the sense of Section 12.1.4:

> Let A be a subset of $\mathbb{N} \times \{0, 1\}$ such that for all α in C there exists n such that $(n, \alpha(n))$ belongs to A.
>
> Using the unrestricted Fan Theorem, we find m such that for all α in C there exists $n \leq m$ such that $(n, \alpha(n))$ belongs to A. Therefore, for all a in $S(2, m + 1)$ there exists $n < m + 1$ such that $(n, a(n))$ belongs to A.

Using the combinatorial lemma from Section 12.1.9.2 we find n such that both $(n, 0)$ and $(n, 1)$ belong to A, and this number obviously is a winning strategy for player I.

In the special case that A is a decidable subset of $\mathbb{N} \times \{0, 1\}$, we obtain the conclusion without using the Fan Theorem, as follows:

Let A be a decidable subset of $\mathbb{N} \times \{0, 1\}$. We use the First Axiom of Countable Choice and define α such that for every n, $\alpha(n) := 1$ if $(n, 0)$ belongs to A and $\alpha(n) := 0$, if $(n, 0)$ does not belong to A. We determine n such that $(n, \alpha(n))$ belongs to A and conclude that $\alpha(n) = 1$ and that both $(n, 0)$ and $(n, 1)$ belong to A.

The next case to consider is that A is not a decidable, but an *enumerable* subset of $\mathbb{N} \times \{0, 1\}$, that is, there exists a function β in $\mathcal{N}$ such that, for each n, i, (n, i) belongs to A if and only if there exists p such that $\beta(\langle n, i, p\rangle) = 0$. The statement that every enumerable subset of $\mathbb{N} \times \{0, 1\}$ is determinate in the above sense is an *equivalent* of the strict Fan Theorem, see Veldman (2005), that is, in a weak formal system BIM for basic intuitionistic analysis introduced in Veldman (2005) the strict Fan Theorem is equivalent to the statement that every enumerable subset of $\mathbb{N} \times \{0, 1\}$ is determinate from the viewpoint of player I. The stronger statements we are to prove in this paper, Lemma 2.2, Corollaries 2.4, 2.5, Lemma 3.3 and Theorem 3.5 also are equivalents of the strict Fan Theorem, see Veldman (2005).

The result that the Fan Theorem implies that every subset of $\mathbb{N} \times \{0, 1\}$ is determinate from the viewpoint of player I in the sense of 12.1.4 occurs already in Section 4 of Veldman (1982). Following a suggestion by J.R. Moschovakis (see Moschovakis, 1980a), we gave here a slightly different proof.

12.1.11 We describe the contents of the remaining sections. In Section 12.2 we introduce *II*-finitary spreads, that is, spreads in which player *II* has only finitely many possibilities for each one of his moves. Using the Fan Theorem, we show that in such spreads, closed sets and open sets are predeterminate from the viewpoint of player I (in the sense of Section 12.1.7).

In Section 12.3 we prove the much stronger result that *every* subset of a *II*-finitary spread is predeterminate from the viewpoint of player I. A slightly different version of this main result occurs already in Chapter 16 of Veldman (1981). In Section 12.4, we give two applications of the main result.

The reader who wants to enjoy the classical story of the notion of determinacy may consult Moschovakis (1980b) and Kechris (1995).

12.2 The safe-move-lemma and the determinacy of closed sets and open sets in *II*-finitary spreads

12.2.1 In this subsection we introduce some notations and some terminology.

Let σ be a spread and let a be a natural number admitted by σ, that is, such that $\sigma(a) = 0$. We define the spread-law $\sigma{\downarrow}a$ by: for all b in $\mathbb{N}$, $(\sigma{\downarrow}a)(b) = \sigma(a * b)$.

Let σ be a spread, and let γ be a strategy for player *II* in σ. Let a be a natural number such that $\sigma(a) = 0$ and $length(a)$ is even, and let δ be a strategy for player *II* in $\sigma{\downarrow}a$. We define: γ *extends* δ, or: δ *extends to* γ, if and only if, a *II*-obeys γ and, for each b in $\mathbb{N}$, if $(\sigma{\downarrow}a)(b) = 0$ and $length(b)$ is odd, then $\delta(b) = \gamma(a * b)$.

Let σ be a spread and let ζ be an anti-strategy for player *I* in σ. Let a be a natural number such that $\sigma(a) = 0$ and $length(a)$ is even. We define: a is ζ*-safe* if and only if every strategy δ for player *II* in the spread $\sigma{\downarrow}a$ extends to a strategy γ for player *II* in the spread σ such that $\zeta|\gamma$ passes through a.

Let σ be a spread. We define: σ is *II-finitary* if and only if, for each a, if σ admits a and $length(a)$ is odd, then there exists n such that, for every m, if σ admits $a * \langle m \rangle$, then $m < n$.

Observe that, if σ is a *II*-finitary spread, then player *II* has only finitely many possibilities for each one of his moves. Therefore, for each strategy γ for player *II* in σ, for each a in $\mathbb{N}$, if $\sigma(a) \neq 0$ or $length(a)$ is even, then $\gamma(a) = 0$, and if $\sigma(a) = 0$ and $length(a)$ is odd, then there are finitely many possible values for $\gamma(a)$. This shows that, if σ is a *II*-finitary spread, then $\mathrm{Strat}_{II}(\sigma)$ is a fan.

12.2.2 **The safe-move-lemma.** *Let σ be a II-finitary spread and let ζ be an anti-strategy for player I in σ. Then:*

(i) *The set of all natural numbers a such that $\sigma(a) = 0$ and* $\mathrm{length}(a)$ *is even and a is ζ-safe is a decidable subset of $\mathbb{N}$.*

(ii) *For every natural number a, if $\sigma(a) = 0$,* $\mathrm{length}(a)$ *is even and a is ζ-safe, then there exists n such that $\sigma(a * \langle n \rangle) = 0$ and, for all m, if $\sigma(a * \langle n, m \rangle) = 0$, then $a * \langle n, m \rangle$ is ζ-safe.*

Proof: Let σ, ζ fulfill the conditions of the lemma.

(i) Let a be a natural number such that $\sigma(a) = 0$ and $length(a)$ is even. Using the strict Fan Theorem, we calculate a natural number N such that for all strategies γ, δ for player *II* in σ, if $\overline{\gamma}N = \overline{\delta}N$, then, for each $i < length(a)$, $(\zeta|\gamma)(i) = (\zeta|\delta)(i)$.

Consider the set B consisting of all natural numbers $\overline{\gamma}N$, where γ is a strategy for player *II* in σ, and observe that B is a finite set of natural numbers.

Let δ be a strategy for player *II* in the spread $\sigma{\downarrow}a$. Considering the number $\overline{\delta}N$ and the set B we may decide whether δ extends to a strategy γ for player *II* in σ such that $\zeta|\gamma$ passes through a, or not. If δ does so indeed, we say that δ *fits* a. Observe that, for all strategies δ, ε for player *II* in the spread $\sigma{\downarrow}a$, if δ fits a and $\overline{\delta}N = \overline{\varepsilon}N$, then ε fits a.

Also observe that the set of all natural numbers $\overline{\delta}N$, where δ is a strategy for player *II* in the spread $\sigma{\downarrow}a$, is a finite set of natural numbers. Therefore, we may decide if it is true that every strategy δ for player *II* in the spread $\sigma{\downarrow}a$ fits a, or not. If so, then a is ζ-safe, and, if not, then a is not ζ-safe.

(ii) Let a be a natural number such that $\sigma(a) = 0$, $length(a)$ is even and a is ζ-safe. Using the strict Fan Theorem we calculate a natural number N such that for each strategy γ for player *II* in σ, $(\zeta|\gamma)(length(a)) < N$. Observe that for each n, m, if $n \geq N$, then $a * \langle n, m\rangle$ is not ζ-safe.

We have to prove that there exists $n < N$ such that σ admits $a * \langle n\rangle$ and, for all m, if σ admits $a * \langle n, m\rangle$, then $a * \langle n, m\rangle$ is ζ-safe. Because of (i) we may argue by contradiction.

Let us assume that, for every n, if $n < N$ and σ admits $a * \langle n\rangle$, then there exists m such that $a*\langle n, m\rangle$ is not ζ-safe. Let $n_0, n_1, \ldots, n_{k-1}$ be an enumeration of the natural numbers n such that $n < N$ and σ admits $a * \langle n\rangle$. Determine $m_0, m_1, \ldots, m_{k-1}$ in $\mathbb{N}$ such that, for all $i < k$, $\sigma(a*\langle n_i, m_i\rangle) = 0$ and $a*\langle n_i, m_i\rangle$ is not ζ-safe. Determine, for each $i < k$, a strategy δ_i for player *II* in $\sigma{\downarrow}(a*\langle n_i, m_i\rangle)$ such that δ_i does not fit $a * \langle n_i, m_i\rangle$. Let γ be a strategy for player *II* in $\sigma{\downarrow}a$ be such that, for each $i < k$, γ extends δ_i and $\gamma(a * \langle n_i\rangle) = m_i$.

As a is ζ-safe, we may determine a strategy γ' for player *II* in σ, extending the strategy γ, and such that $\zeta|(\gamma')$ passes through a. But then there exists $i < k$ such that $\zeta|(\gamma')$ passes through $a * \langle n_i, m_i\rangle$ and this contradicts the fact that γ' extends δ_i and δ_i does not fit $a * \langle n_i, m_i\rangle$.

We thus see that there exists $n < N$ such that σ admits $a * \langle n\rangle$ and, for all m, if σ admits $a * \langle n, m\rangle$, then $a * \langle n, m\rangle$ is ζ-safe.

12.2.3 Let σ be a spread and let A be a subset of σ. We define: A is an *open* subset of σ if and only if there exists a decidable subset C of $\mathbb{N}$ such that for every α in σ, α belongs to A if and only if, for some n, $\overline{\alpha}n$ belongs to C.

We define: A is a *closed* subset of σ if and only if there exists a decidable subset C of $\mathbb{N}$ such that for all α, α belongs to A if and only if, for each n, $\overline{\alpha}n$ belongs to C.

If A itself is a spread, then A is a closed subset of σ, but not every closed subset of σ is a spread, see Veldman (1981). The reason is that, given a decidable subset C of $\mathbb{N}$, it is not always possible to decide if there exists α such that for every n, $\overline{\alpha}n$ belongs to C.

For each a, n such that $n \leq length(a)$, we let $\overline{a}(n)$ be the code number of the finite sequence $(a(0), a(1), \ldots, a(n-1))$.

For each a, for each γ, we define: *a I-obeys γ*, if and only if, for each n, if $2n < \mathit{length}(a)$, then $a(2n) = \gamma(\overline{a}(2n))$.

Similarly, for each a, for each γ, we define: *a II-obeys γ*, if and only if, for each n, if $2n + 1 < \mathit{length}(a)$, then $a(2n + 1) = \gamma(\overline{a}(2n + 1))$.

12.2.4 Corollary. *In II-finitary spreads, closed sets are predeterminate from the viewpoint of player I.*

Proof: Let σ be a *II*-finitary spread and let A be a closed subset of σ. Let C be a decidable subset of $\mathbb{N}$ such that for all α, α belongs to A if and only if, for all n, $\overline{\alpha}n$ belongs to C. Let ζ be an anti-strategy for player I in σ such that for every strategy γ for player II in σ, $\zeta|\gamma$ belongs to A.

We apply the safe-move-lemma 3.6 and determine a strategy γ for player I in σ such that for every a, if σ admits a and $\mathit{length}(a)$ is even and a is ζ-safe, then σ admits $a * \langle\gamma(a)\rangle$, and, for each m, if σ admits $a * \langle\gamma(a), m\rangle$, then $a * \langle\gamma(a), m\rangle$ is ζ-safe.

As the empty sequence $\langle\ \rangle$ is ζ-safe, every α that I-obeys γ will have the property that, for each n, $\overline{\alpha}(2n)$ is ζ-safe. Observe that, for each a, if $\mathit{length}(a)$ is even and a is ζ-safe, then every initial part of a belongs to C. It follows that every α that I-obeys γ belongs to A.

12.2.5 Corollary. *In II-finitary spreads, open sets are determinate from the viewpoint of player I.*

Proof: Let σ be a *II*-finitary spread and let A be an open subset of σ. Let C be a decidable subset of $\mathbb{N}$ such that, for every α in σ, α belongs to A if and only if, for some n, $\overline{\alpha}n$ belongs to C. Suppose that every strategy for player II in σ *II*-governs at least one element of A. Note that, for every strategy γ for player II in σ,v there exists a in C such that a *II*-obeys γ. Applying the strict Fan Theorem, we find N in $\mathbb{N}$ such that for every strategy γ for player II in σ, there exist a such that $a \leq N$ and a belongs to C and a *II*-obeys γ.

Let B be the set of all α in σ such that, for some n, $\overline{\alpha}n \leq N$ and $\overline{\alpha}n$ belongs to C. observe that B is a closed subset of σ and a subset of A. We now define an anti-strategy ζ for player I in σ, as follows. Let γ be a strategy for player II in σ. Let b be the least a such that a *II*-obeys γ and a belongs to C. Let $\zeta|\gamma$ be the sequence β passing through b such that β *II*-obeys γ and for each n, if $2n \geq \mathit{length}(b)$, then $\beta(2n)$ is the least p such that σ admits $\overline{\beta}(2n) * \langle p\rangle$.

It will be clear that, for each strategy γ for player II in σ, the sequence $\zeta|\gamma$ *II*-obeys γ and belongs to B. Applying Corollary 2.4 we conclude that there is a strategy for player I in σ that wins the set B for player I, and therefore also the set A.

12.3 The safe-conjecture-lemma and the intuitionistic determinacy theorem

12.3.1 In Section 12.2 we have seen that, in a *II*-finitary spread σ, if the payoff set A is closed or open, every anti-strategy securing the set A for player I may be effectively transformed in a strategy winning the set A for player I.

In this section, we will strengthen this result considerably: we show that, in any *II*-finitary spread σ, any anti-strategy ζ for player I may be effectively transformed in a strategy γ for player I with the property that to any play α in σ I-obeying γ one may effectively construct a strategy δ for player II such that $\alpha = \zeta|\delta$.

It is not difficult to see that this result solves the determinacy problem for *II*-finitary spreads: every subset of a *II*-finitary spread is predeterminate from the viewpoint of player I.

12.3.2 Let σ be a spread and let ζ be an anti-strategy for player I in σ. We want to refine the notion of a "ζ-safe position", introduced in Section 12.2.1.

Let a be a natural number admitted by σ such that $\mathit{length}(a)$ is even, and let c be a natural number. We define: *a is ζ-safe with conjecture c* if and only if each strategy for player II in the spread $\sigma{\downarrow}a$ extends to a strategy γ for player II in the spread σ *passing through* c such that $\zeta|\gamma$ passes through a.

12.3.3 The safe-conjecture-lemma. *Let σ be a II-finitary spread and let ζ be an anti-strategy for player I in σ.*

(i) *For each c, the set of all natural numbers a such that* $\mathrm{length}(a)$ *is even and $\sigma(a) = 0$ and a is ζ-safe with conjecture c is a decidable subset of* $\mathbb{N}$.

(ii) *For all natural numbers a, c, if $\sigma(a) = 0$,* $\mathrm{length}(a)$ *is even and a is ζ-safe with conjecture c, then there exists n such that $\sigma(a * \langle n \rangle) = 0$ and, for all m, if $\sigma(a * \langle n, m \rangle) = 0$, then $a * \langle n, m \rangle$ is ζ-safe with conjecture c.*

(iii) *For all natural numbers a, c, if $\sigma(a) = 0$,* $\mathrm{length}(a)$ *is even and a is ζ-safe with conjecture c, then, for every strategy δ for player II in the spread $\sigma{\downarrow}a$ there exists d, n such that* $\mathrm{length}(d)$ *is even and d II-obeys δ and $a * d$ is ζ-safe with conjecture $c * \langle n \rangle$.*

Proof: Let σ, ζ fulfill the conditions of the lemma.

(i), (ii): We omit the proofs, as they are similar to the proofs of the corresponding statements in Lemma 2.2.

(iii) Let a, c be natural numbers such that $\sigma(a) = 0$ and $\mathit{length}(a)$ is even and a is ζ-safe with conjecture c. Let δ be a strategy for player II in the spread $\sigma{\downarrow}a$. We determine a strategy γ for player II in the spread σ such that γ extends δ

and γ passes through c and $\zeta|\gamma$ passes through a. We determine p such that for every strategy ε for player *II* in σ, if ε passes through $\overline{\gamma}p$, then $\zeta|\varepsilon$ passes through a.

We now consider $n := \gamma(\mathit{length}(c))$. Observe that γ passes through $c * \langle n \rangle$.

Let m be the greatest one of the two numbers p, $\mathit{length}(c) + 1$. Observe that for every strategy β for player *II* in the spread $\sigma{\downarrow}a$, if β passes through $\overline{\delta}m$, then there exists a strategy η for player *II* in the spread σ such that η extends β and η passes through $\overline{\gamma}m$.

We define $k := 2m + \mathit{length}(a)$. We let B be the set of all numbers $\overline{(\zeta|\eta)}k$, where η is a strategy for player *II* in the spread σ passing through $\overline{\gamma}m$ and extending a strategy δ' for player *II* in the spread $\sigma{\downarrow}a$ with the property: for each e, if $\sigma(a * e) = 0$ and $\mathit{length}(e) < 2m$, then $\delta'(e) = \delta(e)$. The set of all such strategies η is a fan and it follows from the strict Fan Theorem that B is a finite subset of $\mathbb{N}$. Remark that for each b in B there exists d such that $b = a * d$ and $\mathit{length}(d)$ is even and d *II*-obeys δ.

We claim that some member of B must be ζ-safe with conjecture $c * \langle n \rangle$. Because of (i) we may argue by contradiction.

Assume that no member of B is ζ-safe with conjecture $c * \langle n \rangle$. We then choose for each b in B a strategy δ_b for player *II* in the spread $\sigma{\downarrow}b$ such that δ_b does not extend to a strategy η for player *II* in σ with the property that η passes through $c * \langle n \rangle$ and $\zeta|\eta$ passes through b. We then form a strategy β for player *II* in $\sigma{\downarrow}a$ passing through $\overline{\delta}m$ such that, for each b in B, β extends δ_b, and, for each e, if $\sigma(a * e) = 0$ and $\mathit{length}(e) < 2m$, then $\beta(e) = \delta(e)$. We let ε be a strategy for player *II* in σ extending β and passing through $c * \langle n \rangle$ such that $\zeta|\varepsilon$ passes through a.

Consider $b := \overline{(\zeta|\varepsilon)}(k)$ and remark: b belongs to B and ε extends δ_b and ε passes through $c * \langle n \rangle$ and $\zeta|\varepsilon$ passes through b. Contradiction.

We conclude that some element of B must be ζ-safe with conjecture $c * \langle n \rangle$. Let b be such an element of B. Determine d such that $b = a * d$. Observe that $\mathit{length}(d)$ is even and d *II*-obeys δ and that we have obtained the desired conclusion.

12.3.4 For each α, for each n, we let α^n be the element β of $\mathcal{N}$ such that, for all m, $\beta(m) = \alpha(\langle n, m \rangle)$. In the proof of our main theorem, we use the following axiom:

Second Axiom of Countable Choice: *For each subset R of $\mathbb{N} \times \mathcal{N}$, if for each n there exists α such that $nR\alpha$, then there exists α such that, for each n, $nR\alpha^n$.*

This axiom, occurring as $^*2.1$ in Kleene and Vesley (1965), as $\mathbf{AC_{01}}$ in Gielen et al. (1981) and as *AC-NF* in Troelstra and van Dalen (1988), is a consequence of the Second Axiom of Continuous Choice, that we mentioned in Section 12.1.5.

Unlike the Second Axiom of Continuous Choice, the Second Axiom of Countable Choice is, from a classical point of view, a sensible assumption.

12.3.5 Intuitionistic Determinacy Theorem. *Let σ be a II-finitary spread. Every subset of σ is predeterminate from the viewpoint of player I.*

Proof: Let σ be a *II*-finitary spread. Let A be a subset of σ and let ζ be an anti-strategy for player I in σ securing the set A for player I. We prove that there exist a strategy for player I in σ with the property that, for every α in σ, if α I-obeys γ, then there exists a strategy δ for player II in σ such that α coincides with $\zeta|\delta$. Obviously, the strategy γ then wins the set A for player I.

According to Lemma 3.3 and Corollary 2.5 we may determine, for each a, c such that $\sigma(a) = 0$, $length(a)$ is even and a is ζ-safe with conjecture c, a strategy γ for player I in $\sigma{\downarrow}a$ with the property that for every α in σ I-obeying γ there exist p, n such that $a * \overline{\alpha}(2p)$ is ζ-safe with conjecture $c * \langle n\rangle$.

Let B be the set of all numbers $\langle a, c\rangle$ in $\mathbb{N}$ such that $\sigma(a) = 0, length(a)$ is even and a is ζ-safe with conjecture c. According to Lemma 3.3, B is a decidable subset of $\mathbb{N}$.

Using the Second Axiom of Countable Choice we determine ε in $\mathcal{N}$ with the property that, for each $\langle a, c\rangle$ in B, $\varepsilon^{\langle a,c\rangle}$ is a strategy for player I in $\sigma{\downarrow}a$ such that for every α in $\sigma{\downarrow}a$ I-obeying $\varepsilon^{\langle a,c\rangle}$ there exist p, n such that $a * \overline{\alpha}(2p)$ is ζ-safe with conjecture $c * \langle\lambda(\langle a, c\rangle)\rangle$.

We now describe informally the strategy γ that player I should obey in σ.

> Observe that $\langle\ \rangle$ is ζ-safe with conjecture $\langle\ \rangle$. Define $\delta(0) = \lambda(\langle\langle\ \rangle, \langle\ \rangle\rangle)$. Follow the strategy $\varepsilon^{\langle\langle\ \rangle,\langle\ \rangle\rangle}$, until, in cooperation with player *II* a position $\overline{\alpha}(2n_0)$ is reached such that $n_0 > 0$ and, for some n, $\overline{\alpha}(2n_0)$ is ζ-safe with conjecture $\langle n\rangle$. Let $\delta(0)$ be the least such n.
>
> Follow the strategy $\varepsilon^{\langle\overline{\alpha}(2n_0),\ \langle\delta(0)\rangle\rangle}$ until, in cooperation with player *II*, a position $\overline{\alpha}(2n_1)$ is reached such that $n_1 > n_0$ and, for some n, $\overline{\alpha}(2n_1)$ is ζ-safe with conjecture $\langle\delta(0), n\rangle$. Let $\delta(1)$ be the least such n.
>
> And so on.

Lemma 3.2(ii) ensures that it is indeed possible for player I to ensure that $n_1 > n_0$ and $n_2 > n_1$, and so on.

Suppose that α belongs to σ and is played by player I according to this strategy and that δ is the sequence of conjectures formed by player I during the play. Observe that, for all n, there exists a strategy β for player II in σ passing through $\overline{\delta}n$ such that $\zeta|\beta$ passes through $\overline{\alpha}(2n)$. It follows that δ is a strategy for player II in σ with the property: $\zeta|\delta = \alpha$.

12.4 Two applications

12.4.1 For each a, b in $\mathbb{N}$ we define: *the finite sequence (coded by) a is an initial part of the finite sequence (coded by) b* , notation: $a \sqsubseteq b$, if and only if there exists $n \leq \mathit{length}(b)$ such that $a = \overline{b}n$.

For each a, b in $\mathbb{N}$ we define: a, b *form a branching*, notation: $a \perp b$, if and only if a is not an initial part of b and b is not an initial part of a.

Let A be a subset of $\mathcal{N}$. We consider the following game, sometimes called $\mathcal{G}^*(A)$, that has been devised by Morton Davis in Davis (1964).

Player I chooses $\langle \ell_0, r_0 \rangle$ in $\mathbb{N} \times \mathbb{N}$ such that $\ell_0 \perp r_0$.

↘

Player II chooses i_0 in $\{0, 1\}$.
We define $a_0 := \ell_0$ if $i_0 = 0$, and
$a_0 := r_0$ if $i_0 = 1$.

↙

Player I chooses $\langle \ell_1, r_1 \rangle$ in $\mathbb{N} \times \mathbb{N}$ such that $a_0 \sqsubseteq \ell_1, a_0 \sqsubseteq r_1$ and $\ell_1 \perp r_1$.

↘

Player II chooses i_1 in $\{0, 1\}$.
We define : $a_1 := \ell_1$ if $i_1 = 0$, and
$a_1 := r_1$ if $i_1 = 1$.

↙

Player I chooses $\langle \ell_2, r_2 \rangle$ in $\mathbb{N} \times \mathbb{N}$ such that $a_1 \sqsubseteq \ell_2, a_1 \sqsubseteq r_2$ and $\ell_2 \perp r_2$.

↘

Player II chooses i_2 in $\{0, 1\}$
We define $a_2 := \ell_2$ if $i_2 = 0$, and
$a_2 := r_2$ if $i_2 = 1$.

And so on.

In the end, we determine α in $\mathcal{N}$ such that, for all n, α passes through a_n. Player I wins if and only if α belongs to A.

It will be clear that $\mathcal{G}^*(A)$ may be described as a game in a II-finitary spread σ. It follows that for every subset A of $\mathcal{N}$, the game $\mathcal{G}^*(A)$ is predeterminate from the viewpoint of player I.

One may prove constructively that player I has a winning strategy in the game $\mathcal{G}^*(A)$ if and only if there exists an *embedding* of Cantor space C into A, that is: an element ζ of $\mathcal{N}$ coding a continuous function from C into A such that for all α, β in C, if there exists n such that $\alpha(n) \neq \beta(n)$, then there exists p such that $(\zeta|\alpha)(p) \neq (\zeta|\beta)(p)$.

One may prove constructively that, if the set A is *enumerable*, that is, if there exists an element α of $\mathcal{N}$ such that every element of A occurs in the sequence $\alpha^0, \alpha^1, \alpha^2, \ldots$, then player II has a strategy ensuring that the result of a play in $\mathcal{G}^*(A)$ will not belong to A: he makes his n-th move such that the result will differ from α^n.

Classically, it is also true that if player *II* has a successful strategy in $\mathcal{G}^*(A)$, then the set A is at most enumerable. The argument is unconstructive, but, as we hope to show in a future paper, one may prove an intuitionistic counterpart to this result, using Brouwer's Thesis on bars.

The statement that all games $\mathcal{G}^*(A)$ are predeterminate from the viewpoint of player I is an intuitionistic theorem that is in some sense related to the continuum hypothesis, like the different theorem in Section 2 of Gielen et al. (1981), to which it forms a kind of counterpart.

12.4.2 Let A be a subset of the set $\mathbb{Q}$ of rational numbers. We consider the following game that we call $\mathcal{H}(A)$, the letter $\mathcal{H}$ honouring F. Hausdorff.

Player I chooses q_0 in $\mathbb{Q}$.

↘

Player *II* chooses i_0 in $\{0, 1\}$.
We define $H_0 := \{q \in \mathbb{Q} \mid q < q_0\}$ if $i_0 = 0$, and
$H_0 := \{q \in \mathbb{Q} \mid q > q_0\}$ if $i_0 = 1$.

↙

Player I chooses q_1 in H_0.

↘

Player *II* chooses i_1 in $\{0, 1\}$.
We define $H_1 := H_0 \cap \{q \in \mathbb{Q} \mid q < q_1\}$ if $i_1 = 0$, and
$H_1 := H_0 \cap \{q \in \mathbb{Q} \mid q > q_1\}$ if $i_1 = 1$.

↙

Player I chooses q_2 in H_1.

↘

Player *II* chooses i_2 in $\{0, 1\}$.
We define $H_2 := H_1 \cap \{q \in \mathbb{Q} \mid q < q_2\}$ if $i_2 = 0$, and
$H_2 := H_1 \cap \{q \in \mathbb{Q} \mid q > q_2\}$ if $i_2 = 1$.

and so on.

In the end, player I wins if and only if, for each n, q_n belongs to A.

The game $\mathcal{H}(A)$ may be described as a game in a *II*-finitary spread σ. Thus, Theorem 3.3 applies, and, for every subset A of $\mathbb{Q}$, the game $\mathcal{H}(A)$ is predeterminate from the viewpoint of player I.

Observe that player I has a winning strategy in $\mathcal{H}(A)$ if and only if there exists an order-preserving embedding of $(\mathbb{Q}, <)$ into $(A, <)$.

From a classical point of view, the game $\mathcal{H}(A)$ is determinate as it is a closed game, and the class of all subsets A of $\mathbb{Q}$ such that player *II* has a winning strategy in the game $\mathcal{H}(A)$ coincides with the class of all *scattered* subsets of $\mathbb{Q}$, that is, the class of all subsets A of $\mathbb{Q}$ such that it is impossible to embed $(\mathbb{Q}, <)$ into $(A, <)$.

Intuitionistically, it seems wise to restrict oneself to *decidable* subsets A of $\mathbb{Q}$. The statement "*player II has a strategy in the game $\mathcal{H}(A)$, such that, for any resulting sequence $q_0, q_1, q_2, \ldots$, some n may be found with the property $q_n \notin A$*" turns out to be equivalent to "*A is very discrete*", as we hope to show in a future paper. The argument uses Brouwer's Thesis on bars, see Veldman (2006a). The notion of a very discrete subset of $\mathbb{Q}$ is defined inductively, see Rosenstein (1982). A subset A of $\mathbb{Q}$ is *very discrete* if either $A = \emptyset$ or A contains exactly one number, or there exists a sequence $\ldots, A_{-2}, A_{-1}, A_0, A_1, A_2, \ldots$ of very discrete sets such that, for each i in $\mathbb{Z}$, for each q in A_i, for each r in A_{i+1}, $q < r$, and $A = \bigcup_{i \in \mathbb{Z}} A_i$. Scattered sets were first studied by F. Hausdorff (see Hausdorff, 1908) and Rosenstein (1982).

Acknowledgments

I want to express my gratitude to the referee of an earlier version of this paper. He not only detected several inaccuracies but also gave valuable advice leading to further improvements of the paper. I also want to thank my student Takako Nemoto for noticing some inaccuracies in the proofs of Lemma 3.3 and Theorem 3.5.

References

Carroll, L. (1865). *Alice's Adventures in Wonderland, Through the Looking-Glass, and Other Writings*. Collins, London, Glasgow.

Davis, M. (1964). Infinite games of perfect information. *Ann. Math. Stud.*, 52:85–101.

Gielen, W., de Swart, H., and Veldman, W. (1981). The continuum hypothesis in intuitionism. *J. Symb. Logic*, 46:121–136.

Hausdorff, F. (1908). Grundzüge einer theorie der geordneten mengen. *Math. Ann.*, 65:435–505.

Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Springer, New York.

Kleene, S. C. and Vesley, R. E. (1965). *The Foundations of Intuitionistic Mathematics, Especially in Relation to Recursive Functions*. Studies in Logic and the Foundations of Mathematics. North Holland, Amsterdam.

Moschovakis, J. R. (1980a). Review of Veldman, 1982. In *Mathematical Reviews*, **MR** 85g:03089.

Moschovakis, Y. N. (1980b). *Descriptive Set Theory*, volume 100 of *Studies in Logic and the Foundations of Mathematics*. North Holland, Amsterdam.

Rosenstein, J. G. (1982). *Linear Orderings*. Academic, New York.

Troelstra, A. S. and van Dalen, D. (1988). *Constructivism in Mathematics, an Introduction. Volumes I and II*, volumes 121 and 123 of *Studies in Logic and the Foundations of Mathematics*. North Holland, Amsterdam.

Veldman, W. (1981). *Investigations in intuitionistic hierarchy theory*. PhD thesis, Katholieke Universiteit Nijmegen, Nijmegen.

Veldman, W. (1982). On the contraposition of two axioms of countable choice. In Troelstra, A. S. and van Dalen, D., editors, *Brouwer Centenary Symposium*, volume 110, pages 513–523, North Holland, Amsterdam.

Veldman, W. (2005). The fan theorem as an axiom and as a contrast to Kleene's alternative. Report no. 0509, Department of Mathematics, Radboud University, Nijmegen.

Veldman, W. (2006a). The borel hierarchy and the projective hierarchy from Brouwer's intuitionistic perspective. Report no. 0604, Department of Mathematics, Radboud University, Nijmegen.

Veldman, W. (2006b). Brouwer's real thesis on bars. *Philosophia Scientiæ*, 6:21–42.

Symbol Index

Subject Index

Name Index