Logic, Argumentation & Reasoning 5

Erik Weber
Dietlinde Wouters
Joke Meheus *Editors*

Logic, Reasoning, and Rationality

Springer

Logic, Argumentation & Reasoning

Interdisciplinary Perspectives from the Humanities
and Social Sciences

Volume 5

Series Editor

Shahid Rahman

For further volumes:
http://www.springer.com/series/11547

Logic, Argumentation & Reasoning

The series is developed in partnership with the Maison Européenne des Sciences de l'Homme et de la Société (MESHS) at Nord - Pas de Calais and the UMR-STL: 8163 (CNRS). Aims & Scope: The scientific objectives of the series, where humanities and social sciences are conceived as building interdisciplinary interfaces, are:

This series publishes volumes that link practices in the Humanities and Social Sciences, with theories in Logic, Argumentation and Reasoning, such as: Decision theory and Action theory, Argumentation Theories in: cognitive sciences, economy, sociology, law, logic, philosophy of sciences. The series is open towards research from the Analytic and the Continental traditions, and has four main focus areas: Pragmatic models and studies that develop a dynamic approach to reasoning in which argumentation is structured as an interaction or as a game, in which two or more participants' play moves are defined by the type of argumentation in question, communication, language and techniques of argumentation; studies between the practical and theoretical dimensions of argumentation, as well as the relationships between argumentation and other modes of communication, reception, persuasion and power; studies in which reasoning practice is considered from the point of view of its capacity to produce conviction of persuasion, and focusing on understanding what makes an argument performative; Diachronic transformations of reasoning practices studies that emphasize the invention and renewal of reasoning forms, with respect to its performance and its effectiveness.

General Editor
Shahid Rahman (Lille, UMR 8163)

Managing Editor
Laurent Keiff (Lille, UMR 8163)

Area Editors

Argumentation and Pragmatics
Marcelo Dascal (Tel Aviv)
Erik Krabbe (Groningen)
Frans H. van Eemeren (Amsterdam)
John Woods (British Columbia/King's College)

Argumentation and Rhetoric
Fabienne Blaise (Lille, MESHS-Nord Pas de Calais)
Gabriel Galvez-Behar (Lille, MESHS-Nord Pas de Calais)
André Laks (Sorbonne, Paris IV)
Ruth Webb (Lille, UMR 8163)

Decision Theory, Mathematics, Economy
Jacques Dubucs (IHPST-Paris 1)
Fredéric Jouneau (Lille)
Richard Sobel (Lille)

Cognitives Sciences. Computer Sciences
Yann Coello (Lille)
Eric Gregoire (CRIL-Lens)
Henry Prakken (Utrecht)
François Recanati (ENS, Paris)

Epistemology and Philosophy of Science
André Fuhrmann (Frankfurt)
Gerhard Heinzmann (Nancy)
Göran Sundholm (Leyden)

Logic
Michel Crubellier (Lille, UMR 8163)
Dov Gabbay (King's College)
Sven Ove Hansson (Stockholm)
Tero Tulenheimo (Lille, UMR 8163)

Political Science and Sociology
Jean-Gabriel Contamin (Lille)
Franck Fischer (Rutgers)
Josh Ober (Stanford)
Marc Pichard (Lille, MESHS-Nord Pas de Calais)

Erik Weber • Dietlinde Wouters • Joke Meheus
Editors

Logic, Reasoning, and Rationality

Springer

Editors
Erik Weber
Dietlinde Wouters
Joke Meheus
Centre for Logic and Philosophy of Science
Ghent University (UGent)
Ghent, Belgium

ISSN 2214-9120 ISSN 2214-9139 (electronic)
ISBN 978-94-017-9010-9 ISBN 978-94-017-9011-6 (eBook)
DOI 10.1007/978-94-017-9011-6
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014945250

© Springer Science+Business Media Dordrecht 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

According to the members of the Vienna Circle, there was a strong connection between logic, reasoning, and rationality. They believed that human reasoning (and in particular scientific reasoning) is rational in so far as it is based on logic (which meant for them classical logic). It was also believed that scientific reasoning (for them the hallmark of human reasoning) was in general rational. In the second half of the twentieth century, both beliefs came under attack.

One of the motors for this change was the turn in history of science initiated by Alexandre Koyré. In the 'old history of science' success stories were told, usually on the basis of published papers and even textbooks, and only theories that had survived were considered (Galileo's law of free fall, Kepler's three laws, Newton's gravitation theory, and so on). Moreover, no attention was paid to mistaken paths, nor to the contexts in which the original theories were formulated and accepted. So, what happened was that nice and polished reconstructions of scientific episodes were made, with classical logic as the underlying logic, and that the results were deemed to be rational. In the 'new history of science', things changed radically. Theories were studied in their historical setting, and explicit attention was directed not only to theories that were abandoned (such as the phlogiston theory), but also to flaws, and to elements that played a crucial role in the construction of new theories, but that are today considered as non-rational. Examples are Kepler's work on astrology and on the harmony of the spheres, and Newton's work on alchemy.

In the aftermath of Koyré, philosophers of science, such as Hanson and Kuhn, also followed this new trend and started basing their philosophical analyses on actual examples from the history of science. Two central lessons came out of all this. First, the so-called context of justification, which was the sole concern of the members of the Vienna Circle, is less straightforward and less 'logical' than was traditionally accepted. Next, the 'context of discovery' is much more structured and methodical than was believed within the Vienna Circle, even though it is not understandable from the point of view of classical logic. The conclusion was that logic is inadequate to explicate actual examples of human reasoning, whether in the sciences or in everyday life.

There were several reactions to this situation. Some scholars held on to the link between (classical) logic and rationality, but concluded that scientific reasoning (especially as it occurs in the context of discovery) is inherently non-rational or even irrational. Others gave up the connection between logic and rationality. They looked for tools elsewhere (mainly in psychology and cognitive science) to analyse the rational character of scientific reasoning, often at the expense of rigour and formal accuracy. Times have changed, however. Today, a multiplicity of formal frameworks (ranging from non-classical logics over probability theory to Bayesian networks) is available in addition to classical logic. Also, historians and philosophers of science as well as psychologists have described a rich variety of patterns in both scientific and common sense reasoning.

The aim of the congress *Logic, Reasoning and Rationality* (Centre for Logic and Philosophy of Science, Ghent, 20–22 September 2010) was to stimulate the use of formal frameworks to explicate concrete examples of human reasoning, and conversely to challenge scholars in formal studies by presenting them with interesting new examples of actual reasoning. This book contains a selection of papers presented at the congress. Other papers presented at the congress have been published in special issues of the journals *Foundations of Science*, *Logic and Logical Philosophy*, *Logique et Analyse*, and *Philosophica*.

The first paper in this volume is by Diderik Batens. In *Adaptive Logics as a Necessary Tool for Relative Rationality. Including a Section on Logical Pluralism* he shows that adaptive logics are required by his epistemological stand. While doing so, he defies the reader to cope with the problems he is able to cope with. The last section of the paper contains a defense of a specific form of logical pluralism.

In *A New Approach to Epistemic Logic*, Giovanna Corsi and Gabriele Tassi introduce a new language for epistemic logic. They spell out the advantages of this approach. The semantics they present for this language is a generalization of the transition semantics, called epistemic transition semantics in which the possible worlds are states of affairs compatible with the epistemic state of some agent. A calculus is presented and shown to be complete with respect to epistemic transition semantics.

The contribution of Raoul Gervais is entitled *Explaining Capacities: Assessing the Explanatory Power of Models in the Cognitive Sciences*. It has been argued that only those models that describe the actual mechanisms responsible for a given cognitive capacity are genuinely explanatory. On this account, descriptive accuracy is necessary for explanatory power. This means that mechanistic models, which include reference to the components of the actual mechanism responsible for a given capacity, are explanatorily superior to functional models, which decompose a capacity into a number of sub-capacities without specifying the actual realizers. Gervais argues against this view by considering models in engineering contexts. Here, other considerations besides descriptive accuracy play a role. Often, the goal of performance trumps that of adequacy, and researchers are interested in how cognitive capacities *as such* can be realized, rather than how it is realized in a given system.

In *Data-driven Induction in Scientific Discovery. A Critical Assessment Based on Kepler's Discoveries*, Albrecht Heeffer provides a critical assessment of the model of data-driven induction for scientific discovery. The most influential research program using this model is developed by the BACON team. Two of the main claims by this research program, the descriptive and constructive power of data-driven induction, are evaluated by means of two historical case studies: the discovery of the sine law of refraction in optics and Kepler's third law of planetary motion. Heeffer provides evidence that the data used by the BACON program—despite the claims being made—do not correspond to the historical data available to Kepler and his contemporaries. He also shows that for the two cases the method by which the general law was arrived at did not involve data-driven induction, and he questions the value of the data-driven induction as a general model for scientific discovery.

In *Dovetailing Belief Base Revision with (Basic) Truth Approximation*, Theo A. F. Kuipers starts from recent work of Gustavo Cevolani, Vincenzo Crupi and Roberto Festa. They have shown that their account of verisimilitude of 'conjunctive theories' of a finite propositional language can be nicely linked to a variant of AGM belief set revision, viz. belief base revision, in the sense that the latter kind of revision is functional for truth approximation according to the conjunctive account. Kuipers offers a generalization of these ideas to the case of approaching any divide of a (finite or infinite) universe, allowing several interpretations, besides true (false) atomic propositions, notably nomic states (not) in equilibrium, nomic (im)possibilities, (non-)instantiated 'Q-predicates' of a monadic language. The generalization shows how and why approximation of 'the true boundary' takes place by belief base revision guided by evidence.

In their paper *A Method of Generating Modal Logics Defining Jaśkowski's Discussive D2 Consequence*, Marek Nasieniewski and Andrzej Pietruszczak study Jaśkowski's logic $\mathbf{D}_2$. This logic is usually understood as a set of discussive formulae. Studying Jaśkowski's paper one can also find a consequence relation (the $\mathbf{D}_2$-consequence). The logic $\mathbf{D}_2$ was meant to express this consequence relation. Since the logic $\mathbf{D}_2$ was formulated with the help of a modal logic, the consequence relation is also defined in the modal language. It is known that the logic $\mathbf{D}_2$ can be defined by other modal logics than **S5**. A similar question arises as regards the consequence relation. There are modal logics other than **S5** which define exactly the same consequence relation. Nasieniewski and Pietruszczak try to develop a more general method of defining modal logics which also allows to define the $\mathbf{D}_2$-consequence.

In *Frontier Theory of Inquiry: Apparent Conflicts between the Ghent Logical Program and the "Darwinian" Selectionist Program*, Thomas Nickles begins with an appreciation of several pragmatic aspects of Diderik Batens' research program. Then he turns to the apparent conflict with Donald Campbell's evolutionary epistemology, with its generalized Darwinian account of creative problem solving via mechanisms of blind variation plus selective retention. While there are significant differences of emphasis and style, he argues that the two are broadly compatible, just as the problem-solving approach of Herbert Simon and Allen Newell is broadly compatible with that of Campbell, also contrary to first appearances. The paper ends with some questions for the Ghent program concerning open problems.

In their paper *On the Propagation of Consistency in Some Systems of Paraconsistent Logic*, Hitoshi Omori and Toshiharu Waragai offer some results on the propagation of consistency in two systems of Logics of Formal Inconsistency (LFIs). One is the system **Bk** of Avron, which is an extension of the base system **mbC** of Carnielli, Coniglio and Marcos, and the other is an extension of **Bk** to the predicate calculus which will be referred to as **Bk***. Omori and Waragai present a new characterization of the consistency operator in **Bk**. This reflects the intuition of the consistency operator quite well. They then prove that two kinds of propagation of consistency known in the literature are actually equivalent to certain forms of de Morgan's laws without any occurrence of the consistency operator. Finally, they extend the result in **Bk** to **Bk***.

The paper of Francesco Orilia is entitled *Degrees of Validity and the Logical Paradoxes*. We traditionally accept a sharp distinction between deductive and inductive arguments. The former are taken to be undefeasible and thus we accept a principle that, roughly, goes as follows: (D) given a deductively valid argument with premises you believe, you should also believe the conclusion of the argument. Unfortunately, logical paradoxes such as the Liar, Russell's or Curry's cast doubts on (D). For, at least prima facie, they are deductively valid arguments and yet can lead to any conclusion we please, either directly (as in Curry's case) or via Ex Falso Quodlibet. Traditional reactions to this problem question either grammar (e.g., by invoking type-theoretical distinctions) or logic (by regarding as not really deductive some inference rules that are traditionally taken to be deductive) so as to claim that the paradoxes fail to be deductively valid after all. Orilia explores a different approach, based on the intuition that deductive arguments can be treated, on analogy with inductive arguments, as defeasible and as involving degrees of validity.

The aim of Graham Priest's paper *Contradictory Concepts* is to think through a raft of issues that the view known as 'dialetheism' raises. In particular, he is concerned with three inter-related questions: (1) Are the contradictions involved simply in our conceptual/linguistic representations, or are they in reality? And what exactly does this distinction amount to anyway? (2) Assuming that it is only in the former, can we get rid of them simply by changing these? (3) If we can, should we do so? Graham Priest takes up these issues in the three parts of his paper. The journey takes us through a number of important issues in metaphysics, semantics, and epistemology.

In *Bloody Analogical Reasoning*, Dagmar Provijn studies some of William Harvey's applications of analogies in the *Prelectiones Anatomiae Universalis* and the *Exercitatio anatomica de motu cordis et sanguinis in animalibus*. He shows that Harvey applied analogies in many different ways and that some contributed to the discovery of the characteristic 'action' of the heart and pulse and even to the discovery of the blood circulation. The discovery process is approached as a problem solving process as described in Batens' contextual model. The focus on constraints allows to see Harvey both as a modern scholar because of his extensive use of experimental results and as strongly influenced by an Aristotelian 'natural philosophy interpretation' of anatomy and physiology as, for instance, propagated by Fabricius of Aquapendente.

In their paper *Another Look at Mathematical Style, as inspired by Le Lionnais and the OuLiPo*, Jean Paul Van Bendegem and Bart Van Kerkhove offer a less monolithic interpretation of the (mathematical) style concept, in order for it to serve well as a methodological tool in the historiography of mathematics. Drawing inspiration from Le Lionnais, the Bourbaki movement and the French literary-mathematical OuLiPo movement, they introduce an approach along the path of a 'problem solving' conception of mathematics, thus creating room for mathematical style to be significantly 'more' than a mere mode of presentation of immutable content. In their view, this very same approach opens us the possibility for a fruitful comparison between mathematical and literary styles.

In *Internalism Does Entail Scepticism*, Jan Willem Wieland offers an insight into the discussion on the relation between Internalism and Skepticism. Internalism is the view that our inferences are justified depending on whether we have knowledge of the logical rules on which they are based, and Scepticism the view that none of our inferences are justified. Paul Boghossian has shown that Internalism entails Scepticism, Patrice Philie has attempted to block the entailment by an assumption on rationality. Wieland enforces the entailment claim and argues that Philie's solution misses the target: Internalism does entail Scepticism.

The last paper is Andrzej Wiśniewski's *Answering by Means of Questions in View of Inferential Erotetic Logic*. Inferential Erotetic Logic (IEL) gives an account of inferences in which questions play the role of conclusions, and proposes criteria of validity for these inferences. Wiśniewski shows that some tools elaborated within IEL are useful for the formal modeling of (a) replying with questions that are not clarification requests, and (b) question answering based on additional information actively sought for.

The congress was organised in honour of Diderik Batens. It served as an opportunity for him—on the verge of his retirement—to look back on his long and distinguished academic career and clarify his personal views to the audience. Among other things, Batens helped shape paraconsistent logic and was the founder of adaptive logics.

The editors are indebted to Wim Van Rie for his help in preparing the manuscript.

Ghent, Belgium
January 2014

Erik Weber
Joke Meheus
Dietlinde Wouters

Contents

1 **Adaptive Logics as a Necessary Tool for Relative Rationality: Including a Section on Logical Pluralism** 1
Diderik Batens

2 **A New Approach to Epistemic Logic** 27
Giovanna Corsi and Gabriele Tassi

3 **Explaining Capacities: Assessing the Explanatory Power of Models in the Cognitive Sciences** 45
Raoul Gervais

4 **Data-Driven Induction in Scientific Discovery: A Critical Assessment Based on Kepler's Discoveries** 59
Albrecht Heeffer

5 **Dovetailing Belief Base Revision with (Basic) Truth Approximation** 77
Theo A.F. Kuipers

6 **A Method of Generating Modal Logics Defining Jaśkowski's Discussive D_2 Consequence** 95
Marek Nasieniewski and Andrzej Pietruszczak

7 **Frontier Theory of Inquiry: Apparent Conflicts Between the Ghent Logical Program and the "Darwinian" Selectionist Program** 125
Thomas Nickles

8 **On the Propagation of Consistency in Some Systems of Paraconsistent Logic** 153
Hitoshi Omori and Toshiharu Waragai

9 **Degrees of Validity and the Logical Paradoxes** 179
Francesco Orilia

10 Contradictory Concepts 197
Graham Priest

11 Bloody Analogical Reasoning 217
Dagmar Provijn

12 Another Look at Mathematical Style, as Inspired by Le Lionnais and the OuLiPo 233
Jean Paul Van Bendegem and Bart Van Kerkhove

13 Internalism Does Entail Scepticism 247
Jan Willem Wieland

14 Answering by Means of Questions in View of Inferential Erotetic Logic 261
Andrzej Wiśniewski

Chapter 1
Adaptive Logics as a Necessary Tool for Relative Rationality: Including a Section on Logical Pluralism

Diderik Batens

1.1 Aim of This Paper

In most papers on adaptive logics, for example Batens (2004, 2007b), and in the forthcoming book Batens (2014), I try to remain philosophically neutral on whatever is not strictly relevant for adaptive logics. People with different political viewpoints may play the violin, or handle a hammer. Similarly, people with different philosophical viewpoints may apply the same adaptive logics, which may be sensibly classified as reasoning instruments. Tying those logics to my specific philosophical convictions would scare away some readers.

Of course, I have philosophical convictions. Especially those in the realm of epistemology motivated the origin and especially the development of the adaptive logic program. To make this link explicit seems useful. Adaptive logics clarify the notion of defeasible reasoning and highlight its importance. By doing so, they evoke a number of epistemological questions and rule out certain epistemological answers to these questions. The questions are far from specific for my epistemological views, but present interesting problems for any epistemological view.

So the potential interest of the present paper is double. Adaptive logics were developed in view of a philosophical need, which is to make a certain epistemological position meaningful and precise. These logics represent precise formulations of methods and actually of a multiplicity of alternative methods—plurality is easy from an adaptive perspective. The resulting problem for any epistemological stand is: In which way may such methods be integrated?

I shall begin by a, necessarily rough, sketch of my epistemological stand. This is meant as a point of reference and as an example of a stand in which defeasible reasoning, especially as approached by adaptive logics, finds its natural place.

D. Batens (✉)
Centre for Logic and Philosophy of Science, Ghent University (UGent), Ghent, Belgium
e-mail: Diderik.Batens@UGent.be

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_1,
© Springer Science+Business Media Dordrecht 2014

After a quick rehearsal of adaptive logics, formal problem-solving processes are introduced. In these, adaptive logics have their natural place and are able to show their strength. One or more adaptive logics are here combined with a deductive logic as well as with an erotetic logic.

In the next four sections, some typical features will be outlined. I chose those that are presumably most controversial and at the same time most difficult to incorporate within a formal framework. (i) Contextual meaning of logical terms is handled within a formal framework rather than within a linguistic one. The discussion concerns the way in which formal properties of the premises determine the meaning of occurrences of logical terms. (ii) The section on the meaning of logical symbols concerns the distinction that some want to draw between deductive and defeasible reasoning forms and the effects of this distinction on the meaning of logical terms. (iii) The next topic is the contextual meaning of non-logical terms. (iv) The traditional notion of a theory is confronted with an alternative in the section on complex theories. The central issue is that traditional theories may fail to be efficient means to embody the best available human knowledge. While complex theories have the disadvantages of their complexity, including their computational complexity, they are able to describe complex domains that are beyond the reach of traditional theories. (v) The section on logical pluralism is mainly meant to clarify a form of logical pluralism that does not coincide with most positions defended and attacked in the literature.

1.2 Epistemological Stand

This very compressed sketch will consist of a set of theses, each followed by one or more arguments. An extensive description of my epistemological stand is unfortunately only available in Batens (1992b) (and in its Greek translation). Some aspects are also discussed in Batens (1983, 1985, 1992a, 2000a) and Meheus and Batens (1996).

The first thesis reads: *All knowledge is ultimately defeasible.* Note that it says that all knowledge is defeasible, not that all reasoning is defeasible. Still, even non-defeasible reasoning starts always from defeasible premises, whence its conclusions are also defeasible. As we shall see, not all knowledge is defeasible in the same sense or in the same way.

Many will agree that most empirical knowledge is defeasible. Inductive generalizations clearly are, and so are predictions. Most results of abductive reasoning, which includes explanations, are also defeasible. The same holds for knowledge of causal relations, expectancies, results of diagnostic reasoning, and so on. While all this clearly holds for knowledge about the physical world, knowledge about other humans is no exception. Statements made by trustworthy human beings need to be interpreted, for example by means of Gricean maxims, which introduce additional defeasibility—I write "additional" because the statements already rely on defeasible knowledge of the person who utters the statement.

Let us turn to experience. That experience is never uninterpreted seems generally accepted today. I do not know any serious philosopher who identifies the set of experiential data with Mach's *Empfindungen*. Especially the philosophy of science of the second half of the twentieth century clarified this matter. Many reject the consequences attached to the insights of that period, especially the different forms of incommensurability. Yet nobody adduced any good arguments for questioning the insights themselves, which concern theory-ladenness. So even when we may have no reasons to mistrust our eyes or other senses, we may be interpreting what we see, etc., in a mistaken framework. So much more important than occasional optical or other illusions is the fact that every experience is interpreted and that this interpretation may be very mistaken, as the history of the sciences readily reveals.

The defeasible character of experience is enhanced by the fact that most so-called experience is the result of abduction. We think to see that it has rained because we think to see that the grass is wet, and we think to see that it has snowed because we think to see that our environment is covered with white stuff. We think to see a magpie because what we 'see' is compatible with what we know about magpies and not with what we know about any other species of birds. Such abductions occur usually in an unconscious way. Note that we are on a slippery slope here. There is a smooth transition from unconscious abduction to theory-laden observation.

Some will argue that we have means to obtain a higher certainty on our observations: repetition, instruments, experience, and so on—I return to these soon. However, the very fact that observations may be corrected and that means to obtain more reliable observations have been devised highlights the defeasibility of experience.

Methodological knowledge (norms and values in general) have long been claimed to be a priori. Today, informed people have changed their mind basically because history teaches us that methods are historically contingent and hence clearly defeasible. In trying to gather knowledge about the world, we learn how to learn. As Dudley Shapere candidly puts it: "what better basis could we have than what we have learned, including what we have learned about how to learn" (Shapere 2004, p. 52). Shapere couples this with the idea that science is content-guided. He opposes this idea to the views of the Vienna Circle—its members soon turned to logic[1]—as well as to the views of the 'post-classical' philosophers of science (Hanson, Kuhn, Toulmin, and Feyerabend)—these fell into extreme relativism.

Finally, I come to the most touchy and controversial point, logical and mathematical knowledge, even if I do not understand that any sane person could hold such knowledge to be non-defeasible. The point is not whether, for example, $0' + 0' = 0''$ is a theorem of Peano Arithmetic. The point is whether Peano Arithmetic is complete, non-trivial, suitable (in the sense in which Euclidian Geometry is not suitable to describe our universe), and so on. The point is also whether we have

[1] This paper relies on a conception of logic, and even of formal logic, which is different from the Vienna Circle's conception and leaves ample room for content-guidance—see Batens (2007a) for an elaboration.

the right view on what it means to be a theorem of Peano Arithmetic. We know from Gödel's First Incompleteness Theorem that Peano Arithmetic is incomplete if it is consistent. Whether it is non-trivial we do not know. Most mathematicians think it is, but Graham Priest has argued in Priest (2006) that it is not, at least not if it is extended with some obviously correct proof means that do not undermine its semi-recursive character. Moreover, we know from Gödel's Second Incompleteness Theorem that, if Peano Arithmetic is non-trivial, it is impossible to show so by means that can be represented within Peano Arithmetic.[2] Note that, if Peano Arithmetic is trivial, the statement that $0' + 0' = 0''$ is one of its theorems obtains a rather unexpected meaning. If Priest is right, the whole realm of mathematics has to be rethought, presumably starting from an inconsistent set-theory and working our way down to mathematical theories that we apply in everyday life.

We have known more revolutions in the history of mathematics. Many theories originated in a definite period. Their early history was often messy and full of nonsense. The early history of algebra is a ready example. Many years were required before one arrived at symbolic algebra and only then was sense made of isolated negative terms—see for example Heeffer (2007, 2010). Some theories, like Newton's infinitesimal calculus or Cantor's set theory or Frege's set theory were later found to be mistaken, even nonsensical if taken literally. There are the other limitative theorems, all discovered less than a century ago. All this drastically changed the conception of mathematical theories.

So mathematical knowledge is at least defeasible in the sense that one's best insights at some point in time may be later superseded by insights that derive from further study, or from improved conceptual insights, or presumably also from the development of empirical sciences. I shall argue in Sect. 1.8 that the same holds for logic.

The second thesis is a consequence of the preceding one: *No* foundation *is available in any domain.* This means that the only way to arrive at justified convictions is to improve our knowledge by relying on our present knowledge. Of course, we may and should learn from the history of knowledge, especially in methodological respects. But even historiography is the result of our present insights. Obviously, we should collect new data. Still, which data have to be collected and the importance and significance that will be attached to them will depend on present lights. This is what I call relative rationality (and relative justification).

Two other theses are directly connected to the previous one, but space does not permit to explain the connection. The first thesis: *All meaning is contextual.* The (intensional) meaning of words does not reside in some Platonic heaven, but in people's heads. As such, they depend on one's *view* on the domain. Studying *any* outdated theory from the history of the sciences is convincing in this respect. The second thesis: *Actual knowledge systems are neither holistic nor hierarchical.*

[2] Even this Theorem is defeasible. It is proved by means that are only reliable if Peano Arithmetic is non-trivial.

They consist of set of (larger or smaller, more or less vague) clusters of knowledge (about the world, methods, language, ...). These clusters are invoked to solve specific problems. They may be mutually inconsistent. Some of the clusters are related. (i) Some clusters are extended with consequences of others if the need presents itself—as when the knowledge related to handling everyday objects is extended with some physics or geometry to solve a specific problem. (ii) Some of the clusters are associated with others—a set of methodological do's and don'ts may be related to a scientific discipline.

According to this viewpoint, the central epistemological tasks are the following. First of all, one has to solve problems at all 'levels', including the conceptual organization of scientific theories and their unification. Such tasks are subordinate to the central problems that have to be solved. Next, one often has to analyse the context (problem-solving situation) in order to reach a solution. A third and central task is to reduce the role of 'pragmatic factors', factors that depend on properties of the knowing person or group rather than on the domain studied. Note that this reduction takes always place on the basis of available knowledge; it is not a matter of reaching more 'objectivity'. Another central task is furthering intellectual combat. Intellectual fight prevailed in all pivotal periods of the history of the sciences. It is the only means to discover the weak spots in our convictions and to strengthen them.

Subsequent sections are intended as an answer to the question whether the outlined epistemological stand makes sense from a logical point of view and whether it can be backed up by logical means.

1.3 Adaptive Logics: A Quick Rehearsal

An *adaptive logic* is a formal logic that 'adapts itself to the premises' and characterizes a defeasible reasoning form. The ultimate aim of the adaptive logic program is to characterize all defeasible reasoning forms by an adaptive logic *in standard format*.

In standard format, an adaptive logic is defined by a triple: (i) a *lower limit logic* **LLL**, roughly a deductive Tarski logic that is compact and for which there is a positive test,[3] (ii) a set of *abnormalities* Ω, which is characterized by a (possibly restricted) logical form,[4] and (iii) an adaptive *strategy*: *Reliability* or *Minimal Abnormality*.[5]

[3] There is a positive test for a logic **L** iff $\{\langle \Gamma, A\rangle \mid \Gamma \vdash_{\mathbf{L}} A;\ A$ is a formula; Γ is a recursive set of formulas$\}$ is a semi-recursive set. A more general description is that **LLL** is any logic that has *static proofs*, but space prevents me from clarifying this here.

[4] For example, a formula of the form $\exists(A \wedge \neg A)$ or a formula of the form $\exists A \wedge \exists\neg A$, in which $\exists A$ denotes the existential closure of A. A possible restriction is that A is a primitive (or atomic) formula, or that A is a disjunction of primitive formulas and negations of primitive formulas.

[5] These two strategies handle derivable disjunctions of abnormalities in different ways and have a different effect on the proofs and on the selection semantics.

The standard format provides the adaptive logic with dynamic proofs. Typical for the annotated proofs is that each line has a possibly empty *condition*, which is a finite set of abnormalities. The lower limit logic and the set of abnormalities determine the inferential rules (in terms of **LLL**-consequence), the set of abnormalities and the strategy determine the Marking definition. This definition proceeds in terms of the formulas that occur at the stage of the proof. Marked lines are considered OUT—their formula is not derived at the stage—while unmarked lines are considered IN. As the proof proceeds from one stage to the next, a marked line may become unmarked and vice versa. Note that the marks are a means to control the defeasible character of the logic: a formula is derivable or underivable from the premises in view of the insights in the premises that is offered by the stage of the proof. A definition settles which formulas are *finally derivable* from the premises. As (full-blown) defeasible logics have no positive test for final derivability, a finite proof in itself will often not enable one to decide that a formula is finally derivable from the premises; one needs a metatheoretic reasoning about possible extensions of the proof.

The standard format also provides adaptive logics with a semantics, viz. a selection semantics. The lower limit logic assigns a set S of models to the premise set; the set of abnormalities and strategy select a subset of S as the set of adaptive models of the premises.

Finally, the standard format also provides most of the metatheory: Soundness and Completeness of the proof theory with respect to the semantics, and a host of further metatheoretic properties.

Incidentally, the logic obtained by extending **LLL** with an axiom that connects abnormalities to triviality (or by weeding out models that verify an abnormality) is called the upper limit logic **ULL**. A remarkable property of adaptive logics is that $Cn_{\mathbf{AL}}(\Gamma) = Cn_{\mathbf{ULL}}(\Gamma)$ whenever Γ is normal (does not require any abnormalities to be true), whereas $Cn_{\mathbf{LLL}}(\Gamma) \subseteq Cn_{\mathbf{AL}}(\Gamma) \subset Cn_{\mathbf{ULL}}(\Gamma) = \mathcal{W}$ (with $\mathcal{W}$ the set of all formulas) whenever Γ is abnormal.

Adaptive logics are called corrective iff they handle premise sets that have no models in the 'standard logic'; otherwise they are called ampliative. Where **CL** (classical logic) is taken to be the standard, adaptive logics handling inconsistencies are corrective, while adaptive logics for inductive generalization are ampliative. For a more detailed description of the distinction, I refer for example to Batens (2004, 2007b, 2014).

1.4 Formal Problem-Solving Processes

Logics are not applied in isolation. This holds especially when adaptive logics are involved. Such logics are precise characterizations of methods and are at least combined with a deductive logic, usually their own lower limit logic. Obviously, several adaptive logics may be combined. This is the case, for example, when we look for an explanation of a fact in terms of present knowledge, but try to extend our knowledge with new inductive generalizations[6] in case our present knowledge does not provide an explanation.

Reasoning is a goal-directed and problem-oriented process. So we need to be able to express problems and the process should be sensible in view of the solution of the problem or problems. Moreover, we need to be able to ascertain whether a problem is well-formed in view of the available declarative knowledge, we need to be able to split up problems, and to derive problems from given problems in view of declarative knowledge. More often than not, the solution of a problem requires empirical import. This is obviously not provided by any logic, but the adaptive logic should trigger the empirical import; it should instruct one whether new empirical data may be relevant or not.

The required combination is realized by means of formal problem-solving processes, fpsps for short. These were first proposed in Batens (2003). In Batens (2007a) it was shown that fpsps leave ample room for, and actually install, content-guidance; it was also shown that building in observational or experimental means requires an 'oracle' that is different from the one introduced by Hintikka in Hintikka (1999). So let me describe the elements of a fpsp backbone.

An fpsp is a sequence of lines that results from a *procedure*. This is a set of instructions that consist of a rule with a permission or obligation attached to it; the permission or obligation should be defined in terms of the lines that already occur in the fpsp. A fpsp contains two kinds of lines.

Declarative lines are quadruples: a line number, a prospective expression of the form $[B_1, \ldots, B_n]\, A$, a justification, and an adaptive condition. The prospective expression states that A can be obtained from the premises iff the members of the prospective condition, $[B_1, \ldots, B_n]$ $(n \geq 0)$, can be obtained from them. The prospective dynamics is basically a way to push part of the proof heuristics into the proof (see Batens and Provijn 2001); the members of the prospective condition function as *targets* in view of which premises are introduced and analysed. Moreover, the prospective dynamics leads to criteria for final derivability in adaptive logics (see Batens 2005; Verdée 2013). The adaptive conditions are those of the involved adaptive logics; they are essential for the marking and hence for the control of the involved defeasible reasoning forms—see Sect. 1.3.

Problem lines are couples: a problem, which is phrased as a set of yes-no questions $\{?\{A_1, \neg A_1\}, \ldots, ?\{A_n, \neg A_n\}\}$ and a justification. Problems are handled in terms of Andrzej Wiśniewski's erotetic logic (see Wiśniewski 1995, 1996, 2003, 2004), which does not presuppose a specific deductive logic but is defined in a general way. The erotetic logic handles the evocation of questions by declarative premises and the implication of questions by other questions in view of declarative

[6] Many logicians appear to be mesmerized by the grue paradox. However, this concerns the choice of a language (or of a set of primitive predicates), which has obviously to be justified by non-logical means: entrenchment, etc.

premises. As fpsps (by present lights) start from a *main problem* and a set of premises, only question implication plays a role. Note that deriving a sub-problem from a problem is a logical matter, whereas deriving an 'auxiliary' problem requires declarative knowledge.

Apart from Wiśniewski's erotetic logic, which handles questions, there are specific rules to handle problems. Thus, once an answer to $?\{A_i, \neg A_i\}$ $(1 \leq i \leq n)$ is obtained—once A_i or $\neg A_i$ is derived on an empty prospective condition—the problem $\{?\{A_1, \neg A_1\}, \ldots, ?\{A_n, \neg A_n\}\} - \{?\{A_i, \neg A_i\}\}$ is derivable from the problem $\{?\{A_1, \neg A_1\}, \ldots, ?\{A_n, \neg A_n\}\}$. As soon as the former is derived, the latter becomes redundant.

The main problem is essential for the goal-directed character of fpsps. A prospective proof normally starts with a (redundant) *goal* statement of the form $[A]\,A$, which is meant to introduce A as a target. In an fpsp, goal statements are derived from non-redundant problems. This means that $[A]\,A$ can only be introduced if $?\{A, \neg A\}$ is a member of a non-redundant problem. Once targets are present, the prospective dynamics warrants the goal-directed character of the fpsp; a new line can only be added if its prospective expression potentially brings us closer to deriving a target. Whether it actually brings us closer to deriving a target depends on the premises and cannot in general be settled beforehand, given that we are not logically omniscient. Finally, new problems are only introduced if derived declarative statements warrant that their solution is useful for solving given problems in view of the declarative premises.

Note that fpsps *guides research*. The introduction of a goal statement $[A]\,A$ will, after a number of steps, lead to the presence of expressions of the form $[B_1, \ldots, B_n]\,A$. This suggests observations, experiments, conceptual analysis, or bringing in other information, and all of these will lead to new premises. New premises will be required if $?\{A, \neg A\}$ is a member of a problem and the prospective conditions of A and $\neg A$ cannot be derived from the premises. Even if the question can be answered by deductive means, a new premise may be more easily obtainable by observation than by deduction.

I only presented an outline of the backbone of fpsps. Still, having referred the reader to other papers, I should stress that lots of work still has to be carried out. Thus the incorporation of some adaptive logics in fpsps requires that the relevant heuristics is elaborated. All this, however, is pretty standard or at least does not require much ingenuity.

1.5 Contextual Meaning

The first point I shall make is that adaptive logics introduce contextual meanings. To see this, consider a simple propositional example of a proof for the logic $\mathbf{CLuN}^r$. Its lower limit logic (generic name: **LLL**) is **CLuN**, which is full positive propositional logic together with excluded middle. The set of propositional abnormalities comprises the formulas of the form $A \wedge \neg A$—the predicative abnormalities are the existential closure of those formulas, in which A is then possibly open. The strategy is Reliability—see below.

The (generic) rules of inference are the same for all (non-combined) adaptive logics. Let

$$A \qquad \Delta$$

abbreviate that A occurs in the proof on the condition Δ, which is a set of abnormalities (so a subset of Ω). There are three generic rules. In RC, $\check{\bigvee}\Theta$ denotes the classical disjunction of the members of $\Theta \subset \Omega$ and the symbol $\check{\vee}$ is the classical disjunction.[7]

$$\text{Prem} \qquad \text{If } A \in \Gamma\text{:} \qquad \begin{array}{ll} \dots & \dots \\ \hline A & \emptyset \end{array}$$

$$\text{RU} \qquad \text{If } A_1, \dots, A_n \vdash_{\mathbf{LLL}} B\text{:} \qquad \begin{array}{ll} A_1 & \Delta_1 \\ \dots & \dots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \end{array}$$

$$\text{RC} \qquad \text{If } A_1, \dots, A_n \vdash_{\mathbf{LLL}} B \check{\vee} \check{\bigvee}\Theta\text{:} \qquad \begin{array}{ll} A_1 & \Delta_1 \\ \dots & \dots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \cup \Theta \end{array}$$

Let the premise set be $\{(\neg p \wedge \neg q) \wedge t, p \vee r, q \vee s, p \vee q, t \supset p\}$. Here is the proof up to stage 8[8]:

1	$(\neg p \wedge \neg q) \wedge t$	PREM	$\emptyset$	
2	$p \vee r$	PREM	$\emptyset$	
3	$q \vee s$	PREM	$\emptyset$	
4	$p \vee q$	PREM	$\emptyset$	
5	$t \supset p$	PREM	$\emptyset$	
6	r	1, 2; RC	$\{p \wedge \neg p\}$	$\checkmark$
7	s	1, 3; RC	$\{q \wedge \neg q\}$	$\checkmark$
8	$(p \wedge \neg p) \vee (q \wedge \neg q)$	1, 4; RU	$\emptyset$	

Lines 6 and 7 are marked at this stage of the proof. Where Reliability is the strategy, a line is marked at a stage iff its condition contains a disjunct of a *minimal*

[7] I skip some related complications. They are not relevant to the point I am trying to make here.

[8] A stage is a sequence of lines and a proof is a chain of stages.

disjunction of abnormalities. At stage 8 of the proof, the only minimal disjunction of abnormalities is the formula of line 8. Let us now consider stage 9 of the proof—I rewrite the sequence of lines from line 6 on.

6	r	1, 2; RC	$\{p \wedge \neg p\}$	$\surd$
7	s	1, 3; RC	$\{q \wedge \neg q\}$	
8	$(p \wedge \neg p) \vee (q \wedge \neg q)$	1, 4; RU	$\emptyset$	
9	$p \wedge \neg p$	1, 5; RU	$\emptyset$	

Line 6 is still marked, but line 7 is unmarked. Indeed, the only minimal disjunction of abnormalities at stage 9 is the formula of line 9, viz. $p \wedge \neg p$.[9]

In all subsequent stages of the proof—in more traditional terms: in all extensions of this proof—line 6 is marked and that line 7 unmarked. In other words, the proof at stage 9 is stable with respect to lines 6 and 7 (and actually with respect to lines 1–7). So s is finally derivable from Γ, whereas r is not.[10]

The above proof nicely illustrates the contextual meaning of negation. The negation in $\neg p$ is clearly the paraconsistent **CLuN**-negation because both p and $\neg p$ are **CLuN**^{r}-derivable from the premises. The negation in $\neg q$ has the force of a **CL**-negation. Precisely this is why s is a consequence of the premises: s follows from $\neg q$ and $q \vee s$ because the premises do not require that $q \wedge \neg q$ is a disjunct of a true and minimal disjunction of abnormalities.[11]

To understand what is going on, let us have a look at the preferred application context of inconsistency-adaptive logics. Consider a theory T, that is intended as consistent and has **CL** as its underlying logic, but turns out to be inconsistent. Often, one will try to find a consistent replacement T', and one will try to obtain it by reasoning from T. One typically will want a T' that is as rich as T, except that T' should be consistent. While **CL** is obviously useless for this purpose (it identifies T with the trivial theory), (static) paraconsistent logics are too weak; many 'good' consequences of T will not be derivable by the paraconsistent logic because it is much weaker than **CL**.

Inconsistency-adaptive logics are obviously not intended to remove the inconsistencies. This should be done on the basis of non-logical arguments: new empirical data or new results of conceptual analysis. Incidentally, it is not difficult to devise adaptive logics that remove inconsistencies. However, some such logics lead to arbitrary results (removing, for example, p in favour of $\neg p$) and some leave one with too poor a theory (when it removes both 'halves' of every inconsistency).

In preparation of removing the inconsistencies by non-logical means, we need to obtain a maximally consistent interpretation of T, from which the inconsistencies

[9] The Minimal Abnormality strategy leads, for some premise sets, to a richer consequence set than the Reliability strategy. For this proof, however, both strategies lead to the same marks.

[10] The definition of final derivability is slightly more sophisticated, but this is what it comes to for the present propositional example.

[11] If the negation is paraconsistent and A as well as $\neg A$ are true, then $A \vee B$ and $\neg A$ are true even if B is false.

may then be removed. Inconsistency-adaptive logics should provide us with such an interpretation. The example proof illustrates the way in which they do so. Inconsistency-adaptive logics, like $\mathbf{CLuN}^r$, interpret premise sets as consistently as possible.[12] In other words, inconsistencies are considered as false, except when the premises prevent this; for Reliability this comes to: except when the inconsistency is a disjunct of a minimal disjunction of abnormalities that is derivable from the premises by the lower limit logic. Precisely because inconsistency-adaptive logics interpret premise sets as consistently as possible, they assign a contextual meaning to negation. It is worth noting that the meaning that is assigned to a negation depends on the *content* of the premise set.

Incidentally, why do we need a maximal consistent interpretation of the theory? The theory T was *intended* as closed under **CL**: every **CL**-consequence of theorems of T is itself a theorem of T. This causes T to be trivial. We cannot look for a consistent replacement of T itself, but we can look for a consistent replacement of the non-trivial theory that it closest to the original intention, and this is a maximal consistent interpretation of T. In it the inconsistencies are localized and all 'parts' in which no inconsistency is involved are closed under **CL**.[13]

There is a variety of inconsistency-adaptive logics. Each of them offers a variety of maximally consistent interpretations of T. In doing so, each of them assigns, for every formula of the form $\neg A$, one of (at least) two meanings to the $\neg$ in view of the premise set. These are the meanings of negation as fixed by respectively the lower limit logic and the upper limit logic. By varying the lower limit logic and the set of abnormalities—the latter may have effects on the upper limit logic—one varies the couples of negations that is chosen from. By varying the strategy and by certain variations of the set of abnormalities,[14] one may vary the choices made between a same couple of negations. A similar (but different) variation is obtained by combined adaptive logics.[15] Incidentally, some combined inconsistency-adaptive logics assign contextual meanings to negations by choosing from more than two negations.

Do not complain about the large variety of possibilities. As was mentioned in Sect. 1.1, adaptive logics form precise formulations of methods, in the present case methods for handling inconsistency. The choice between such methods requires a philosophical justification, which should be contextual in that it should depend on the properties of the specific situation. More variation is spelled out in Batens (2014, Chap. 7).

[12] This expression is ambiguous and is disambiguated by the strategy.

[13] A maximal consistent interpretation can only be defined by a defeasible inference relation and the adaptive logic program aims at characterizing all such relations.

[14] Given a lower limit logic, there are often several sets of abnormalities that lead to the same upper limit logic.

[15] Several combined adaptive logics have been described. The simplest combination, which obviously works only under certain conditions, is where a consequence set of Γ is defined as the union of the consequence sets different simple adaptive logics assign to Γ.

There is, however, a very different multiplicity of adaptive logics for handling inconsistency. An inconsistency, viz. that some A is true together with $\neg A$, may be seen as a negation *glut*. That A as well as $\neg A$ are false may be seen as a negation *gap*. In a similar way, all logical terms (including the quantifiers and identity) may be said to display gluts or gaps (or both). One way to define the gluts and gaps is with respect to the **CL**-truth conditions. For example, there is an implication glut if $A \supset B$ is true while A is true and B is false, and there is an implication gap if $A \supset B$ is false while A is false or B is true. The interesting point is that many inconsistent theories have models in which no negation gluts but other gluts or gaps occur. Whenever this is the case, the adaptive approach (which 'minimizes' abnormalities) leads to 'interpretations' of the inconsistent theory that are as normal as possible. This was explained already in Batens (2000b) and the matter is studied at some length in Batens (2014, Chap. 8). There I also consider the ambiguity of non-logical terms[16] as well as its combination with kinds of gluts and gaps for logical terms. The combination of ambiguity with all kinds of gluts and gaps leads to zero logic, by which nothing is derivable from any premise set, not even the premises themselves. All the (Tarski) logics and combinations of them lead to adaptive logics that define a maximally normal interpretation of the premises. The adaptive logic that has zero logic as its lower limit may not be very interesting in itself, but it is an excellent instrument for surveying which choices (of gluts, gaps and ambiguities) are sufficient to obtain a minimally abnormal interpretation of the premises.

So each of these adaptive logics offers a minimally abnormal interpretation of premise sets and introduces contextual meanings for all the logical terms for which they tolerate gluts or gaps. The same holds for ampliative adaptive logics—the matter is just a trifle more complicated.

1.6 Meaning of Logical Symbols

According to the official doctrine, a logic determines the meaning of its logical terms. By "logic" is meant a deductive logic here. What comes of this if adaptive logics are applied? Three positions seem sensible: the two-logic view, the dialetheist view, and the direct view.

According to the two-logic view, the meaning of logical terms is defined by the lower limit logic and the upper limit logic, which may be seen as deductive logics. The adaptive logic picks the right choice for each occurrence of the logical term. It picks the meaning from the upper limit logic whenever this is possible—a matter disambiguated and determined by the strategy.

If the adaptive logic **AL** is corrective, it offers a minimally abnormal interpretation of a theory or premise set. If the theory is normal, **AL** offers a normal

[16] Present ambiguity-adaptive logics do not assign any specific meanings to occurrences of non-logical terms. They merely minimize the number of occurrences of the same term that require a different meaning.

interpretation—say the **CL**-interpretation. If the theory is abnormal, **AL** offers an interpretation according to which some logical terms have a meaning that is weaker than the **CL**-meaning. In the preferred application context, the result will eventually be replaced, on non-logical grounds, by a normal theory. So the contextual meanings are provisional; they apply in a transitory period in which problems have still to be solved in order to reach the 'finished' and normal theory.

It is worth noting that even the transitory stage[17] can be made fully transparent from a logical point of view. Suppose for example that we are dealing with a simple adaptive logic which has **CL** as its upper limit and some weaker logic **LLL** as its lower limit. Let the standard logical symbols be those of **LLL**. It is obviously possible to enrich the language of **LLL** with a set of logical symbols that have the same meanings as the symbols of **CL**—the Ghent standard is to use 'checked' symbols for these: $\check{\neg}$, $\check{\wedge}$, $\ldots$, $\check{\exists}$, and $\check{=}$. From the premises of the example proof in Sect. 1.5, $\check{\neg}q$ is finally derivable whereas $\check{\neg}p$ is not.[18] Note that only the standard symbols occur in the premises and that the adaptive logic determines which negations may be replaced by the **CL**-negation.

As announced in Sect. 1.5, the matter is slightly more complicated for ampliative adaptive logics (such as adaptive logics for inductive generalization, for abduction, and so on). These adaptive logics offer a richer consequence set than **CL** (which I here consider as the standard of deduction for merely pragmatic reasons). To take a concrete example, consider inductive generalization and let the set of abnormalities be $\exists A \wedge \exists\neg A$ in which $\exists A$ is the existential closure of A.[19] The upper limit logic is the so-called uniform classical logic (it has $\exists A \supset \forall A$ as a theorem and, in its models, the assignment value of every unary predicate is either the empty set or the whole domain; and similarly for other predicates). A typical application of the conditional rule, RC, is that from Pa on the empty condition follows $\forall x\, Px$ on the condition $\exists x\, Px \wedge \exists x \neg Px$. Let us compare the meaning of the universal quantifier in the upper limit logic with its meaning in the lower limit logic **CL**. The former is only stronger than the latter in that less information is required for a universally quantified formula to hold true (a single instance and even the corresponding existentially quantified formula is sufficient). The upper limit meaning is typically invoked by the conditional rule, RC, which introduces a new condition in view of its defeasible character. Once the universally quantified formula is obtained, other formulas may be derived from it by the unconditional rule, RU, (carrying over the condition). However, the upper limit meaning of the universal quantifier is not different from its lower limit meaning in this respect, viz. with respect to the formulas that are derivable *from* a universally quantified formula. In other words,

[17] See the previous paragraph: the stage at which the adaptive theory is not yet replaced by a novel normal theory.

[18] Except for negation, all logical symbols of **CLuN** have the same meaning as the classical logical symbols.

[19] I simplify here. The actual adaptive logic I have in mind here (see Batens and Haesaert 2001; Batens 2011) imposes certain restrictions on A.

once the adaptively derivable generalizations have been added to the theory, one may 'reaxiomatize' the theory and the result may be seen as a **CL**-theory.[20]

So far for the two-logic view. The dialetheist view is radically different from it, but it is more restricted because it only concerns negation. For a dialetheist like Graham Priest, there is a 'true logic', viz. the paraconsistent **LP**.[21] An adaptive logic that has **LP** as its lower limit logic is ampliative for the dialetheist (because **LP** is the standard of deduction). Applications of RC are justified by the so-called consistency presumption: that most inconsistencies are false and hence that they may be taken to be false unless it is found that the premises require the opposite. So the true logic alone determines the *meaning* of the logical terms (in a sense this is a hyper-classical position). The consistency presumption offers reasons to accept additional consequences. These, however, do not follow by logic but by logic together with the consistency presumption. Put differently, the consequence set is changed, but the meaning of negation is not.

I do not know what a dialetheist would make of the meaning of logical terms in other, for example ampliative, adaptive logics. I guess that the answer would be that deductive logic determines the meaning of the logical symbols whereas methodological steps do not affect meanings, even if they allow one to derive certain conclusions that do not follow by logic.

A more interesting question to be answered by a dialetheist concerns theories that were intended to be handled by the 'true logic', but turned out trivial. Especially if a detachable implication is around, a mathematical theory, for example, may turn out to be trivial. So suppose that the dialetheist's set theory is found to be trivial. In order to replace it by a non-trivial improvement, the dialetheist will reason from the trivial set theory. The only way to do so, as far as I see, is by considering an adaptive logic that has the 'true logic' as its upper limit and that has as lower limit a logic according to which the set theory is non-trivial. Only this approach will lead to a maximal non-trivial interpretation of the set theory. If such an approach is followed, the meaning of the logical symbols cannot be defined by the 'true logic' because this results in triviality.

If the true logic is **LP** (without a detachable implication) as was Graham Priest's view before the first edition of Priest (2006), the problem can be neglected because $B_1 \wedge \ldots \wedge B_m$ ($m \geq 1$) is a **LP**-consequence of $\{A_1, \ldots, A_n\}$ iff every B_i ($1 \leq i \leq m$) is a **LP**-consequence of a single A_j ($1 \leq j \leq n$). So a theory is only trivial iff, for every formula A, there is an axiom B of the theory from which A is **LP**-derivable. The presence of a relevant implication, however, changes the matter drastically.

[20] I here consider the case in which no new premises are added to the theory. Indeed, if new data are gathered, these may falsify the adaptively derived generalizations and hence trivialize the so extended 'reaxiomatized' theory.

[21] The implication is defined in terms of the paraconsistent negation, viz. by $\neg A \vee B$, whence it is not detachable. In later versions, Priest added a modal implication to **LP**, which he later replaced by a relevant implication.

Finally, let us turn to the direct view. According to this, the adaptive logic itself determines the meaning of logical symbols. So meaning is explicitly contextual. Let me explain in which sense this view is different from the two-logic view.

First a technical matter. Some adaptive logics, for example the logic of inductive generalization **LI**, have the trivial logic **Tr** as their upper limit.[22] To say that a logical term receives the meaning assigned to it by **Tr** is a bit of a nonsense. So it seems more sensible that **LI** allows one to derive a universally quantified statement $\forall A$ from one or more instances of it, unless and until a statement falsifying $\forall A$ (for example, $\exists\neg A$) is derived from the premises.[23] This only affects the universal quantifier and, as long as the line on which the universally quantified formula is derived is unmarked (viz. as long as no formula falsifying $\forall A$ is derived), the change to the meaning of the universal quantifier reduces to the fact that the formula can be introduced on the basis of an instance.

The relevant philosophical question, which is much more general, is whether deductive logic can be separated from defeasible logic. Technically this separation is obviously possible. The question, however, is meant in the epistemological sense. Is it possible, within a given problem solving context in which reasoning occurs, to construct theories about the meaning of logical terms in such a way that these theories are independent from the meanings of the non-logical terms? The standard format of adaptive logics[24] was devised in such a way that this separation is maintained. It was not shown, however, that all defeasible reasoning will eventually be integrated in the present standard format of adaptive logics or in a format that allows for the separation. In other words, it is possible that our theories about the world fix the meaning of logical terms in such a way that no separate theory about the logical terms can be split off. This might especially be the case in ‘provisional’ stages of those theories, in which lots of *theoretical* problems are still to be solved by means of defeasible reasoning forms.[25] If such a situation obtains, one may still devise deductive logics but their application will be restricted if not empty and their use spurious.

Needless to say, the meaning of the logical symbols is determined by the derivability relation, not by what is actually derived (at a stage or finally) in a proof from a premise set. This is not any different from deductive logics. For example, that p was not derived from $p \wedge q$ in a given proof does nor affect the meaning of (this) conjunction.

[22] Do not confuse **Tr** with the modal logic **Triv**. **Tr** is characterized by $\Gamma \vdash_{\mathbf{Tr}} A$ for all Γ and A. It either has *no* models or only one, viz. the trivial model.

[23] This is not fully accurate. A disjunction of statements of the form $\exists\neg A$ will have the same effect in view of the marking definition. This is one of the reasons why the **LI**-consequence set of a consistent premise set is always consistent.

[24] A first attempt to formulate it was made in Batens (2001). The present version is in Batens (2007b) and especially in Batens (2014, Chap. 4).

[25] Some defeasible reasoning forms only concern application problems. Abduction is a ready example.

1.7 Complex Theories

Until now, I considered adaptive logics as methods, so as mainly relevant for the development of theories and for their application. But might such logics not also be employed as underlying logics of theories? I shall offer a brief argument to show that this is meaningful and, especially in view of Gödel's incompleteness theorems, offers interesting perspectives.

A theory is often seen as a couple comprising a decidable set of axioms and a *logic*, $T = \langle \Gamma, \mathbf{L} \rangle$. The set of theorems of the theory, with which the theory is sometimes identified, is taken to be $Cn_{\mathbf{L}}(\Gamma)$.

Up to the nineteenth century, the logic of most theories was implicit and only rarely were the axioms listed. Yet theories were considered as well-defined and apparently also as effectively decidable (although the concept itself was not explicitly around). In the early twentieth century, it was discovered that **CL** is only semi-recursive. So (predicative) theories, even if well-defined, are only semi-recursive consequence sets of a recursive set of axioms. At the same time, Gödel's incompleteness theorems revealed grave restrictions—see Sect. 1.2. Theories with other underlying logics were proposed. These were either also semi-recursive, for example when the logic was intuitionistic or relevant, or else they were much more complex, for example when the underlying logic was second order—the reader will remember that second-order logic requires infinitary rules.

By an adaptive theory I obviously mean a theory $T = \langle \Gamma, \mathbf{L} \rangle$ in which $\mathbf{L}$ is an adaptive logic. For such theories $Cn_{\mathbf{L}}(\Gamma)$ is not in general semi-recursive—it may be up to Π^1_1-complex—see Verdée (2009).

The reasons for introducing adaptive theories is that the world may be so complex that it cannot be captured by semi-recursive theories. Or rather, we know that the world is so complex in view of Gödel's first incompleteness theorem. If it is captured by an adaptive theory, this theory does not have certain nice properties of traditional first-order theories, but no nice theory can capture the domain anyway. Moreover, adaptive theories have certain relatively nice properties. For one thing, they define theories (in the sense of sets of theorems) just as second-order logics. Their proofs at a stage are simple in that they proceed in terms of finitary rules; in this sense they are much simpler than second-order theories. These proofs explicate actual reasoning and introduce a kind of control, in terms of the conditions and the Marking definition, that is absent in actual reasoning. Finite adaptive proofs-at-a-stage are not more complex than, for example, **CL**-proofs. Even infinite adaptive proofs are relatively simple, given that they consist of a denumerable set of lines and that all applied rules are finitary (every conclusion is drawn from finitely many formulas preceding it). Heuristic procedures for adaptive proofs are available. Moreover, there are procedural criteria for final derivability (see Batens 2005 and especially Verdée 2013), whence certainty can be gained about at least a number of theorems of the theories. Note that, where a criterion applies, it establishes final derivability in a finite proof.

These criteria are worth a further comment. Establishing final derivability requires in principle that one offers an argument about all (finite or infinite) extensions of a finite proof. There are clearly more than countably many such extensions. Establishing final derivability in terms of a procedure reduces the complexity of the required reasoning. It is sufficient to establish that a finite set of formulas cannot be derived from the (decidable) premise set. The premise set is always countable and non-derivability is established in terms of the 'positive part' relation, which is decidable.

Even when no criterion applies, proofs at a stage give us the best estimate of the theory that can be obtained in view of present insights, which are the insights provided by the present proofs. This is a basis for drawing a defeasible conclusion.

I have some results that relate to Gödel's second incompleteness theorem: how to deal with a possibly inconsistent axiom system for arithmetic? As these results are in print, I do not mention them here. Moreover, the reader will be more interested in the question whether adaptive theories may be complete with respect to such domains as *true arithmetic* (the formulas verified by the standard model). Right now, Peter Verdée has ideas on the matter that look extremely promising to me—I advise the reader to look out for forthcoming results. For now, let me restrict myself to a promise. If true arithmetic is consistent, it is likely that an adaptive theory is complete with respect to it and that every formula which one can show to be verified by the standard model, for example the Gödel sentence, can be shown to be a theorem of the adaptive theory. If true arithmetic is inconsistent, the standard model (and most of model theory) is nonsense. So the classicist looses everything she has. In that case, it is extremely likely that there is an inconsistency-adaptive theory which is non-trivial and actually behaves for all natural numbers (the numbers of the standard block) as true arithmetic was intended to behave.

1.8 A Form of Logical Pluralism

In Sect. 1.2, I claimed that all meaning is contextual. In this section, I offer some further arguments for that claim. These arguments go along with my epistemological stand, but may be considered independently of it. Let me admit at once that this section is somewhat touchy. Some papers opposing logical pluralism dragged me into the scene, but I felt deeply misunderstood, accused of things I never stated, associated with positions I consider utterly mistaken. If one is misunderstood, one may attack the 'opponent'. One may also feel guilty. After all, if X writes out Y's position, X will construct Y's view on the basis of X's view. Who is to blame for misunderstanding? So let me try to be constructive and make another attempt to state my position. Part of the statement is determined by certain misunderstandings, but I shall not bother the reader with them.

The aim of logic is to explicate *reasoning*. What is 'out there' is actual reasoning and it has to be *explicated*. It is not a matter of fact. It is not a platonic heaven. It is not a domain that has to be *described*. So no descriptive theory of actual reasoning will do. The *explicandum* contains mistakes and there is a normative dimension.

As our culture likes distinctions, let us make them. Some of the reasoning is deductive, some is defeasible. Deductive reasoning can be separated into formal reasoning and informal reasoning.[26]

Formal deductive reasoning concerns logical terms. Its correctness is judged in view of the meanings of logical terms. This is the reason why deductive (formal) logics are supposed to fix the meanings of logical terms.

Among the logical terms that extremely frequently occur in actual reasoning are causal relations, time and tense, deontic operators, and sundry kinds of other modalities. All these are neglected by **CL** and actually by most other Tarski logics. More importantly, there is obviously a manifold of each of these. Just think, for example, of logical modalities, nomological (or 'physical') modalities, practical modalities (it is physically possible but practically impossible to bring the moon into a different orbit tomorrow), and so on. In order to make their case, monologists should articulate a single logical system that deals with all logical terms, a matter far from realized today. Moreover, they should be able to use their logical system to axiomatize all required mathematical theories as well as all empirical theories.

Some monologists will argue that there is no objection against axiomatizing a mathematical theory by means of another logic **L**, as a merely technical realization as it were. The idea is that the theorems of the so obtained theory are then combined with other, for example empirical, statements in order to forge empirical theories. Note that this will only do if **L** is at least as strong as the true logic. In other words, the true logic should be conservative with respect to every mathematical theory. If it is not, non-theorems of the mathematical theory (and of its language) will be derivable by the true logic and hence will ruin the applicability of the original mathematical theory.

For the sequel of this section, let us restrict attention to the traditional logical terms, say those of the predicative language schema. I shall argue that even with this restriction there are reasons for logical pluralism.

Why should the traditional logical terms be unique? Why should only one negation, one implication, one universal quantifier, ... occur in reasoning? Everyday practice clearly points to the opposite. Some negations are paraconsistent while others clearly are not. Some implications are contrapositive or transitive, while others clearly are not—see also below, where I come to the distinction between formalization and logical inference, but daily practice clearly favours a multiplicity of logical terms. So the burden of proof is on those that argue for uniqueness and, claims apart, they did not produce any sound arguments.

Once we grant that there is a multiplicity of unambiguous logical terms, why should all unambiguous logical terms occur in all contexts? Whether "context" is understood here as linguistic context or as problem-solving situation, the facts

[26] Defeasible reasoning can also be separated into formal reasoning and informal reasoning. Adaptive logics, for example, characterize formal defeasible reasoning forms. However, I shall not need this distinction in the sequel.

plea in favour of a negative answer to the question. So the burden of the proof is again with those that favour a positive answer. Again, prejudice apart, they failed to produce any sound arguments.

Once we grant that not all unambiguous logical terms occur in all contexts, why should a *unique* logic **L** be a *suitable* explication of the logical terms that occur in all contexts? Let **L** be a suitable explication for the logical terms that occur in a context in which we reason about beers. Why should **L** also be a *suitable* explication of the logical terms that occur in the context in which we reason about **L**? The burden of the proof ...

Let me interrupt this for a moment. Many classical logicians, relevance logicians, dialetheists, ... just *take it for granted* that there is a 'true logic' **L** and that **L** should be the logic of the metatheory of **L**. I tried to stepwise spell out what they take for granted in order to arrive at this conclusion. I stepwise asked for arguments. All I got, looking at the literature, is the well-known "it is obvious that" (sometimes phrased as "it is reasonable to take it that"). But let me go on to the final step on deductive logic.

As soon as we grant that a logic may differ from the logic of its metatheory, the following seems justified. There is no 'true logic' of which *parts* are used in a 'context'. In other words, there is no 'true logic' that comprises the logical terms used in all possible contexts ($\neg_1, \neg_2, \ldots, \supset_1, \supset_2, \ldots, \wedge_1, \wedge_2, \ldots, \ldots$) and from which the right logical terms are chosen according to the context. While the burden of the proof is still on those who claim there is such a logic, let me add some arguments to show that the burden is heavy. (i) Joining logics may ruin the meaning of the involved logical terms and, worse, may have tonk-like effects (cause the joined logic to be identical to the aforementioned trivial logic **Tr**). (ii) The 'true logic' *cannot* itself determine the contextual choice of terms. (iii) Two unambiguous logical terms may be equally suitable explications for the same bit of reasoning. Note that (ii) and (iii) are the reason why monologists candidly separate the formalization of natural language arguments from logical inference and, equally candidly, leave the formalization part unexplicated. Proceeding thus, they put themselves into a quite comfortable position. Formalizing your statements by means only known to God, they then decide what follows from your formalized statements according to their 'true logic'. In this way, they move the burden of the proof to you. If you drew a conclusion they reject, you have to find, in their logical system, a formalization of your statements from which follows your conclusion. Similarly if you reject a conclusion they draw.

From here, we move on quickly. Let us first move to informal deductive reasoning. This concerns reasoning that is correct in view of the meanings of non-logical (or referring) terms. To these applies all that was said about logical terms, but there is more. Referring terms are vague, ambiguous, etc. If in doubt, open a dictionary. Referring terms are also theory-laden. It does not follow that they are incommensurable (theory-ladenness need not prevent communication).

Finally turning to defeasible reasoning, note that all that was said before (about logical and referring terms) applies here as well. Even more than for deductive reasoning, defeasible reasoning requires an *explication*. The idea that anything *a priori* would be involved is as crazy as outdated—see the quotation from Dudley Shapere.

The first line of argument started from human reasoning. Let me briefly follow a second line of argument. Many people take it that knowledge about logical terms should not be obtained by starting from actual reasoning. They hold that there are other ways for obtaining such knowledge and that these refer to a more objective basis for logical terms.

For a start, let it be mentioned that logical terms are not hardwired in our brains. In a sense, classical logic is hardwired in digital computers, but that does not prevent one from writing programs for implementing other logical terms. While our knowledge about the functioning of human brains is far from perfect, we have reasons to believe that human brains are very different from digital computers. Our present knowledge about brains does not reveal any hardwired logical terms and rather supports the claim that logical terms are neurologically complex entities. Moreover, we are able to handle a variety of logics (classical, intuitionistic, relevant, . . .) and nothing suggests that one of these is the basis from which the other logical terms can be implemented in our brains.

It is sometimes claimed that truth-preservation provides access to the true meanings of logical terms. This seems putting the cart before the horse. Indeed, in order to find out which inferences are truth-preserving, one needs to know the truth conditions of the logical terms. Thus, if $\supset$ is the implication of **CL** or of intuitionistic logic, then the inference from A to $B \supset A$ is truth-preserving; if the implication is relevant, the inference is not truth-preserving. A (coherent) semantics fixes those truth conditions and hence fixes which inferences are sound. A (coherent) formal system also fixes those truth conditions because it determines an inferential semantics. This is usually described in terms of a set of two-valued valuation functions v. The transition from the formal system to the semantics is obtained by translating every correct inferential statement $A_1, \ldots, A_n \vdash B$ to: "for all v, $v(A_1) = 0$ or . . . or $v(A_n) = 0$ or $v(B) = 1$".[27] The conclusion of all this is that a logic fixes (its own) truth-preservation and hence that truth-preservation cannot be used as a criterion for finding 'the true logic'.[28]

Others claim that conceptual analysis provides access to the true meanings of logical terms. The criticism to this view is all in line with the one in the previous paragraph. The central question is which concepts are analysed. Intuitionistic disjunction is clearly different from classical or relevant disjunction, but both seem equally coherent. Similarly for implication: the classical, intuitionist, and relevant concepts are different (and relevant implication has many variants). This is not only

[27] Such an inferential semantics is often 'ugly' and sometimes not even recursive. Nicer results are sometimes obtained by translating the inferential statements to a worlds semantics. In Suszko (1977), Suszko has shown that every 'logic' has a two-valued semantics; in Routley and Meyer (1976), Routley and Meyer have shown the same for a two-valued worlds semantics.

[28] This holds even for logics phrased in natural language, like Aristotle's syllogistics. The syllogistics determines to which occurrences of "all", "some", etc. it is legitimately applied.

typical for logical terms but also for mathematical ones. Cantor's and Frege's set theories were shown trivial and hence incoherent. Possibly consistent replacements that have **CL** as their underlying logic are **ZF**, **NF**, and several others. Each of these clearly concerns (or rather introduces) a different concept of set membership, and hence a different set concept. A nice example in this realm is Zach Weber's set theory from Weber (2010). Here the concept of set inclusion is intensional. Where x and y are sets, $x = y$ iff the condition for being a member of x is *equivalent* to the condition for being a member of y, where equivalence is relevant. It follows that there are infinitely many sets that, for example, have no member, but are not identical to each other—and similarly for other sets. This clearly is not in line with the tradition that locates set theory in the domain of extensionality, and it restricts the possible application contexts. Nevertheless, the underlying concept seems quite coherent and well-analysed.

Returning to the main point, it is possible that, in the end, only one concept of each kind (one negation, one implication, etc.) would turn out to be coherent—together they would form 'the true logic'. This does not seem very likely, however. That there are different notions of coherence makes it even less likely. The stronger form of coherence one adopts, the less likely is the warrant for uniqueness. If coherent is taken to mean non-trivial, then uniqueness becomes very unlikely. Indeed, it would mean that at most one logic known today would be non-trivial.[29] Moreover, the non-trivial logical concepts would have to be such that they tolerate no weakening—every weakening of the 'unique' logical terms is bound to warrant non-triviality—see next paragraph. If coherent is meant as stronger than non-trivial, I am not sure that incoherence is very fatal. Nearly all creative episodes, in empirical and logico-mathematical sciences alike, were incoherent according to some notions of coherence.

There is a limit to weakening logical terms. Consider a logical term that has no meaning at all. So it may be deleted from every string of symbols in which it occurs, without the meaning of the string being changed. Such a symbol clearly serves no purpose and does not contribute anything to logic. Phrased differently, empty logic, viz. the logic **L** according to which $\Gamma \nvdash_{\mathbf{L}} A$ for all Γ and A, is coherent but does not explicate or enable any reasoning. Note, however, that another possibility reveals itself at this point. We have seen in Sect. 1.5 that there is an adaptive logic based on zero logic. Applied to a consistent set of premises, this adaptive logic delivers exactly the same consequence set as classical logic. Applied to an inconsistent set of premises, the adaptive logic will interpret the premise set as normally as possible—read this as: as much as possible in agreement with **CL** (or with whatever the upper limit logic is chosen to be). So the lower limit logic assigns no meaning to logical terms, but the adaptive logic interprets them as much as possible in agreement with **CL** (or with whatever the upper limit logic is chosen to be). This means that the meaning of all logical terms becomes context-dependent, viz. depends on the contents of the premises.

[29] A logic **L** is trivial iff $\Gamma \vdash_{\mathbf{L}} A$ for all Γ and A.

Incidentally, I consider brain science, truth-preservation and conceptual analysis as important instruments (of which the first is beyond logicians' competence). I only argued that they are not sound means to arrive at 'the true logic'.

After having attacked means invoked by monologists, I now turn to means that I would consider conclusive. Actually, I see only one. We (humans) are striving for obtaining a body of useful knowledge. I write "useful" rather than "complete", because completeness seems out of reach anyway. By writing "useful", I also mean to exclude unimportant and irrelevant knowledge. *Ideally*, this body of knowledge should form a single theory. We are far away from that stage today. Also, we had better stick to partial and problematic theories rather than opting for a unified but weaker theory—unification is only a relative merit. Nevertheless, striving for unification is important because it reveals problems and sometimes enables us to solve them. That the adequate body of knowledge is located at the proverbial end of time, should not prevent us from striving towards it.

Note that the adequacy of a body of knowledge is a function of the world as well as of our knowledge capacities. The latter are limited. This is why the world may be so complex that we cannot consistently describe it by the means available to us, for example denumerable languages. If the inconsistent description is richer and more precise than any consistent one, then every scientist will obviously opt for the inconsistent description.

In the ideal body of knowledge, all theories (logical, mathematical, and empirical) should form a coherent structure. What will matter most are obviously the empirical theories. But these will require mathematical theories and the empirical theories, together with the mathematical ones, will require logical theories. If this would reveal that there is a unique true logic, then I shall gladly admit that there is a true logic.

The reader may have read the previous paragraph as saying that mathematics is the servant of empirical theories and that logic is the servant of both (remember that Thomas Aquinas saw philosophy as the servant—"ancilla" he said, which is a female and not the female of "servus"—of theology). This is not what I meant. What I did mean, however, is that the world, as knowable by us, who are parts of the world, is the correct criterion.

Does this mean that, in the end, I favour monologism? By no means. First, *we* are not and never shall be at the proverbial end of time. So we have no idea of what 'the true logic' is, if there is one. Next, even at the proverbial end of time, it is still possible that different theories will require different logics. Even at the proverbial end of time, some theories may require a different underlying logic than others. Coherence only supposes that, if a theory T is used to formulate a theory T', then T has a logic that is at least as strong as the underlying logic of T'.[30]

The third argument against end-of-time monologism deserves a separate paragraph. Suppose that a unique logic turns out to be revealed by the adequate body of

[30] If the underlying logic of T is adaptive—see Sect. 1.7—then the logic of T' should not be stronger that the adaptive logic's lower limit.

knowledge at the end of time. So this is ‘the true logic’. Yet, what use is it to us? We are not in the ideal end-of-time situation. We have to cope with the present transitory theories. Handling these may require (defeasible as well as deductive) logics that are very different from the true logic. This is an understatement. Even physics is not unified today. There never was as much disagreement on its fundamentals than in our era—just compare string theories (and their difficulties) with particle physics. That the end-of-time logic might be adequate for all contemporary theories seems (put politely) unlikely or (put bluntly) nonsensical.

I have presented two lines of argument. I hope they made the reader doubt, and hence think—all a philosopher can hope is to make his audience think. Yet, I realize very well that there is a question that should be answered by me. If there indeed is this plurality of logics, defining meanings for logical terms, in which way should we choose which logic applies in which context?

The “we” being ambiguous, let me disambiguate it. Non-logicians make the choice intuitively in terms of their learned implicit reasoning competence. Whether this is better or worse is not our concern. If it is worse, we logicians have to teach them. For us logicians, the task is straightforward but not simple: we have to study properties of logics to make a justified *choice* possible. Needless to say, lots of work still has to be done. I realize this from personal experience. For many years I have been teaching my freshmen the logic **PCR**, which is the extension of propositional **CL** with a very simple but relevant implication. I think this logic is able to capture most of natural language reasoning, but I admit that a systematic study is lacking.

By all means, the criterion for choosing between logics is *satisficing* rather than *optimizing*. Given the obviously lacking survey of all possibilities, optimizing is simply out of reach; satisficing to the contrary is sufficient. This solves many problems. For example, it relieves us from the impracticable task to find a weighed average of the different merits of different logics.

In practice, things are rather simple; the motto is: pick a choice and look for counterarguments. Even if this leads to a choice that is ‘pragmatic’ and provisional, not much harm will be done provided one keeps track of what happens if things go wrong. Allow me to give a rather personal example here. I think I found a way to preserve most of Peano Arithmetic even if Peano Arithmetic is inconsistent—I referred to this in Sect. 1.7. The means are to replace Peano Arithmetic by a **CL**-equivalent axiom system and to replace **CL** by a specific inconsistency-adaptive logic. From a pragmatic point of view, however, I would advise a mathematician to go on using **CL** as underlying logic (and hence not to keep track of the conditions that the inconsistency-adaptive logic requires). I know this will sound outrageous to my paraconsistent friends. And yet, I have a good reason for this advice. As long as no inconsistency is derived from Peano Arithmetic, **CL** will enable us to derive theorems of Peano Arithmetic that would also be derivable at-a-stage by the inconsistency-adaptive logic—never mind what this logic precisely is. Once an inconsistency is (some would like to say “were”) derived from the Peano axioms, there is an algorithm for transforming every **CL**-proof into the inconsistency-adaptive proof and for deciding which lines of the inconsistency-adaptive proof are marked. So writing **CL**-proofs is harmless, provided we realize that an inconsistency may turn up and that we know what we have to do in that case.

Acknowledgements Research for this paper was supported by subventions from the Fund for Scientific Research—Flanders and from Ghent University. I am indebted to Peter Verdée and to the two referees for careful comments on previous drafts.

References

Batens, D. (1983). Incommensurability is not a threat to the rationality of science or to the anti-dogmatic tradition. *Philosophica, 32*, 117–132.

Batens, D. (1985). Meaning, acceptance, and dialectics. In J. C. Pitt (Ed.), *Change and progress in modern science* (pp. 333–360). Dordrecht: Reidel.

Batens, D. (1992a). Do we need a hierarchical model of science? In J. Earman (Ed.), *Inference, explanation, and other frustrations: Essays in the philosophy of science* (pp. 199–215). Berkeley: University of California Press.

Batens, D. (1992b). *Menselijke kennis. Pleidooi voor een bruikbare rationaliteit* (2nd ed., 2004; 3rd ed., 2008). Antwerpen/Apeldoorn: Garant.

Batens, D. (2000a). On the epistemological justification of pluralism and tolerance. In *Philosophie et Tolérance. Philosophy and tolerance. Actes des Entretiens de Rabat, vol. I* (Philosophica, Vol. 65, pp. 33–54). Gent: Ghent University Appeared 2002.

Batens, D. (2000b). A survey of inconsistency-adaptive logics. In D. Batens, C. Mortensen, G. Priest, & J. P. Van Bendegem (Eds.), *Frontiers of paraconsistent logic* (pp. 49–73). Baldock: Research Studies.

Batens, D. (2001). A general characterization of adaptive logics. *Logique et Analyse, 173–175*, 45–68. Appeared 2003.

Batens, D. (2003). A formal approach to problem solving. In C. Delrieux & J. Legris (Eds.), *Computer modeling of scientific reasoning* (pp. 15–26). Bahia Blanca: Universidad Nacional del Sur.

Batens, D. (2004). The need for adaptive logics in epistemology. In D. Gabbay, S. Rahman, J. Symons, & J. P. Van Bendegem (Eds.), *Logic, epistemology and the unity of science* (pp. 459–485). Dordrecht: Kluwer Academic.

Batens, D. (2005). A procedural criterion for final derivability in inconsistency-adaptive logics. *Journal of Applied Logic, 3*, 221–250.

Batens, D. (2007a). Content guidance in formal problem solving processes. In O. Pombo & A. Gerner (Eds.), *Abduction and the process of scientific discovery* (pp. 121–156). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Batens, D. (2007b). A universal logic approach to adaptive logics. *Logica Universalis, 1*, 221–242.

Batens, D. (2011). Logics for qualitative inductive generalization. *Studia Logica, 97*(1), 61–80.

Batens, D. (2014). *Adaptive logics and dynamic proofs. Mastering the dynamics of reasoning*. In E. Weber, D. Wouters, & J. Meheus (Eds.), *Logic, reasoning, and rationality* (Vol. 5). Dordrecht: Springer. (In progress, parts available at http://logica.ugent.be/adlog/book.html).

Batens, D., & Haesaert, L. (2001). On classical adaptive logics of induction. *Logique et Analyse, 173–175*, 255–290. Appeared 2003.

Batens, D., & Provijn, D. (2001). Pushing the search paths in the proofs. A study in proof heuristics. *Logique et Analyse, 173–175*, 113–134. Appeared 2003.

Heeffer, A. (2007). Abduction as a strategy for concept formation in mathematics: Cardano postulating a negative. In O. Pombo & A. Gerner (Eds.), *Abduction and the process of scientific discovery* (pp. 179–194). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Heeffer, A. (2010). The symbolic model for algebra: Functions and mechanisms. In L. Magnani, W. Carnielli, & C. Pizzi (Eds.), *Model-based reasoning in science and technology. Abduction, logic, and computational discovery* (Studies in computational intelligence, Vol. 314, pp. 519–532). Heidelberg: Springer.

Hintikka, J. (1999). *Inquiry as inquiry: A logic of scientific discovery*. Dordrecht: Kluwer.

Meheus, J., & Batens, D. (1996). Steering problem solving between Cliff incoherence and Cliff solitude. *Philosophica, 58*, 153–187. Appeared 1998.

Priest, G. (2006). *In contradiction: A study of the transconsistent* (Second expanded ed.; 1st ed., 1987). Oxford: Oxford University Press.

Routley, R., & Meyer, R. K. (1976). Every sentential logic has a two-valued worlds semantics. *Logique et Analyse, 71*, 345–364.

Shapere, D. (2004). Logic and the philosophical interpretation of science. In P. Weingartner (Ed.), *Alternative logics: Do sciences need them?* (pp. 41–54). Berlin/Heidelberg: Springer.

Suszko, R. (1977). The Fregean axiom and Polish mathematical logic in the 1920s. *Studia Logica, 36*, 377–380.

Verdée, P. (2009). Adaptive logics using the minimal abnormality strategy are Π^1_1-complex. *Synthese, 167*, 93–104.

Verdée, P. (2013). A proof procedure for adaptive logics. *Logic Journal of the IGPL, 21*, 743–766.

Weber, Z. (2010). Transfinite numbers in paraconsistent set theory. *Review of Symbolic Logic, 3*, 71–92.

Wiśniewski, A. (1995). *The posing of questions: Logical foundations of erotetic inferences*. Dordrecht: Kluwer.

Wiśniewski, A. (1996). The logic of questions as a theory of erotetic arguments. *Synthese, 109*, 1–25.

Wiśniewski, A. (2003). Erotetic search scenarios. *Synthese, 134*, 389–427.

Wiśniewski, A. (2004). Erotetic search scenarios, problem-solving, and deduction. *Logique et Analyse, 185–188*, 139–166. Appeared 2005.

Chapter 2
A New Approach to Epistemic Logic

Giovanna Corsi and Gabriele Tassi

2.1 Introduction

Reasoning about knowledge by the help of logical notions and tools has originated a mess of different approaches to knowledge depending, among other things, on the intended applications: ordinary language, artificial intelligence, game theory, communication protocols. Various types of logics have been introduced starting with epistemic logics in the style of Hintikka (1962), then multi-agent logics and common knowledge logics in the style of Fagin et al. (1995). This last book has set the agenda for future research up to the present days and this paper locates itself in its wake.

Typically, the first step of every approach considered consists in setting the appropriate language in order to deal with the chosen aspect or variant of the notion of knowledge under study. As a matter of fact most of the languages are propositional languages obtained by adding to the boolean connectives a finite set of *modal* operators. In the case of epistemic logic these operators are indexed by agents $K_i, K_j, \ldots$

$$K_i(A)$$

agent i knows that A

When we move to first-order level, quantification is allowed with respect to A but not with respect to the agents, we can say that 'i knows that someone is P', but not that '*someone* knows that someone is P'.

We will take a quite different approach by introducing epistemic operators indexed by terms analogous to the indexed modal operators for alethic modalities.

G. Corsi (✉) • G. Tassi
Dipartimento di Filosofia e Comunicazione, Università di Bologna, Bologna, Italy
e-mail: giovanna.corsi@unibo.it; gabriele.tassi@studio.unibo.it

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_2,
© Springer Science+Business Media Dordrecht 2014

In the case of alethic modalities, see Corsi (2010), $\Box P(x)$ is not a well-formed formula since x is free in $P(x)$ and it has to be replaced by

$$|\, x \,|\, P(x)$$

to be read as

'it is necessary for x to be $P(x)$'.

$|\, x \,|$ is a box-operator indexed by x. A more complex form of the box-operator is the following one

$$|\, {}^{i}_{x} \,|\, P(x)$$

'it is necessary for the individual i to have the property $\lambda x.P(x)$'.

Dually,

$$\langle\, {}^{i}_{x} \,\rangle\, P(x)$$

'it is possible for i to have the property $\lambda x.P(x)$'.

Again,

$$|\, {}^{i}_{x}\; {}^{j}_{y} \,|\, R(x, y)$$

'it is necessary for i and j to stand in the relation $\lambda x \lambda y.R(x, y)$'.

In the case of epistemic modalities we need to distinguish the agent of the act of knowing from the objects of knowledge, therefore epistemic operators will have the form

$$|\, t \;:\; {}^{t_1}_{x_1} \ldots {}^{t_n}_{x_n} \,|\, A$$

t knows of $t_1 \ldots t_n$ that A.

where $x_1 \ldots x_n$ is a list of variables without repetitions that may contain also variables occurring in t, and A contains at most the variables $x_1 \ldots x_n$.

Features of the notation just introduced:

- The epistemic operator binds the variables $x_1, \ldots, x_n$ occurring in A
- The variables occurring in $t, t_1, \ldots, t_n$ are the free variables of $|t \;:\; {}^{t_1}_{x_1} \ldots {}^{t_n}_{x_n} \,|\, A$
- If A is a sentence $|\, t \;:\; |\, A$ is well formed, 't knows that A'
- By convention $|\, x \;:\; x_1 \ldots x_n \,|\, A$ stands for $|\, x \;:\; {}^{x_1}_{x_1} \ldots {}^{x_n}_{x_n} \,|\, A$
- *de re/de dicto* distinction
  de re $|\, t : {}^{i}_{x} \,|\, Px$ 't knows of i that (s)he is P'
  de dicto $|\, t : \,|\, Pi$ 't knows that Pi'

- Substitution is indicated inside the epistemic operator, it is not carried out in A

$$(|\,x\ :\ x_1 \dots x_n\,|A)[t/x, t_1/x_1 \dots t_n/x_n] := |\,t\ :\ {}^{t_1}_{x_1} \dots {}^{t_n}_{x_n}\,|\,A$$

$$(|\,t\ :\ {}^{t_1}_{x_1} \dots {}^{t_n}_{x_n}\,|A)\,[s/y] := |\,t[s/y] :\ {}^{t_1[s/y]}_{x_1} \dots {}^{t_n[s/y]}_{x_n}\,|\,A$$

- Substitution does not commute with epistemic operators

$$|\,t :\ {}^{i}_{x}\,|Px \not\leftrightarrow |\,t :\ |\,Pi$$

  We need to add specific axioms if we want substitution to commute with epistemic-operators.

Before giving the formal definition of a first-order epistemic language with indexed knowledge operators, let us look at some examples.

All Mary's friends know that she likes Paul

$$\forall x(\textsc{friend}(x, \mathit{Mary}) \rightarrow |\,x :\ {}^{\mathit{Mary}}_{y}\,|\,\textsc{likes}(y, \mathit{Paul}))$$

and this sentence is not equivalent to

$$\forall x(\textsc{friend}(x, \mathit{Mary}) \rightarrow |\,x :\ \ |\,\textsc{likes}(\mathit{Mary},\ \mathit{Paul}))$$

In the latter sentence *Mary* is in a *de dicto* position, in the former sentence in a *de re* position.

Someone knows that all Peter's friends know that he likes Mary

$$\exists x\,|\,x :\ {}^{Peter}_{y}\,|\,\forall z(\textsc{friend}(z, y) \rightarrow |\,z : y\,|\,\textsc{likes}(y, \mathit{Mary}))$$

Someone knows somebody who is late

$$\exists x \exists y\,|\,x : y\,|\,\textsc{late}(y)$$

Someone knows who Dr Smith is

$$\exists x \exists y\,|\,x : y\,|\,(y = \mathit{Dr\ Smith})$$

Peter knows that he is Peter

$$|\,Peter :\ {}^{Peter}_{x}\,|\,(x = \mathit{Peter})$$

All experts known by Peter know that smoking is dangerous

$$\forall x(\textsc{expert}(x) \wedge \exists y\,|\,\mathit{Peter} : x, y\,|\,(x = y) \rightarrow |\,x\ :\,|\,\textsc{dangerous}(\mathit{smoking}))$$

2.2 Language

Definition 1.

- *Terms* are either variables or individual constants and the set of free variables occurring in a term t, $fv(t)$, is either $\{t\}$ if t is a variable or the empty set, otherwise.
- The *logical symbols* are $\bot, \rightarrow, \forall, |\, t : \genfrac{}{}{0pt}{}{t_1}{x_1} \ldots \genfrac{}{}{0pt}{}{t_n}{x_n} \,|$, $n \geq 0$, where $x_1, \ldots, x_n$ is a list of pairwise distinct variables and $t, t_1, \ldots, t_n$ are terms. When $n = 0$ we write $|\, t : \,|$.

Definition 2. *Well formed formula* and *free variable in a wff.*

wff	*free variables*
$\bot$	$fv(\bot) = \emptyset$
$P^n t_1, \ldots, t_n$	$fv(P^n t_1, \ldots, t_n) = fv(t_1) \cup \cdots \cup fv(t_n)$
$A \rightarrow B$	$fv(A \rightarrow B) = fv(A) \cup fv(B)$
$\|\, t : \genfrac{}{}{0pt}{}{t_1}{x_1} \ldots \genfrac{}{}{0pt}{}{t_n}{x_n} \,\| A$	where $fv(A) \subseteq \{x_1, \ldots, x_n\}$
	$fv(\|\, t : \genfrac{}{}{0pt}{}{t_1}{x_1} \ldots \genfrac{}{}{0pt}{}{t_n}{x_n} \,\| A) = fv(t) \cup fv(t_1) \cup \cdots \cup fv(t_n)$
$\forall x A$	$fv(\forall x A) = fv(A) - \{x\}$

$\neg A$, $A \vee B$, $A \wedge B$, $A \leftrightarrow B$, $\exists x A$, $\langle t : \genfrac{}{}{0pt}{}{t_1}{x_1} \ldots \genfrac{}{}{0pt}{}{t_n}{x_n} \rangle A$ are defined as usual and $|\, x : x_1 \ldots x_n | A$ and $\langle x : x_1 \ldots x_n \rangle A$ stand for $|\, x : \genfrac{}{}{0pt}{}{x_1}{x_1} \ldots \genfrac{}{}{0pt}{}{x_n}{x_n} \,| A$ and $\langle x : \genfrac{}{}{0pt}{}{x_1}{x_1} \ldots \genfrac{}{}{0pt}{}{x_n}{x_n} \rangle A$, respectively.

Definition 3. *Simultaneous substitution.* Given a wff A containing the free variables $x_1, \ldots, x_k$, we define the wff $A[s_1/x_1 \ldots s_k/x_k]$ where the term s_i is substituted for x_i, $1 \leq i \leq k$. Let $[\mathbf{s}/\mathbf{x}] =_{df} [s_1/x_1 \ldots s_k/x_k]$.

- $\bot\, [\mathbf{s}/\mathbf{x}] = \bot$
- $(P^n t_1, \ldots, t_n)[\mathbf{s}/\mathbf{x}] = P^n(t_1[\mathbf{s}/\mathbf{x}], \ldots, t_n[\mathbf{s}/\mathbf{x}])$, where
  - $t_i[\mathbf{s}/\mathbf{x}] = s_i$ if $t_i = x_j \in \{x_1, \ldots, x_k\}$
  - $t_i[\mathbf{s}/\mathbf{x}] = t_i$ if $t_i \not\in \{x_1, \ldots, x_k\}$
- $(A \rightarrow B)[\mathbf{s}/\mathbf{x}] = (A[\mathbf{s}/\mathbf{x}] \rightarrow B[\mathbf{s}/\mathbf{x}])$
- $(\forall y A)[\mathbf{s}/\mathbf{x}] =$

$$= \begin{cases} \forall y (A[\mathbf{s}/\mathbf{x}]) & \text{if } y \notin (\{x_1, \ldots, x_k\} \cup \{s_1, \ldots, s_k\}) \\ \forall z ((A[z/y])[\mathbf{s}/\mathbf{x}] & \text{where } z \text{ doesn't occur in } \forall y A \textit{ and } z \notin \{s_1, \ldots, s_k\} \\ & \text{if } y \not\in \{x_1, \ldots, x_k\} \text{ and } y \in \{s_1, \ldots, s_k\} \\ \forall y A & \text{if } y \in \{x_1, \ldots, x_k\} \end{cases}$$

- $(|\, t : {}^{t_1}_{y_1} \ldots {}^{t_n}_{y_n} \,|\, A)[\mathbf{s}/\mathbf{x}] = |\, t[\mathbf{s}/\mathbf{x}] : {}^{t_1[\mathbf{s}/\mathbf{x}]}_{y_1} \ldots {}^{t_n[\mathbf{s}/\mathbf{x}]}_{y_n} \,|\, A$

2.3 Semantics

The main idea behind the *epistemic transition semantics* is that

$$|\, t : {}^{s}_{x} |\, P(x)$$

is true at a world w if t is an individual existing at w, s is an individual existing at w and in all worlds compatible with the epistemic state of t the t-counterparts of s (the counterparts of s according to t) in those worlds satisfy $P(x)$.

$$|\, t : \,|\, P(s)$$

is true at a world w if t is an individual existing at w and in all worlds compatible with the epistemic state of t whoever is s in those worlds satisfies $P(x)$.

An *epistemic transition model* (in brief, an *epistemic model*) is a family of classical models endowed with (1) a relation of *compatibility* between individuals and models and (2) a counterpart relation between individuals of different models or of the same model. We will call *worlds* the classical models, following the terminology of possible world semantics. In details, let W be a not empty set of worlds, so each $w \in W$ is a pair $\langle D_w, I_w \rangle$ where D_w is a not-empty set, the *domain* of w and I_w is an interpretation function such that:

- For every relation P^n, $I_w(P^n) \subseteq (D_w)^n$
- $I_w(=) = \{\langle a, a \rangle : a \in D_w\}$
- For every individual constant i, $I_w(i) \in D_w$

We assume that $D_w \cap D_v = \emptyset$ when $w \neq v$. By $\prec$ we denote a relation between elements of $\mathscr{E} = \bigcup\{D_w\}_{w \in W}$ and elements of W:

$$\prec \subseteq (\mathscr{E} \times \mathscr{W}).$$

If $a \prec v$ holds, then we say that the world v is *epistemically compatible with the individual* a or that v is compatible with the epistemic state of a. By $\overset{a}{\rightarrowtail}$ we denote the counterpart relation parametrized by the individual a:

$$\overset{a}{\rightarrowtail} \subseteq \bigcup\{D_w \times D_v : a \in D_w \wedge a \prec v\}$$

If $a, b \in D_w$, $c \in D_v$ and $b \overset{a}{\rightarrowtail} c$ holds, then we say that c is a counterpart of b according to a (in a world epistemically compatible with a).

Definition 4. An epistemic transition model $\mathscr{M} = \langle W, \prec, \rightarrowtail, D, I\rangle$ is a quintuple where W and $\prec$ are defined as above, $\rightarrowtail \;=\; \bigcup\{\overset{a}{\rightarrowtail}\}_{a\in\mathscr{E}}$, D is a function that associates to any $w \in W$ its domain D_w and I is a function that associates to any $w \in W$ its interpretation function I_w.

Definition 5. For every $w \in W$, a *w-assignment* is a function $\sigma : VAR \rightarrow D_w$. If σ is a w-assignment, $\sigma^{x \rhd d}$ denotes the w-assignment which behaves exactly like σ except that it maps x to $d \in D_w$.

Given a w-assignment σ the *interpretation* of t in w under σ, $I^{\sigma}_{w}(t)$, is defined in the standard way:

- $I^{\sigma}_{w}(x) = \sigma(x)$
- $I^{\sigma}_{w}(i) = I_w(i)$

Notational convention. When no ambiguity can arise, we write $\sigma(t)$ instead of $I^{\sigma}_{w}(t)$.

Definition 6. *Satisfaction.* We define when a wff A is *satisfied at* w *by* a w-assignment σ in an epistemic model $\mathscr{M}$, $\sigma \models^{\mathscr{M}}_{w} A$.

$\sigma \not\models^{\mathscr{M}}_{w} \bot$

$\sigma \models^{\mathscr{M}}_{w} P^k(t_1 \dots t_k)$ iff $\sigma(t_1, \dots, t_k) \in I_w(P^k)$

$\sigma \models^{\mathscr{M}}_{w} B \rightarrow G$ iff $\sigma \not\models^{\mathscr{M}}_{w} B$ or $\sigma \models^{\mathscr{M}}_{w} G$

$\sigma \models^{\mathscr{M}}_{w} \forall x G$ iff for all $d \in D_w$, $\sigma^{x \rhd d} \models^{\mathscr{M}}_{w} G$

$\sigma \models^{\mathscr{M}}_{w} |t : {}^{t_1}_{y_1} \dots {}^{t_n}_{y_n}| G$ iff for all v such that $\sigma(t) \prec v$, and all v-assignments τ such that $\sigma(t_1)\overset{\sigma(t)}{\rightarrowtail}\tau(y_1), \dots, \sigma(t_n)\overset{\sigma(t)}{\rightarrowtail}\tau(y_n)$, then $\tau \models^{\mathscr{M}}_{v} G$

Definition 7.

- A wff A is *true at* w in $\mathscr{M}$, $\models^{\mathscr{M}}_{w} A$, iff for every w-assignment σ, $\sigma \models^{\mathscr{M}}_{w} A$.
- A wff A is *true* in $\mathscr{M}$, $\models^{\mathscr{M}} A$, iff for every w, $\models^{\mathscr{M}}_{w} A$.
- A wff A is *valid* on a class C of epistemic transition models iff A is true in each of them.

Lemma 1. Substitution and satisfaction for terms and formulas. *Let σ be a w-assignment.*

$$\sigma(t[s/x]) \;=\; \sigma^{x \rhd \sigma(s)}(t)$$

$$\sigma \models_w A[s/x] \quad iff \quad \sigma^{x \rhd \sigma(s)} \models_w A$$

Proof. By induction on A.

- $A = P^n(t_1, \dots, t_n)$
  $\sigma^{x \rhd \sigma(s)} \models_w P^n(t_1, \dots, t_n)$ iff $\langle \sigma^{x \rhd \sigma(s)}(t_1), \dots, \sigma^{x \rhd \sigma(s)}(t_n) \rangle \in I_w(P^n)$ iff $\langle \sigma(t_1[s/x]), \dots, \sigma(t_n[s/x]) \rangle \in I_w(P^n)$ iff $\sigma \models_w P^n(t_1[s/x], \dots, t_n[s/x])$ iff $\sigma \models_w P^n(t_1, \dots, t_n)[s/x]$.
- $A = \forall y B$ and $y \neq s$ and $y \neq x$
  $\sigma^{x \rhd \sigma(s)} \models_w \forall y B$ iff for all $d \in D_w$, $\sigma^{x \rhd \sigma(s), y \rhd d} \models_w B$ iff for all $d \in D_w$, $\sigma^{y \rhd d, x \rhd \sigma(s)} \models_w B$ iff by induction hypothesis for all $d \in D_w$, $\sigma^{y \rhd d} \models_w B[s/x]$ iff $\sigma \models_w \forall y (B[s/x])$ iff by def. of substitution $\sigma \models_w (\forall y B)[s/x]$.
  The cases in which either $y = s$ or $y = x$ are similar.
- $A = |\genfrac{}{}{0pt}{}{t_1}{y_1} \dots \genfrac{}{}{0pt}{}{t_n}{y_n}|B$
  $\sigma^{x \rhd \sigma(s)} \models_w |t : \genfrac{}{}{0pt}{}{t_1}{y_1} \dots \genfrac{}{}{0pt}{}{t_n}{y_n}|B$ iff

  $\tau \models_v B$ for all v-assignment τ such that $\sigma^{x \rhd \sigma(s)}(t_i) \overset{\sigma^{x \rhd \sigma(s)}(t)}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$, iff

  $\tau \models_v B$ for all v-assignment τ such that $\sigma(t_i[s/x]) \overset{\sigma(t[s/x])}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$, iff
  $\sigma \models_w |t[s/x], \genfrac{}{}{0pt}{}{t_1[s/x]}{y_1} \dots \genfrac{}{}{0pt}{}{t_n[s/x]}{y_n}|B$ iff
  $\sigma \models_w (|t, \genfrac{}{}{0pt}{}{t_1}{y_1} \dots \genfrac{}{}{0pt}{}{t_n}{y_n}|B)[s/x]$.

□

2.3.1 Validity

The epistemic semantics we have seen so far is a generalization of the transition semantics presented in Corsi (2010) and at the same time a particular case of a more general semantics called *cone transition semantics* due to Gabriele Tassi (Tassi forthcoming; Corsi and Tassi 2010). Most of the results proved in Corsi (2010) hold for the epistemic case. The main difference with respect to transition semantics is that the accessibility relation among worlds is parametrized by individuals. We do not say anymore that a world w is related to or accessible to another world v, but rather that v is compatible with the epistemic state of an individual a living in w. Moreover, as we have seen, also the counterpart relation is parametrized by individuals, so we speak of the a-counterpart of b, meaning the counterpart of b according to a, parametrized by a.

Notice first that no condition has been put in order to establish some connections between the counterparts in a world v of an individual b living in w and the interpretation of b in v. This fact has the consequence that the following two types of knowledge are quite different:

$$|i : \genfrac{}{}{0pt}{}{t}{x_1} \genfrac{}{}{0pt}{}{s}{x_2}|(x_1 = x_2) \qquad i \text{ knows of } t \text{ and } s \text{ that they are equal}$$

$$|i : \;|(t = s) \qquad i \text{ knows that } t \text{ is equal to } s$$

The first sentence is true at a world w iff in all worlds v compatible with the epistemic state of $\underline{i}$, all the $\underline{i}$-counterparts in v of $\underline{t}$ and $\underline{s}$ (the interpretation of t and s in w) are identical. The second sentence is true at w iff in all worlds v compatible with the epistemic state of $\underline{i}$ the interpretation of t and s in v are identical.

For particular individual constants i, t and s we can assume that the $\underline{i}$-counterparts in a world v of $\underline{t}$ in w include the interpretation of t in v. A consequence is that the wff

$$|i : {}^{t}_{x_1}\, {}^{s}_{x_2}\,| (x_1 = x_2) \rightarrow |i : \ | (t = s)$$

is valid. When this is the case we say that the terms t and s are i-rigid, i.e. are *rigid terms* from the point of view of i. For some student i it might well be that if (s)he knows of Walter Scott and Ivanhoe that the first is the author of the second, than (s)he knows also the fact that Walter Scott is the author of Ivanhoe, because in the worlds (s)he can envisage the counterparts of both Walter Scott and Ivanhoe include the interpretations of both names in those worlds.

We can impose even stronger constraints on the counterpart relation, e.g. that the $\underline{i}$-counterpart in a world v of the interpretations in w of t and s coincide with the interpretations of t and s, respectively, in v. For example if t and s are numbers, say 9 and 7, we may want that for any individual i, the $\underline{i}$-counterpart in a world v of the interpretations of 9 and 7 in w coincide with the interpretation of 9 and 7 in v, respectively. When this is the case the terms 9 and 7 are said to be *i-stable* and the following formula is valid

$$|i : {}^{9}_{x_1}\, {}^{7}_{x_2}\,| A(x_1, x_2) \leftrightarrow |i : \ | A(9, 7)$$

This equivalence doesn't hold in general, not even for variables, instead the following implication, say from *de re* to *de dicto*, holds for variables:

$$RG_e^v \qquad |t : {}^{y_1}_{x_1} \ldots {}^{y_n}_{x_n}\,| A \rightarrow |\, y : v_1 \ldots v_k| (A[y_1/x_1 \ldots y_n/x_n])$$

where $v_1 \ldots v_k$ are the variables $y_1 \ldots y_n$ without repetitions.

Therefore we say that variables are rigid designators. In the case of aletic modalities it is often assumed that all terms, not just variables, are rigid designators and so the following formula is taken as an axiom

$$RG \qquad |{}^{t_1}_{x_1} \ldots {}^{t_n}_{x_n}\,| A \rightarrow |\, v_1 \ldots v_k\,| (A[t_1/x_1 \ldots t_n/x_n])$$

where $v_1 \ldots v_k$ are the variables occurring in $t_1, \ldots t_n$.

The rigidity axiom is untenable, in general, in the epistemic case:

$$RG_e \qquad |t : {}^{t_1}_{x_1} \ldots {}^{t_n}_{x_n}\,| A \rightarrow |\, v_1 \ldots v_k\,|\, y : (A[t_1/x_1 \ldots t_n/x_n])$$

Let us stress that the converse of RG_e^v is not valid, just consider the following instance:

$$|\, i : y \,|\, (y = y) \rightarrow |\, i : {}^{y}_{x_1}{}^{y}_{x_2} \,|\, (x_1 = x_2)$$

It is certainly true that in all worlds compatible with the epistemic state of $\underline{i}$, each individual is identical with itself, but at the same time if $\underline{y}$ has two different $\underline{i}$-counterparts in a world v, then $(x_1 = x_2)$ may be falsified in v.

2.4 The Epistemic Logic $Q.K_e$

Now we present a calculus for epistemic logic which makes no assumptions either on the compatibility relation or on the counterpart relation. $Q.K_e$ intends to be the core system of any quantified logic either of belief or of knowledge or of obligation. We can think of weaker systems than $Q.K_e$ in the style of Gabriele Tassi (forthcoming), where the greater generality of Tassi's systems resides in the fact that the epistemic operators are indexed by lists of terms and not by pairs composed of a term and a set of terms, as we do, see axiom PRM_e.

Here are the axioms and inference rules of $Q.K_e$.

PRM$_e$ $|x : x_1 \ldots x_n \,|\, A \leftrightarrow |\, x : x_{i_1} \ldots x_{i_n} \,|\, A$
for every permutation $x_{i_1} \ldots x_{i_n}$ of $x_1 \ldots x_n$

K$_e$ $|\, x : x_1 \ldots x_n |\, (A \rightarrow B) \rightarrow (|\, x : x_1 \ldots x_n \,|\, A \rightarrow |\, x : x_1 \ldots x_n \,|\, B)$

UI $\forall x A(x) \rightarrow A$

LNGT$_e$ $|\, x : x_1 \ldots x_n \,|\, A \rightarrow |\, x : x_1 \ldots x_n, x_{n+1} \,|\, A$

RG_e^v $|\, y : {}^{y_1}_{x_1} \ldots {}^{y_n}_{x_n} \,|\, A \rightarrow |\, y : v_1 \ldots v_k \,|\, (A[y_1/x_1 \ldots y_n/x_n])$
where $v_1 \ldots v_k$ are the variables $y_1 \ldots y_n$ without repetitions.

ID $x = x$

LBZ $t = s \rightarrow (A[t/x] \rightarrow A[s/x])$

$$\frac{A \quad A \rightarrow B}{B} \quad (MP)$$

$$\frac{A}{|\, x : x_1 \dots x_n \,|\, A} \quad (N_e) \quad \text{provided } \{x_1, \dots, x_n\} \supseteq fv(A).$$

$$\frac{A \to B}{A \to \forall x B} \quad (UG) \quad \text{provided } x \notin fv(A)$$

$$\frac{A}{A[s/x]} \quad (SFV)$$

The notions of *proof* and *theorem* are defined in the usual way.

2.5 Completeness of $Q.K_e$

The completeness proof we present follows the same strategy of the proof given in Corsi (2010) and in Braüner and Ghilardi (2007). Given a language with indexed operators $\mathscr{L}$, we define a classical first-order language $\overline{\mathscr{L}}$ which contains the same predicate and constant symbols of $\mathscr{L}$, and moreover for each formula $|\, t : x_1 \dots x_n \,|\, A$ of $\mathscr{L}$ a new predicate symbol $P^{n+1}_{|\,:x_1 \dots x_n \,|\, A}$. Then we translate each formula of $\mathscr{L}$ into a formula of $\overline{\mathscr{L}}$ according to the following definition:

Definition 8.

$$\begin{array}{lcl}
\overline{\bot} & = & \bot \\
\overline{P^n(t_1, \dots, t_n)} & = & P^n(t_1, \dots, t_n) \\
\overline{s = t} & = & s = t \\
\overline{(A \to B)} & = & \overline{A} \to \overline{B} \\
\overline{\forall x_i A} & = & \forall x_i \overline{A} \\
\overline{|\, t : {}^{t_1}_{x_1} \dots {}^{t_n}_{x_n} \,|\, A} & = & P^{n+1}_{|\,:x_1 \dots x_n \,|\, A}(t, t_1, \dots, t_n)
\end{array}$$

Lemma 2. $\overline{A[s/x]} = \overline{A}[s/x]$, *for all formulas* $A \in \mathscr{L}$.

Proof. By induction on A.

- $\overline{P(t_1, \dots, t_n)[s/x]} = \overline{P(t_1[s/x], \dots, t_n[s/x])} = P(t_1[s/x], \dots, t_n[s/x]) = P(t_1, \dots, t_n)[s/x] = \overline{P(t_1, \dots, t_n)}[s/x]$
- Let $y \neq x$ and $y \neq s$. $\overline{(\forall y B)[s/x]} = \overline{\forall y (B[s/x])} = \forall y \overline{(B[s/x])} = \forall y (\overline{B}[s/x]) = (\forall y \overline{B})[s/x] = \overline{\forall y B}[s/x]$

  The other cases relative to quantified formulas are similar.
- $\overline{(|\, t : {}^{t_1}_{x_1} \dots {}^{t_n}_{x_n} \,|\, B)[s/x]} = \overline{(|\, t[s/x] : {}^{t_1[s/x]}_{x_1} \dots {}^{t_n[s/x]}_{x_n} \,|\, B)} =$
  $P^{n+1}_{|\,:x_1 \dots x_n \,|\, B}(t[s/x], t_1[s/x], \dots, t_n[s/x]) =$
  $(P^{n+1}_{|\,:x_1 \dots x_n \,|\, B}(t, t_1, \dots, t_n))[s/x] = \overline{(|\, t : {}^{t_1}_{x_1} \dots {}^{t_n}_{x_n} \,|\, B)}[s/x]$

□

We now define a classical theory $\overline{Q.K_e}$ whose specific axioms are

$$\{\overline{A} \,:\, Q.K_e \vdash A\}$$

Lemma 3. $X \vdash_{Q.K_e} A$ *iff* $\overline{X} \vdash_{\overline{Q.K_e}} \overline{A}$.

Proof. We show that $\vdash_{Q.K_e} B_1 \wedge \cdots \wedge B_n \rightarrow A$ iff $\vdash_{\overline{Q.K_e}} \overline{B_1 \wedge \cdots \wedge B_n \rightarrow A}$, where $B_1, \ldots, B_n \in X$.

$\Rightarrow$ holds by definition of $\overline{Q.K_e}$.

$\Leftarrow$ holds because the specific axioms of $\overline{Q.K_e}$ are the translations of the theorems of $Q.K_e$ and the inference rules of $\overline{Q.K_e}$ are also inference rules of $Q.K_e$. □

Let $\mathscr{S}$ be a family of classical models for $\overline{Q.K_e}$. Each model w is a pair $\langle D_w, I_w\rangle$ where D_w is a not-empty set, the *domain* of w and I_w is an interpretation function. With $\langle \sigma, w\rangle \overset{c}{\models} B$ we denote that the formula B is satisfied by the assignment σ in the model w according to the standard classical definition and with $w \overset{c}{\models} B$ that B is (classically) true in the model w.

Lemma 4. *Let σ be a w-assignment and A a wff of $\mathscr{L}$.*

$$\langle \sigma, w\rangle \overset{c}{\models} \overline{A[s/x]} \qquad \textit{iff} \qquad \langle \sigma^{x \rhd \sigma(s)}, w\rangle \overset{c}{\models} \overline{A}$$

Proof. By induction on A. We examine the case when A is $|\, t : \genfrac{}{}{0pt}{}{t_1}{y_1} \ldots\ldots \genfrac{}{}{0pt}{}{t_n}{y_n} \,|\, C$.

Then $\langle \sigma, w\rangle \overset{c}{\models} \overline{|\, t : \genfrac{}{}{0pt}{}{t_1}{y_1} \ldots\ldots \genfrac{}{}{0pt}{}{t_n}{y_n} \,|\, C\,[s/x]}$ iff by Lemma 2

$\langle \sigma, w\rangle \overset{c}{\models} \overline{|\, t : \genfrac{}{}{0pt}{}{t_1}{y_1} \ldots\ldots \genfrac{}{}{0pt}{}{t_n}{y_n} \,|\, C}\,[s/x]$ iff

$\langle \sigma, w\rangle \overset{c}{\models} (P^{n+1}_{|:y_1 \ldots y_n\,|\,C}(t, t_1, \ldots, t_n))[s/x]$ iff by Lemma 1

$\langle \sigma^{x \rhd \sigma(s)}, w\rangle \overset{c}{\models} P^{n+1}_{|:y_1 \ldots y_n\,|\,C}(t, t_1, \ldots, t_n)$ iff

$\langle \sigma^{x \rhd \sigma(s)}, w\rangle \overset{c}{\models} \overline{|\, t : \genfrac{}{}{0pt}{}{t_1}{y_1} \ldots\ldots \genfrac{}{}{0pt}{}{t_n}{y_n} \,|\, C}$. □

Definition 9. Let w, v be $\overline{Q.K_e}$-models. For any $a \in D_w$ we say that

$$a \prec v \qquad \text{iff} \qquad v \models^c \{\overline{A} :\, \langle \sigma^{x \rhd a}, w\rangle \overset{c}{\models} \overline{|x :\ |A}\}$$

In words, v *is compatible with the epistemic state of* a iff every sentence known by a is true in v.

Definition 10. Let w, v be $\overline{Q.K_e}$-models. For any $a \in D_w$, a relation $\overset{a}{\rightarrowtail} \subseteq D_w \times D_v$ is said to be a *transition relation admissible for a* iff for every $k \geq 0$, every w-assignment σ and every v-assignment τ,

$$\langle\, \sigma(x_i),\, \tau(x_i)\,\rangle \in \overset{a}{\rightarrowtail} \qquad \text{for } i = 1, \ldots, k^{[1]}$$

[1] We also write $\sigma(x_i) \overset{a}{\rightarrowtail} \tau(x_i) \qquad \text{for } i = 1, \ldots, k$.

only if

$$\langle \sigma^{x \triangleright a}, w \rangle \overset{c}{\models} \overline{|x : x_1 \ldots x_k|A} \qquad \Rightarrow \qquad \langle \tau, v \rangle \overset{c}{\models} \overline{A}$$

holds for every formula A containing (at most) the variables $x_1, \ldots, x_k$.

In words, if $\tau(x_i)$ is a counterpart of $\sigma(x_i)$, $1 \leq i \leq k$, according to a, then if a knows of $\sigma(x_1) \ldots \sigma(x_k)$ that A, then A is satisfied in v by $\tau(x_1) \ldots \tau(x_k)$.

Lemma 5. *Let w be a $\overline{Q.K_e}$-model and $\langle \sigma, w \rangle \overset{c}{\not\models} \overline{|x : x_1 \ldots x_m|A}$ for some formula $|x : x_1 \ldots x_m|A$ and w-assignment σ. Then there is a $\overline{Q.K_e}$-model v and a v-assignment τ such that:*

1. $\langle \tau, v \rangle \overset{c}{\not\models} \overline{A}$;
2. $\sigma(x) \prec v$;
3. *The set* $\overset{\sigma(x)}{\rightarrowtail} = \{\langle \sigma(x_1), \tau(x_1) \rangle, \langle \sigma(x_2), \tau(x_2) \rangle, \ldots, \langle \sigma(x_m), \tau(x_m) \rangle\}$ *is a transition relation admissible for* $\sigma(x)$.

Proof.

- Let Γ be the following set of (classical) formulae:

$$\Gamma = \{\neg \overline{A}\} \cup \{\overline{B} : \langle \sigma, w \rangle \overset{c}{\models} \overline{|x : x_{j_1} \ldots x_{j_h}|B}, \text{where } \{x_{j_1} \ldots x_{j_h}\} \subseteq \{x_1, \ldots, x_m\}\}.$$

First we show that Γ is $\overline{Q.K_e}$-consistent. Assume *by reductio* that it is not, then:

(1) $\vdash_{\overline{Q.K_e}} \overline{B_1} \wedge \ldots \wedge \overline{B_r} \rightarrow \overline{A}$

(2) $\vdash_{Q.K_e} B_1 \wedge \ldots \wedge B_r \rightarrow A$ (3)

(3) $\vdash_{Q.K_e} |x : x_1 \ldots x_m|B_1 \wedge \ldots \wedge |x : x_1 \ldots x_m|B_r \rightarrow |x : x_1 \ldots x_m|A$ (N_e)

(4) $\vdash_{Q.K_e} |x : x_{j_1} \ldots x_{j_{h_1}}|B_1 \wedge \ldots \wedge |x : x_{j_1} \ldots x_{j_{h_r}}|B_r \rightarrow |x : x_1 \ldots x_m|A$ $(LNGT_e)$

(5) $\vdash_{\overline{Q.K_e}} \overline{|x : x_{j_1} \ldots x_{j_{h_1}}|B_1} \wedge \ldots \wedge \overline{|x : x_{j_1} \ldots x_{j_{h_r}}|B_r} \rightarrow \overline{|x : x_1 \ldots x_m|A}$ (3)

Therefore, we would have that $\langle \sigma, w \rangle \overset{c}{\models} \overline{|x : x_1 \ldots x_m|A}$ contrary to the fact that $\langle \sigma, w \rangle \overset{c}{\not\models} \overline{|x : x_1 \ldots x_m|A}$.

Since Γ is $\overline{Q.K_e}$-consistent, by classical model theory there is a model v and a v-assignment τ such that $\langle \tau, v \rangle \overset{c}{\models} \Gamma$, therefore $\langle \tau, v \rangle \overset{c}{\not\models} \overline{A}$.

- By the way Γ is defined, Γ contains all the formulae $\overline{B}$ without free variables such that $\langle \sigma, w \rangle \overset{c}{\models} \overline{|x : \ |B}$, therefore $\sigma(x) \prec v$.
- We have to show that the set $\overset{\sigma(x)}{\rightarrowtail}$ is a counterpart relation admissible for $\sigma(x)$, i.e. for any $k > 0$, any formula $C(y_1, \ldots, y_k)$, any w-assignment π and any v-assignment μ, if

$$(i) \quad \pi(y) = \sigma(x) \qquad\qquad (ii) \quad \langle \pi, w \rangle \overset{c}{\models} \overline{|y : y_1 \ldots y_k|C}$$

$$(iii) \quad \pi(y_i) \overset{\sigma(x)}{\rightarrowtail} \mu(y_i) \quad i = 1 \ldots k$$

then

$$\langle \mu, v\rangle \overset{c}{\models} \overline{C(y_1, \ldots, y_k)}.$$

By the definition of $\overset{\sigma(x)}{\rightarrowtail}$, if $\pi(y_i) \overset{\sigma(x)}{\rightarrowtail} \mu(y_i)\,, i = 1 \ldots k$, then for some $x_{j_i} \in \{x_1, \ldots, x_m\}$,

$$(a) \quad \pi(y_i) = \sigma(x_{j_i})$$

and

$$(b) \quad \mu(y_i) = \tau(x_{j_i})$$

It follows from (ii) that:

$$\langle \pi^{y_1 \triangleright \sigma(x_{j_1}) \ldots y_k \triangleright \sigma(x_{j_k})}, w\rangle \overset{c}{\models} \overline{|y : y_1 \ldots y_k|C}$$

Given that $y_1, \ldots, y_k$ are all the free variables in $\overline{C}$ and that $\pi(y) = \sigma(x)$, this is equivalent to:

$$\langle \sigma^{y \triangleright \sigma(x),\, y_1 \triangleright \sigma(x_{j_1}) \ldots y_k \triangleright \sigma(x_{j_k})}, w\rangle \overset{c}{\models} \overline{|y : y_1 \ldots y_k|C}$$

By Lemma 1 we get that:

$$\langle \sigma, w\rangle \overset{c}{\models} \overline{|y[x/y] : \substack{x_{j_1} \\ y_1} \ldots \substack{x_{j_k} \\ y_k} |C}$$

Then by MP with the (translation of the) axiom RG^v it obtains that:

$$\langle \sigma, v\rangle \overset{c}{\models} \overline{|x : v_1 \ldots v_h|(C[x_{j_1}/y_1 \ldots x_{j_k}/y_k])}.$$

Given that $\{v_1, \ldots, v_h\} \subseteq \{x_{j_1}, \ldots, x_{j_k}\} \subseteq \{x_1, \ldots, x_m\}$, it follows that

$$\overline{C[x_{j_1}/y_1 \ldots x_{j_k}/y_k]} \in \Gamma.$$

Therefore

$$\langle \tau, v\rangle \overset{c}{\models} \overline{C[x_{j_1}/y_1 \ldots x_{j_k}/y_k]}.$$

By Lemma 1 we get that:

$$\langle \tau^{y_1 \triangleright \tau(x_{j_1}) \ldots y_k \triangleright \tau(x_{j_k})}, v\rangle \overset{c}{\models} \overline{C(y_1, \ldots, y_k)}.$$

But all the free variables of $\overline{C}$ are among $y_1, \ldots, y_k$, therefore this is equivalent to:

$$\langle \mu^{y_1 \triangleright \tau(x_{j_1}) \ldots y_k \triangleright \tau(x_{j_k})}, v\rangle \overset{c}{\models} \overline{C(y_1, \ldots, y_k)}.$$

By the definition of $\overset{\sigma(x)}{\rightarrowtail}$, if $\pi(y_i) \overset{\sigma(x)}{\rightarrowtail} \mu(y_i)$ for all $i = 1 \dots k$, then, for all $i = 1 \dots k$ there is a $x_{j_i} \in \{x_i, \dots, x_m\}$ such that $\tau(x_{j_i}) = \mu(y_i)$. Therefore we have:

$$\langle \mu^{y_1 \triangleright \mu(y_1) \dots y_k \triangleright \mu(y_k)}, v\rangle \overset{c}{\models} \overline{C(y_1, \dots, y_k)}$$

i.e.

$$\langle \mu, v\rangle \overset{c}{\models} \overline{C(y_1, \dots, y_k).}$$

□

The set $\overset{\sigma(x)}{\rightarrowtail}$ as defined in Lemma 5 gives the minimal counterpart relation that links the model w to the model v in dependence of the formula A, the w-assignment σ and the individual $\sigma(x)$. Between D_w and D_v no other counterpart relation is taken into account even if extensions of $\overset{\sigma(x)}{\rightarrowtail}$ may be admissible. If $\sigma(x) = a$ for some $a \in D_w$, we call the set $\overset{a}{\rightarrowtail}$ the *canonical counterpart relation relative to a, w and v*, in brief *CNTP(a,w,v)*. Notice that if $CNTP(a, w, v) \neq \emptyset$, then $a \prec w$.

Definition 11. Let $\mathscr{S}$ be a set of $\overline{Q.K_e}$-models. We say that:

- *$w \in \mathscr{S}$ is realized in $\mathscr{S}$* iff for each w-assignment σ and each formula $|\, x : x_1 \dots x_m \,|\, A$ of $\mathscr{L}$, if $\langle \sigma, w\rangle \overset{c}{\not\models} \overline{|\, x : x_1 \dots x_m \,|\, A}$, then there is a $\overline{Q.K_e}$-model $v \in \mathscr{S}$ and a v-assignment τ such that:
  - $\sigma(x) \prec v$;
  - $\sigma(x_i) \overset{\sigma(x)}{\rightarrowtail} \tau(x_i)$, for every $x_i \in \{x_1, \dots, x_m\}$;
  - $\langle \tau, v\rangle \overset{c}{\not\models} \overline{A}$.
- *$\mathscr{S}$ is fully realized* iff every member of $\mathscr{S}$ is realized in $\mathscr{S}$ and for any $z, w \in \mathscr{S}$, if $z \neq w$ then $D_z \cap D_w = \emptyset$.

Lemma 6. *For every $\overline{Q.K_e}$-model w there is a set $\mathscr{S}^w$ of $\overline{Q.K_e}$-models such that:*

- *$w \in \mathscr{S}^w$;*
- *$\mathscr{S}^w$ is fully realized.*

Proof. We define a chain $\mathscr{S}_0, \mathscr{S}_1 \dots, \mathscr{S}_n, \dots$ of sets of classical models such that $\mathscr{S}_0 = \{w\}$ and $\mathscr{S}_{n+1}$ is obtained fron $\mathscr{S}_n$ by adding to it new $\overline{Q.K_e}$-models so as to realize the models already present in $\mathscr{S}_n$. This step is performed according to Lemma 5 taking care to choose models whose domains do not overlap the domains of the models already present in $\mathscr{S}_n$. Let $\mathscr{S}^w$ be the union of the chain. □

The fully realized set $\mathscr{S}^w$ whose elements are constructed according to Lemma 5 is said to be *canonical*. In a canonical (fully realized) set the relation $CNTP(a, w, v)$ is uniquely determined given w and v, in fact if $CNTP(a, w, v) = CNTP(b, w, v)$ then $a = b$, so as far as canonical sets are concerned, we will talk of the relation $CNTP(w, v)$.

Given a canonical set $\mathscr{S}^w$, the model $\mathscr{M}^{\mathscr{S}^w} = \langle \mathscr{S}^w, \mathscr{D}, \prec, \rightarrowtail, \mathscr{I} \rangle$ is said to be *a canonical epistemic model* if

- $\mathscr{D}$ is a function such that for every $z = \langle D_z, I_z \rangle \in \mathscr{S}^w$, $D(z) = D_z$
- $\prec = \{\langle a, v \rangle : a \in D_w, v \models^c \{\overline{A} : \langle \sigma^{x \rhd a}, w \rangle \overset{c}{\models} \overline{|x : \ |A}$, for all $A \in \mathscr{L}\}$, for some $w, v \in \mathscr{S}^w$, and w-assignment $\sigma\}$
- $\rightarrowtail \ = \bigcup \{CNTP(w, v)\}_{w,v \in S^w}$
- $\mathscr{I}$ is a function such that for every $z = \langle D_z, I_z \rangle \in \mathscr{S}^w$, $I(z) = I_z$

Lemma 7. *Given a canonical epistemic model* $\mathscr{M}^{\mathscr{S}^w} = \langle \mathscr{S}^w, \mathscr{D}, \prec, \rightarrowtail, \mathscr{I} \rangle$*, for every formula* B *of* $\mathscr{L}$ *and every z-assignment* σ*,*

$$\sigma \models_z^{\mathscr{M}^{\mathscr{S}^w}} B \quad \textit{iff} \quad \langle \sigma, z \rangle \overset{c}{\models} \overline{B}$$

for all $z \in \mathscr{S}^w$*.*

Proof. By induction on B. We examine just two cases.

- If B is atomic, the Lemma holds thanks to the definition of the interpretation function $\mathscr{I}$ of $\mathscr{M}^{\mathscr{S}^w}$.
- $B = |\, t : {}^{t_1}_{y_1} \ldots {}^{t_n}_{y_n} \,|\, C$, where $fv(B) = \{y, y_1, \ldots, y_n\}$.

If $\sigma \not\models_z^{\mathscr{M}} \ |\, t : {}^{t_1}_{y_1} \ldots\ldots {}^{t_n}_{y_n} \,|\, C$, then by Lemma 1

$$\sigma^{y \rhd \sigma(t), y_1 \rhd \sigma(t_1), \ldots\, y_n \rhd \sigma(t_n)} \not\models_z^{\mathscr{M}^{\mathscr{S}^w}} \ |\, y : y_1 \ldots y_n \,|\, C$$

where y is a variable different from $y_1 \ldots y_n$.

To simplify the notation, let $\pi = \sigma^{y \rhd \sigma(t), y_1 \rhd \sigma(t_1), \ldots\, y_n \rhd \sigma(t_n)}$, then

$$\pi \not\models_z^{\mathscr{M}^{\mathscr{S}^w}} \ |\, y : y_1 \ldots y_n |\, C.$$

By Definition 6 of satisfaction there is a v such that $\pi(y) \prec v$, a v-assignment τ such that $\tau \not\models_v^{\mathscr{M}^{\mathscr{S}^w}} C$, and moreover $\pi(y_i) \overset{\pi(y)}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$. By induction hypothesis $\langle \tau, v \rangle \overset{c}{\not\models} \overline{C}$. Since $\pi(y_i) \overset{\pi(y)}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$,

$$\langle \pi, z \rangle \overset{c}{\not\models} \overline{|\, y : y_1 \ldots y_n |\, C}$$

thanks to Definition 10. Consequently $\langle \sigma, z \rangle \overset{c}{\not\models} \overline{|\, t : {}^{t_1}_{y_1} \ldots {}^{t_n}_{y_n} \,|\, C}$ by Lemma 1 and the definition of π.

Conversely, if $\langle \sigma, z \rangle \overset{c}{\not\models} \overline{|\, t : {}^{t_1}_{y_1} \ldots\ldots {}^{t_n}_{y_n} |\, C}$, then by Definition 3,

$$\langle \sigma, z \rangle \overset{c}{\not\models} \overline{|\, y : y_1 \ldots\ldots y_n |\, C[t/y, t_1/y_1 \ldots t_n/y_n]}$$

where y is a variable different from $y_1 \dots y_n$, hence by Lemma 4

$$\langle \sigma^{y \rhd \sigma(t), y_1 \rhd \sigma(t_1) \dots\dots y_n \rhd \sigma(t_n)}, z \rangle \overset{c}{\not\models} \overline{|\, y : y_1 \dots y_n \,|\, C}.$$

To simplify the notation, let $\pi = \sigma^{y \rhd \sigma(t), y_1 \rhd \sigma(t_1), \dots\, y_n \rhd \sigma(t_n)}$, then

$$\langle \pi, z \rangle \overset{c}{\not\models} \overline{|\, y : y_1 \dots y_n \,|\, C}.$$

Since $\mathscr{S}^w$ is fully realized, there is a classical model v such that $\pi(y) \prec v$ and there is a v-assignment τ such that $\langle \tau, v \rangle \overset{c}{\models} \{\overline{B} : \langle \pi, z \rangle \overset{c}{\models} \overline{|\, y : y_1 \dots y_n \,|\, B}\} \cup \{\neg \overline{C}\}$ and moreover $\pi(y_i) \overset{\pi(y)}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$. Hence $\langle \tau, v \rangle \models_v \neg \overline{C}$, $\langle \tau, v \rangle \not\models_v \overline{C}$, therefore by induction hypothesis $\tau \not\models_v^{\mathscr{M}^{\mathscr{S}^w}} C$. Since $\pi(y_i) \overset{\pi(y)}{\rightarrowtail} \tau(y_i)$, $1 \leq i \leq n$, $\pi \not\models_z^{\mathscr{M}^{\mathscr{S}^w}}$ $|\, y : y_1 \dots y_n \,|\, C$ by Definition 10. Consequently $\sigma \not\models_z^{\mathscr{M}^{\mathscr{S}^w}} |\, t : {}^{t_1}_{y_1} \dots {}^{t_n}_{y_n} \,|\, C$.

□

Theorem 1 (Completeness). *If a wff $A \in \mathscr{L}$ is not a theorem of $Q.K_e$, then it is not valid on the class of transition epistemic models.*

2.6 Correspondence

- WHAT IS KNOWN IS TRUE
  (T_e) $\quad |\, x : x_1 \dots x_n \,|\, A \rightarrow A$

  It corresponds to the following conditions:

  - $a \in D_w$ only if $a \prec w$
  - For all $a, b \in D_w$, $b \overset{a}{\rightarrowtail} b$

  Let $a = \sigma(y)$ for some y and σ. If axiom T_e holds, then $w \overset{c}{\models} \{\overline{A} : \langle \sigma, w \rangle \overset{c}{\models} \overline{|\, y : |\, A}\}$, therefore $a \prec w$. Moreover, since $\langle \sigma, w \rangle \overset{c}{\models} \overline{|\, t : x_1 \dots x_n |\, A}$ only if $\langle \sigma, w \rangle \overset{c}{\models} \overline{A}$, then $\sigma(x_i) \overset{\sigma(t)}{\rightarrowtail} \sigma(x_i)$.

- POSITIVE INTROSPECTION
  (4_e) $\quad |\, x : x_1 \dots x_n \,|\, A \rightarrow |\, x : x, x_1 \dots x_n \,||\, x : x_1 \dots x_n \,|\, A$

  It corresponds to the following conditions:

  - Given $a \in D_w$ and $b \in D_v$, if $a \overset{a}{\rightarrowtail} b$ and $b \prec z$, then $a \prec z$
  - For all $a, b \in D_w$, $c, d \in D_v$, $e \in D_z$, if $a \overset{a}{\rightarrowtail} d$ and $b \overset{a}{\rightarrowtail} c$ and $c \overset{d}{\rightarrowtail} e$, then $a \overset{a}{\rightarrowtail} e$

- NEGATIVE INTROSPECTION
  (5_e) $\quad \neg |\, x : x_1 \dots x_n \,|\, A \rightarrow |\, x : x, x_1 \dots x_n \,| \neg \,|\, x : x_1 \dots x_n \,|\, A$

It corresponds to the following conditions:

- Given $a \in D_w$ and $b \in D_z$, if $a \prec v$, $a \prec z$ and $a \overset{a}{\rightarrowtail} b$, then $b \prec v$
- For all $a, d \in D_w$, $c, b \in D_v$, $e \in D_z$, if $d \overset{a}{\rightarrowtail} c$ and $d \overset{a}{\rightarrowtail} e$ and $a \overset{a}{\rightarrowtail} b$ then $c \overset{b}{\rightarrowtail} e$

As shown in Corsi (2010), some conditions of the counterpart relation correspond to quantified formulas.

- The Barcan formula: $\forall y|x : y, x_1, \ldots, x_n| A \to |x : x_1, \ldots, x_n| \forall y A$ corresponds to the property of the counterpart relation of being *surjective*.

  If Peter knows of all his friends that they are trustworthy, then Peter knows that all his friends are trustworthy.

  $$\forall y(\text{BEST FRIEND}(y, Peter) \to |Peter : y\,|\,\text{TRUSTWORTHY}(y)) \to |Peter :$$
  $$|\,\forall y(\text{BEST FRIEND}(y, Peter) \to \text{TRUSTWORTHY}(y))$$

  This sentence can be falsified if in worlds compatible with the epistemic state of Peter now, Peter has friends apart from the Peter-counterparts of his friends now.
- The Ghilardi formula : $\exists y\,|x : y, x_1, \ldots, x_n\,| A \to |x : x_1, \ldots, x_n\,| \exists y A$ corresponds to the property of the counterpart relation of being *everywhere defined.*

  If Peter knows of his best friend that he is trustworthy, then Peter knows that someone is trustworthy.

  $$\exists y(\text{BEST FRIEND}(y, Peter) \wedge |\,Peter : y\,|\,\text{TRUSTHWORTHY}(y)) \to |\,Peter :$$
  $$|\,\exists y\ \text{TRUSTWORTHY}(y)$$

  This sentence can be falsified if in worlds compatible with the epistemic state of Peter now, there are no Peter-counterparts of Peter's best friend now.
- The knowledge of identity: $x = y \to |z : x, y\,|\,(x = y)$ corresponds to the property of the counterpart relation of being *functional*.

  If Peter's best friend is Brian's father, then Peter knows of his best friend that he is Brian's father.

  $$P'bf = B'f \to |\,Peter : {}^{P'bf}_{x}, {}^{B'f}_{y}\,|\,(x = y)$$

  This sentence can be falsified if in worlds compatible with the epistemic state of Peter now, Peter-counterparts of Peter's best friend now are different from Peter-counterparts of Brian's father.
- The knowledge of diversity : $x \neq y \to |z : x, y\,|\,(x \neq y)$ corresponds to the property of the counterpart relation of not being *convergent*.

If Peter's best friend is not Brian's father, then Peter knows of his best friend that he is not Brian's father.

$$P'bf \neq B'f \rightarrow | p : {}^{P'bf}_{x}, {}^{B'f}_{y} | (x \neq y)$$

This sentence can be falsified since Peter-counterparts of Peter's best friend now can be the same as Peter-counterparts of Brian's father in all worlds compatible with the epistemic state of Peter now.

References

Braüner, T., & Ghilardi, S. (2007). First-order modal logic. In P. Blackburn, J. van Benthem, & F. Wolter (Eds.), *Handbook of modal logic* (pp. 549–620). Amsterdam/Boston: Elsevier.

Corsi, G. (2010). Necessary for. In C. Glymor, W. Wei, & D. Westerståhl (Eds.), *Logic methodology and philosophy of science: Proceedings of the thirteen international congress* (pp. 162–184). London: King's College Publications.

Corsi, G., & Tassi, G. (2010). *Agenti e transizioni*. Presented at Incontro di Logica in onore di Annalisa Marcja. Department of Mathematics, University of Florence, Florence, 6–7 May 2010.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge: MIT.

Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca: Cornell University Press.

Tassi, G. (forthcoming). Cone transition semantics.

Chapter 3
Explaining Capacities: Assessing the Explanatory Power of Models in the Cognitive Sciences

Raoul Gervais

3.1 Introduction

As Robert Cummins notes, capacities are an important type of explanandum addressed by psychologists (Cummins 2000). In fact, this does not only hold with respect to psychology, but seems to apply in equal measure to the other disciplines that fall under the label 'cognitive sciences'. All kind of cognitive capacities are in need of explanation, from face recognition to the ability to play chess; from motor skills to language acquisition. Now whereas most other types of explanandum (events, occurrences, states of affairs etc.) are, at least intuitively, explained by identifying their *causes*, capacities are typically explained in terms of a *model*.[1,2] To put the difference between these two explanations in pragmatic or erotetic terms, the former are answers to why-questions (Van Fraassen 1980), the latter to how-questions.[3]

[1] Throughout this paper, the term 'model' is used in a loose sense, to encompass any schema that mimics a certain pattern of behaviour that constitutes the explanandum. Of course, not all such models are scientifically or even philosophically interesting. However, in what follows, some specific *types* of models that *are* of interest will be considered in more detail.

[2] Of course, this is not to say that models cannot be causal in themselves, or that we cannot model causes. Rather, the difference is that the explanation of an event, occurrence or state of affairs typically refers to the cause of that event, occurrence or state of affairs, while the explanation of a capacity refers to a model, which may include *descriptions or simulations* of causes, but not the actual cause responsible for the capacity. In the former case, the explanans is located in reality, in the latter, it is a description or simulation of the cause, not the cause itself that does the explaining.

[3] This is not to say that one cannot ask how-questions about events, or why-questions about capacities (evolutionary explanations of biological traits provide examples of the latter strategy).

R. Gervais (✉)
Centre for Logic and Philosophy of Science, Ghent University (UGent),
Blandijnberg 2, 9000 Ghent, Belgium
e-mail: Raoul.Gervais@UGent.be

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic,
Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_3,
© Springer Science+Business Media Dordrecht 2014

In the cognitive sciences, two types of model are used to explain capacities: functional and mechanistic models. Functional models explain capacities by decomposing them into ever smaller sub-capacities or -routines, and then attempt to show how the overall capacity arises as a result of the way these sub-routines are organized (a useful metaphor here is that of the assembly line, where a complex task is divided into several simpler ones). These functional models can be highly abstract, putting more emphasis on the function to be performed than what actually performs it. Mechanistic models on the other hand, are less abstract. They too involve decomposing a capacity into a hierarchy of sub-functions or -capacities, but also include data on what type of entity is actually responsible for this or that (sub-)function (I will explain these two types of models in more detail in Sect. 3.2).

According to some authors, mechanistic models are superior to functional models precisely because they incorporate this additional information. While the latter are merely loose conjectures, the former are, at least in the ideal case, complete descriptions of the mechanism responsible for the explanandum. Indeed, Craver goes so far as to say that only to the degree it describes the actual entities by means of which a mechanism performs a capacity, can the model be said to *explain* that capacity (Craver 2006). Functional models can be useful for the purposes of prediction and control (they can successfully map the input-output patterns of the target system) but explanation requires something further. In the case of cognitive capacities, the model should at least be somewhat accurate ('plausible') from a neurophysiological point of view, if it is to explain those capacities. In short, it seems that on this view, *accuracy with regard to a mechanism's components is necessary for a model to have explanatory power*.

In this paper, I will argue against this view. Of course Craver is right in stating that in cases where we try to explain a capacity as it is realized in some particular system (which, of course, is what Craver and the mechanists in general are interested in), mere phenomenal models are not explanatory. However, this conclusion does not carry over to models in general: it is not correct to claim that descriptive accuracy is necessary in every context. The argument I present takes the form of a reductio: if it were necessary, this would exclude a whole range of models that are not only useful in the phenomenal sense (for the purposes of control or prediction), but intuitively also have explanatory power. These models are found in the context of *engineering*. A particularly promising way to account for these models is to employ the pragmatic perspective on explanation I hinted at above. We should realize that models need not be answers to how-questions relative to some set of systems S, but can also answer how-questions about capacities *as such*. The picture that emerges suggests that explaining capacities is a much more dynamic affair—consequently, a simple insistence on descriptive accuracy is too simplistic and does no justice to scientific practice.

The point is simply that in the cognitive sciences, explaining how a capacity comes about by constructing a model is simply a very prominent research strategy, which makes it philosophically interesting.

3.2 Functional Versus Mechanistic Explanations

Traditional functional explanations work by decomposition. They explain a capacity by breaking it down into sub-capacities or -functions, and then show how the overall capacity is a result of the organization of these sub-functions. Returning to the metaphor of the assembly line, let us consider a factory churning out radios. This factory effectively performs the function of taking parts as input and producing radios as output. This function can be explained by dividing the assembly process into several sub-routines carried out by workers standing alongside a conveyor belt, where each subsequent worker adds a specific component to the radio, until the finished product appears at the end of the belt, ready for transport. Once we know all the sub-routines that make up the assembly process, and understand the way they are organized (the order in which the parts are added) we can explain how the factory performs its function by means of a flow-chart or box diagram.

This explanatory strategy was widely used in the cognitive sciences, especially in the 1980s and 1990s. Cognitive capacities like memory storage, face recognition and numerical cognition were explained by construing models of how these capacities might be divided up into sub-functions. In psycholinguistics for example, a particularly influential functional model for the capacity of speech production was offered by Levelt (1989). Roughly, the process was divided into three steps: first, the person conceptualizes what he wants to say, second, he formulates this into language (this step is in turn divided into two sub-tasks, one of lexicalization, which produces the words needed, and one of syntactic planning, which provides order and grammatical structuring to these words) and finally, he engages in articulation (see Fig. 3.1).

Of course, this is a rough sketch of how the capacity might be realized, but it need not be wholly speculative. For example, the distinction between lexicalization and syntactic planning may be grounded in experimental evidence: some test subjects might be able to produce the right words, but fail to put them in the correct order. In general then, functional models need not be merely phenomenal (input-output mapping devices): with respect to the partitioning of a capacity into sub-routines,

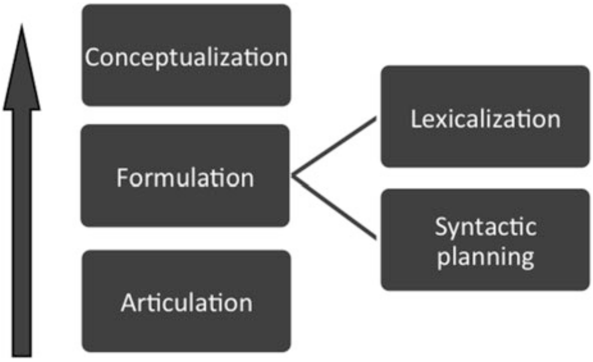


Fig. 3.1 An adaptation of Levelt's functional model of speech production. The sequence of operations is depicted vertically, with one step (formulation) being divided into two sub-routines (lexicalization and syntactic planning)

one can be detailed or abstract, and this partitioning might be supported by experimental evidence to a greater or lesser degree.

Yet however much informed a functional model like this might be, there is one issue with respect to which it remains silent: it has nothing to say about what actually performs all these sub-tasks. To put the point differently, it specifies functions, but not the realizers of these functions. In the example of the assembly line, imagine that in another factory, the different assembly tasks are realized by robots instead of workers. From a certain level of abstraction, the two factories are functionally equivalent, as they both perform the function of taking in parts as input and producing radios as output. More formally, if we want to explain a capacity C of a system S, we have to construct a functional model M which performs C, such that for each input, output and input-output relation in S there is a corresponding input, output and input-output relation in M.

In philosophy, this abstraction from what performs a function is often paired with the thesis of multiple realizability, and has been a key motivator to argue in favour of the autonomy of the special sciences (Fodor 1981). However, what was once hailed as an advantage is now increasingly criticised as a weakness. To be sure, functional models may succeed in correctly mapping the input-output relation of the target system, and for the purposes of control or prediction this may suffice, but does that make the model explanatory? Even though a particular partitioning of a function into subroutines is supported by evidence, if we want to understand how we, as humans, perform some kind of cognitive capacity, it seems imperative that we know something of the brain regions involved. Too often, the critics say, researchers are at a loss about what is really behind the boxes in their diagrams. For heuristic purposes, e.g. when we are just mapping out a certain capacity, this may be fine,[4] but if the original status of these boxes as mere placeholders is forgotten, they only serve to mask gaps in our understanding (hence the derogatory term ‘boxology’ that is sometimes applied to pure functional analysis).

In any case, a growing body of literature is devoted to an alternative approach to explaining cognitive capacities: mechanistic explanations. Like functional explanations, mechanistic explanations decompose the target capacity into several subcapacities. Unlike functional explanations however, mechanistic explanations also incorporate information about *what* performs a certain (sub-)function. They explain a capacity of a system by modelling the mechanism responsible for it: its operations, its entities or parts and the way the operation and parts are organized come into play.[5] Of course, this model need not be a complete description of the mechanism.

[4] See for example Machamer et al., who write that a mechanistic explanation typically starts by providing a mechanism sketch, which is “…an abstraction for which bottom out entities and activities cannot (yet) be supplied or which contains gaps in its stages. The productive continuity from one stage to the next has missing pieces, black boxes, which we do not yet know how to fill in” (Machamer et al. 2000, p. 18).

[5] Another way to put the difference is that mechanistic explanations, besides decomposition, also involve localization, where the latter notion is understood as the identification of activities with parts (Bechtel and Richardson 1993).

Ideally complete descriptions only serve as a regulative ideal: the degree of completeness required depends on our purposes at the time.

So far so good. But some authors do not stop at that. They believe that if our purposes are *explanatory*, then the model cannot afford to remain silent about the parts or entities of a mechanism:

> In order to explain a phenomenon, it is insufficient merely to characterize the phenomenon and to describe the behavior of some underlying mechanism. It is required, in addition, that the components described in the model should correspond to components in the mechanism.... (Craver 2006, p. 361)

Note that in this quote, Craver no longer talks about capacities as they are realized by humans, or indeed by any specific system: the claim he makes is about explaining 'a phenomenon', that is, about the explanatory power of models in general, not as they apply to any particular system. Thus Craver seems to endorse the following thesis:

(T) *For a model to have explanatory power, it is necessary that it corresponds to the target system, both with respect to its operations and the parts carrying out these operations.*

Now I agree that if we want to explain a capacity *as it is performed by some system or set of systems*, we must say something about the parts or components involved and, what is more, what we say should be correct. That is, the accuracy of the model should extend beyond the input-output relations to the actual mechanism itself. However, if from this concession **T** follows, we are in trouble, for not only do the traditional functional models described above not give accurate descriptions of a system's components, they typically remain silent about them altogether! According to **T** then, purely functional models are not explanatory. Nevertheless, from the 1970s onward, they have been used in cognitive psychology to explain all kind of capacities. With this discrepancy in mind, in Sect. 3.3, I will try to account for explanatory, yet purely functional models by considering some pragmatic aspects of explanation, while in Sect. 3.4, I will give an example of an explanatory context in which these aspects typically play a role.

3.3 Pragmatic Aspects of Explanation Considered

Although traditional functional models like the one sketched above are more abstract than mechanistic explanations in that they remain silent about a system's components, it would be wrong to infer from this that they have no explanatory power at all. To make this point, I will turn to a pragmatic account of explanation. The account I shall develop is pragmatic in the sense that it elaborates on van Fraassen's erotetic model of explanation.

According to van Fraassen, explanations are answers to why-questions (Van Fraassen 1980). However, as I have mentioned in the introduction, when dealing with *capacities*, it is often more appropriate to say that explanations are answers to how-questions. Fortunately, it has been argued persuasively that how-questions are valid explanation-seeking questions in their own right (Scriven 1962; Salmon 1989). Again, while answers to the former typically consist of identifying or referring to causes, the answers to the latter take the form of models. Recall how functional models work: if we want to explain a capacity C of a system S, we have to construct a functional model M which performs C, such that for each input, output and input-output relation in S there is a corresponding input, output and input-output relation in M. That is, if we want to answer a question like:

(1) How is C realized in S?

we should construct a model M that maps the input-output relations that make up C. Having done that, we can answer (1) by saying:

(2) C is realized in S the same way that C is realized in M.

Note that although it looks like (2) just restates the mystery, it does not, for we must remember that M is not a mechanism or system in nature, but a model that we have constructed ourselves, so that we know in detail how it realizes C. However, and this is where I agree with Craver, the question seems to ask something beyond input-output mapping. For a simple example, consider:

(3) How is the capacity to recognize faces realized in the human brain?

Now some face-recognition systems have been developed that perform this capacity very well, in that they are able, in experimental setups, to map the input-output relations of the brain (they are presented with examples of faces and non-faces and are able to tell the difference with more or less the same degree of accuracy as humans), but do so in a fundamentally different way. Up until recently for example, they could only use two-dimensional geometrical data. Of course we do not want to count:

(4) The capacity to recognize faces is realized in the human brain by applying algorithms to exclusively 2-D geometrical data.

as an answer to (3). As we know ourselves to see, e.g., chins and noses as protrusions, (4) is clearly inaccurate. Beyond this appeal to ‘first person knowledge’ however, there is also some ‘harder’ evidence. For example, 2-D face systems notoriously suffer from what is known as the ‘lighting problem’: their ability to recognize faces deteriorates significantly when the strength of the light coming from the image they are presented with is varied, while humans tend to retain their abilities in such circumstances. No matter how perfectly such systems may mimic our performance in this task, we have to concede that, being 2-D, they are not explanatory models for face recognition as it is performed by humans.

Granted then, a model may to a certain extent map the human input-output relation for a capacity, without being explanatory with respect to the human realization of that capacity. However, **T** makes a stronger claim than that. Craver went beyond models for capacities as they are performed by humans or systems, to claim that *any* model that does not offer an adequate description of a system’s components has no explanatory power. But do models always have to be models of a capacity as it is performed in a specific (set of) system(s)? The erotetic approach we have explored so far says that if a capacity is the explanandum, the explanans can be viewed as an answer to a how-question. There is nothing to restrict this type of question to include only capacities *as they are realized in some system*, we can also ask how-questions about capacities *as such*, that is, without any particular descriptive or correspondence constraints. Instead of (1), we might ask:

(5) How is C (as such) possible?[6]

The point here is not that researchers will actually be interested in how capacities could be realized without *any* constraints: capacities are of course always realized in some system. Rather, the point is that one can have legitimate motives in placing *as little constraints on the system as possible*. In Sect. 3.4, I will consider one context in which this strategy is commonplace, namely the context of engineering. For now, note that at least in psychology and the cognitive sciences, asking explanatory questions about capacities as such forms an important part of scientific practice, if only as a preliminary strategy (that is, preliminary to the business of answering the question how the capacity is realized in some particular system). In fact, this was already noted by Dennett back in 1978:

> Faced with the practical impossibility of answering the empirical questions of psychology by brute inspection (how *in fact* does the nervous system accomplish X or Y or Z), psychologists ask themselves an easier preliminary question: How could any system (...) possibly accomplish X? This question is easier because it is 'less empirical'; it is an engineering question, a quest for a solution (*any* solution) rather than a discovery. (...) Seeking an answer to such a question can sometimes lead to the discovery of general constraints on all solutions (...), and therein lies the value of this style of aprioristic theorizing. (...). For instance, one can ask how any neuronal network with such-and-such physical features could possibly accomplish human color discriminations (...). Or, one can ask, with Kant, how anything at all could possibly experience or know anything at all. Pure epistemology, thus viewed (...) is simply the limiting case of the psychologist's quest. (Dennett 1978, pp. 110–111)

Thus viewed, the 'Kantian' question (How is X possible at all?) can be interpreted as constituting the extreme end of a continuum, while enquiries about how a particular system performs that function occupies the opposite end (Fig. 3.2).

As Dennett notes, it is possible to begin with more general questions, discovering constraints having to do more with C itself, and work your way to a particular realization of C in S. However, explanation can also work in the opposite direction.

[6] Note that this question does not fall into the category of Craver's how-possibly questions (Craver 2006). For Craver, how-possibly questions are loose inquiries that are made in the early stages of an investigation, in which a lot of data is still missing: they are attempts to put some initial constraints on the explanandum, prior to constructing a more informed (how-plausibly), and ultimately ideally complete description (how-actually). Nevertheless, how-possibly questions in Craver's sense are still asked with respect to a capacity as it is performed by some system. The question under consideration differs because it is asked about a capacity *as such*, regardless of any particular realization.

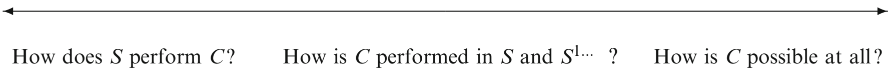

How does S perform C?
How is C performed in S and $S^{1\cdots}$?
How is C possible at all?

Fig. 3.2 Different levels of abstraction at which one might seek to explain a capacity

As one moves to the right of the spectrum, the number of constraints will decrease. This means that somewhere along the line you get to the point where C is described in such a general way that it applies to more than one system. In other words, the scope increases. Examples of this can be found in medicine. If an impaired capacity in a brain damaged patient has somehow been restored by the brain, we might be interested to know just exactly how that capacity is carried out in this damaged brain. In circumstances like these, we are actually looking to move toward the right end of the spectrum. Of course, detail matters: as soon as we reach the point where all the relevant systems fall under the scope of that capacity, we stop. In the example, as soon as we have described the capacity in such general terms that it applies both to healthy patients and the brain damaged patient, we stop jettisoning constraints. This stopping has to do with our methodological interests: it is simply the act of eliminating variables.[7]

In Sect. 3.4, I will give a more detailed example of this explanatory strategy. For now, the point to note here is that abstraction is a matter of degree. How many constraints one places on the system responsible for a certain capacity will be decided by pragmatic issues. This however, seems at odds with **T**, which endorses descriptive accuracy about implementational details as necessary for a model to have explanatory power. Of course, this is particularly striking for questions located near the right end of the spectrum: surely, one cannot expect a model answering (5) to excel in descriptive accuracy, for there is no mechanism specified to describe. In fact, *any* model of *any* system that realizes C is a valid answer. Again, scientists are rarely (if ever) interested in capacities under no constraint whatsoever. Nevertheless, the continuum sketched above suggests a more dynamic and more tolerant picture of model-explanation; a picture which **T**, with its simple assertion that descriptive accuracy about entities and parts is necessary for a model to have explanatory power, is too rigid to encompass.

3.4 Explaining Capacities in Engineering Contexts

Explanation-seeking how-questions about capacities as such are often asked in cases where the research is driven by engineering interests. In the case of the cognitive sciences for example, type (5) questions might arise in artificial intelligence. Let us consider one specific example of a cognitive capacity: exact calculation.

[7] Also, think of animal testing: here we continue to drop constraints until the capacity is described in such a way as to apply across species. Again, S can be any system, natural or artificial.

Humans are endowed with the capacity to perform exact calculations accurately, up to a certain level of complexity. If we ask how we perform this capacity, the model that answers this question indeed derives its explanatory power from (among other things) its neurophysiological accuracy. That is, if we want to answer:

(6) How is the capacity to perform exact calculations realized in humans?

the model that we use to answer (6) has to reproduce the capacity under a number of constraints. For example, some artificial computing devices might make poor models, as they are disanalogous to human brains in important respects: they might be neurophysiologically implausible, or they might fail to reproduce the capacity to perform exact calculations (e.g., they might be less exact, or they might take far longer to solve arithmetic problems).

However, although these respects are important to contexts like the one referred to in question (6), there are other contexts in which they are less important, or even irrelevant, and these other contexts might still have to do with explaining the capacity. In other words, descriptive accuracy or correspondence is not the only explanatory context in which we could be interested in the capacity: there are other reasons we might want to explain the capacity to perform exact calculations. Suppose an engineer wants to construct a desk calculator. Now of course, his goal is not to construct a model of how humans perform complex calculations: after all, he is designing a tool that, hopefully, surpasses our own ability. In fact, he seeks to *duplicate* the capacity. Motivated by this interest of duplication, he might ask:

(7) How is exact calculation as such possible?

However, this is somewhat artificial. In fact, when constructing a desk calculator, there are all kinds of constraints he needs to take into account.[8] The point is that these constraints are different from the ones applying to exact calculations as it is performed by humans. Thus, a sensible strategy would be to put fewer constraints on the capacity, until the scope is broad enough to apply to both humans and certain artificial devices. In terms of the continuum sketched above, we stop somewhere in the middle, at the point where the scope is just broad enough to encompass both the human realization of the capacity and an artificial one. To put it in other terms, we stop where the forces pulling in opposite directions, namely level of detail (to the left) and duplication (to the right), balance out for the task at hand.

But that is not all. In engineering contexts, it is not uncommon to jettison the requirement of descriptive accuracy completely. To appreciate this, let us continue to pursue the example of the engineer trying to construct his desk calculator. Now there are a number of models that can perform exact calculations. For reasons of clarity, let us consider classic computationalism and connectionism. The symbolic architecture of classic computationalism, where symbols are manipulated according

[8] Examples of such constraints are: the materials available, convenience of use and time considerations (we want the calculator to perform calculations rapidly—within a timeframe that is of use to us, that is).

to a pre-programmed set of rules, is very good at performing very complex calculations with great accuracy, far surpassing that of any human. On the other hand, as a model of the mind, computationalism is outdated. The serial nature of its operations and its consequent brittleness does not compare to the robustness of our brains. Connectionism on the other hand, resembles our brains more closely. In fact, in the original debate between computationalism and connectionism as candidate models for the mind, the latter's neural plausibility (in the form of distribution of activity over a network of nodes, graceful degradation, its ability to recognize patterns etc.) counted as an important point in its favour (McClelland and Rumelhart 1986).[9] However, despite all these advantages, they perform poorly when it comes to exact calculations. In fact, connectionist networks have been ridiculed for answering a question like "What is two plus two?", after much crunching, with "About four".

Clearly, exactness is a virtue when it comes to desk calculators. In fact, when engineering interests drive model construction, *performance trumps accuracy*. Duplication therefore, is only a subsidiary goal: it is really the desire to make a system that outperforms humans that motivates the engineer, and the model he finally constructs will reflect this. Of this model, that is of the flow chart representing how the calculator performs the exact calculations, we can say three things. First, with regard to how humans perform exact calculations, it is an inaccurate model and fails to explain it. Second, with regard to how the calculator performs it, it is an ideally complete description and explains it, but that is hardly surprising, since it is the very blueprint the engineer used to make the calculator in the first place. Third, with regard to the capacity to perform *exact calculations* as such, it explains how that capacity *can be* performed. When the engineer asked (7) and started decomposing exact calculation down into sub-routines, he was looking for an explanation, only not with neurophysiological accuracy on his mind, but performance.

Yet there are other interests besides duplication or performance that might prompt the search for an explanation of such capacities. Another interest is *unification*. Once an artificial system has been designed and constructed, then to anyone besides the engineers involved in this process of designing and construction, the explanatory question might arise as to what these artificial systems have in common with, e.g., natural systems. Again, the term 'system' has been chosen to reflect the fact that we might not only be interested in a capacity as performed by humans (or natural systems in general), but also by artificial ones. Thus, one might ask the following question:

(8) How is the capacity to perform exact calculations performed in this desk calculator and in humans?

This question is situated somewhere in the middle of the continuum presented in Sect. 3.3. In effect, what we are asking for here is what two realizations of the capacity of exact calculations have in common with each other. These comparative

[9] As the debate currently stands though, connectionist networks are considered to be highly idealized models too—but still more plausible than classic computationalist architectures.

question-types are often motivated by unification: in revealing features that are common to the operations of both types of systems, an answer to (8) brings together information from multiple and diverse sources. And of course, an answer to comparative question-types like (8) will typically take the form of a model—precisely the kind of functional model introduced in Sect. 3.2. In the case of question (8), this is especially clear, since any similarity between humans performing complex calculations and desk calculators exercising the same capacity will not be found in the entities, but will be confined solely to the domain of the operations. Yet, despite its abstract nature, and *pace* **T**, such a model would clearly be of explanatory value to those who are interested in the similarities between human and artificial performances of exact calculation.

Again, all this does not tarnish the explanatory importance of mechanistic models when it comes to explaining capacities as they are realized in particular systems. Of course we need the models of, e.g., biological functions to be accurate, and not only phenomenally adequate. It might even follow that for particular systems, this accuracy is necessary for a model to have any explanatory power regarding that capacity. What does not follow however, is that phenomenal and functional models have no explanatory power in *any* context. Reiterating Dennett's point, asking about capacities under fewer constraints can be a valuable research strategy. Ultimately, how many constraints one takes into account is decided by one's interests: in the case of performance, an interest typical of engineering contexts, these constraints will surely be determined by practical considerations, rather than empirical adequacy. Nevertheless, this does not undermine the explanatory power of answers to such questions. Hence, it seems that Craver's thesis **T** is false as it stands. However, although strictly speaking correct, this conclusion should not be the main point to take away from this discussion, if only for the fact that Craver and the mechanists have a very different context in mind from some of the ones considered in this paper. Of greater importance is the observation, borne out by the continuum sketched in Sect. 3.3 and illustrated in this section, that the business of explaining capacities by constructing models is far more diverse and dynamic than Craver suggests. This more constructive conclusion might serve as a starting point to reformulate **T** in a way that either restricts its scope, so that it applies only to those contexts which Craver had in mind, or to drop the requirement of descriptive accuracy, so that it does justice to the practice of explaining capacities by constructing models.

3.5 Some Concluding Remarks

Two final remarks are in order. First, although distinct, engineering and accuracy interests are often present at the same time and can even be complementary. This is especially the case when a model has to be constructed of a capacity at which, unlike exact calculations, humans are particularly good. Face recognition for example, is a capacity in which we excel, and many of the early artificial systems badly underperformed compared to us, being sensitive to all kind of distortions (we already encountered the lighting problem, faces presented at angles is another one) that human test persons just see right through. In such cases of course, an engineer wanting to design such an artificial system has everything to gain by first asking how the capacity is realized in us. The point is though, that even here, accuracy is only a sub-goal. As soon as artificial systems are starting to equal or outperform us, engineers will drop accuracy as a goal, as it no longer serves the greater goal of performance.[10]

Finally, one may wonder whether the capacities targeted by functional explanations in engineering contexts, such as the one described in Sect. 3.4, are still properly called *cognitive* capacities. Can we still talk of subtraction as a cognitive capacity when it is performed by a humble desk calculator instead of a person? Here, one might point out that the engineering sciences (artificial intelligence in particular) have a history of fruitful interaction with the cognitive sciences. Artificial systems can help us understand our own capacities, while knowledge of these may in turn lead engineers to improve the performance of these systems. After all, the point made in this article is that accuracy and explanatory power can, and in some cases do, operate separately from each other, not that they always do so.

Acknowledgements The research for this paper was supported by the Research Fund Flanders (FWO) through project nr. G.0031.09.

References

Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.

Cummins, R. (2000). "How does it work?" versus "What are the laws?" Two conceptions of psychological explanations. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117–145). Cambridge: MIT.

Dennett, D. (1978). Artificial intelligence as philosophy and as psychology. In D. Dennett (Ed.), *Brainstorms* (Philosophical essays on mind and psychology, pp. 109–126). Montgomery: Bradford Books.

Fodor, J. (1981). Special sciences. In *Representations: Philosophical essays on the foundations of cognitive science* (pp. 127–145). Harvester: Hassocks.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

McClelland, J., & Rumelhart, D. (1986). *Parallel distributed processing: Explorations in the micro-structure of cognition* (Vol. 2). Cambridge: MIT.

[10] And in fact, with the example of face recognition systems we considered earlier, this is beginning to happen right now; see the results from the 2006 Face Recognition Vendor Test (available for download at: http://www.frvt.org/).

Salmon, W. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Minnesota studies in philosophy of science* (Vol. VIII, pp. 3–10). Minneapolis: University of Minnesota Press.

Scriven, M. (1962). Explanation, predictions and laws. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space and time* (Minnesota studies in the philosophy of science, Vol. III, pp. 170–229). Minneapolis: University of Minnesota Press.

Van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press.

Chapter 4
Data-Driven Induction in Scientific Discovery: A Critical Assessment Based on Kepler's Discoveries

Albrecht Heeffer

4.1 Introduction

Any rational approach to scientific discovery has to deal with ampliative reasoning. One of the most important ampliative mechanisms to expand our knowledge about nature is inductive reasoning. Since Hempel (1945), however, many philosophical problems with induction have been formulated. Some have dismissed the role of induction in science altogether: "A theory of induction is superfluous. It has no function in a logic of science" (Popper 1959). Starting with *Experience and Prediction* by Reichenbach (1938), the distinction between the context of discovery and the context of justification has provided a subterfuge to deal with the latter only and to move the difficult problem of induction outside the realm of scientific explanation. From the late 1970s onwards, cognitive scientists and researchers within the domain of artificial intelligence (AI) started formulating models for dealing with the context of discovery. Based on the seminal work by Newell and Simon (1972), creativity, such as scientific discovery, was approached as a rational process—using the same kind of mechanisms as one does in puzzle solving. Most of the research on scientific discovery focused on one particular kind of induction, named data-driven induction. The central tenet of data-driven induction in science is that scientists discover quantitative laws of nature by a process of inductive generalization from observational data. Using Simon's model of a goal-directed state-space search, the problem of discovery in science was reduced to finding the right heuristics in a general, generic and domain-independent model of problem solving. Thomas Nickles formulated this approach pertinently as "the neo-enlightment counterpart of universal Reason, a faculty that could in principle solve any (solvable) problem in any domain" (Nickles 1994). Several systems have

A. Heeffer (✉)
Center for the History of Science, Ghent University (UGent), Ghent, Belgium
e-mail: Albrecht.Heeffer@UGent.be

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_4,
© Springer Science+Business Media Dordrecht 2014

been built to model such processes. The most influential one has been BACON, authored by Pat Langley, Gary Bradshaw, Jan Zytkow and Herbert Simon. This research program spanned a period of 1978–1990, involving research groups from several universities and resulting in many publications. The most representative overview is a book titled *Scientific Discovery. Computational Explorations of the Creative Processes* (Langley et al. 1987). In the rest of this paper I will mostly refer to this publication although my arguments also apply to other KDD approaches which endorse the model of data-driven induction.

4.2 Aims and Motivation

My aim is to show that the approach taken by the BACON team is not an adequate one for explaining and modelling scientific discovery *in general*. This is exactly the claim made by the BACON team: not only is their book "concerned more with describing and explaining scientific discovery than with providing a normative theory of the process", they also call their model constructive, "it exhibits a set of processes that, when executed, actually make scientific discoveries" (Langley et al. 1987, p. 7). I will present here a critical assessment of these two main presuppositions, (1) their model describes and explains historical cases of scientific discovery and (2) the BACON program is sufficient to make scientific discoveries. I will do so, not so much from the viewpoint of philosophy of science but more from a contextual historical perspective. We will focus on two "successes" of the BACON program: the rediscovery of Kepler's third law and the sine law of refraction.

Before discussing the main criticisms against these claims, I have to make a disclaimer. The arguments discussed below deal with one particular form of inductive reasoning: data-driven induction. In no way is it implied that inductive reasoning has no function in scientific discovery. Some arguments will focus on the use of observational data by the BACON program. Neither is suggested here that observational data does not play a significant role in the formulation of general quantitative laws. The possibility of inducing general laws from empirical data is not questioned either. Even the claim that BACON is able to find some laws from data fed to the program is not at issue here. Furthermore, I fully acknowledge the importance of the BACON program towards a rational explanation of the process of scientific discovery and endorse the main starting point that creativity and discovery in science can and must be explained as problem-solving processes.

One more question has to be answered. What is the relevance today of formulating methodological objections against the BACON program? While a simplified reduction of scientific discovery as data-driven induction has been mostly dismissed in currently prevailing philosophy of science, the idea is still very prominent in recent artificial intelligence research. Two new technologies have given a new impetus to data-driven induction: data mining since the early 1990s and knowledge discovery from data (KDD). In 1995 a Special Interest Group on Knowledge discovery and Data Mining (SIGKDD) was created. In 1999 this group founded a new journal on this topic. Several new conferences were created to communicate new research results within this domain: Research Issues on Data Mining and Knowledge Discovery (DMKD), the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), and The annual Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKKD). The Langley book is often quoted within these scientific communities. ISI Web of Science lists 290 citations, Google scholar 841 articles. These provide us reasons enough to reassess the philosophical and methodological foundations of the program.

4.3 The Descriptive and Explanatory Power of the BACON Program

While early research in computer chess was motivated by the study of human problem solving (as with Newell and Simon 1972) the domain rapidly developed into a race to beat the best humans in chess. It became clear that the calculating power of computers received more benefit from the model of state-space search than humans did. In computer chess there is a tradeoff between the amount of domain-dependent knowledge you add and the search depth. Chess knowledge in computer programs is kept to a minimum in order to gain some more moves in search depth. Current programs reach up to 20 ply (half moves) in middle-game positions, which is sufficient to beat the world champion in chess. Place this against the famous quote by the Cuban chess grandmaster Capablanca when asked how many moves he can think ahead: “Only one move ahead—the right one”, he replied. The moral of this story is that computer chess works very different from the cognitive processes involved in human chess. Therefore, chess programs do not explain much about the cognitive capacities of chess masters. In other words: computer chess is no model for human problem solving in chess.

Let us now move to the question if BACON provides an adequate model for discovery in science. The first discussion point is whether the BACON researchers intended to produce a model of scientific discovery. The answer is affirmative. Not only is their intention to provide a model for scientific discovery in general, they also claim that it explains the historical cases covered in the book. It is not without some pride that Langley et al. conclude their presentation of BACON with the words “We would like to imagine that the great discoverers, the scientists whose behavior we are trying to understand, would be pleased with this interpretation of their activity” (Langley et al. 1987, p. 340). Unexpectedly, they have also claims about the amount of time scientists spend on the activities of data-driven induction, modeled by the BACON program: “Although we have no quantitative data on which to build an estimate, it is reasonable to suppose that the induction of laws from data might typically occupy *only* from 1 to 10 % of a scientist’s time” (Langley et al. 1987, p. 112, emphasis mine). Since quantitative laws of nature are discovered only so often, even the lower estimate of 1 % seems to be a very high occurrence of data-driven induction by scientists.

In Sect. 4.4 I will assess to what degree two of the successful discoveries by BACON contribute to an understanding of the actual historical discoveries: the sine law and Kepler's third law. There are two reasons for selecting these two cases. The first is that both are related to Johannes Kepler. Kepler is a gratifying subject for any historical study. In contrast with his contemporaries as Descartes he abundantly documented his line of reasoning including his failures as well as his successes. While Kepler did not succeed in formulating the sine law he came very close to it and we may suspect that others came to the discovery following the same lines of investigation. The second reason is given by the BACON team itself. According to Langley et al. (1987, p. 224) the sine law is one of the "three instances where, in the actual history of the matter, the data—essentially the same data that were available to BACON—were interpreted erroneously before the "correct" law was discovered". The other two are Kepler's third law and Boerhaave's law. The fact that scientists fail to arrive at a quantitative law when the correct data is available to them appears to be an interesting phenomenon. If only data-driven induction is at play, such failures should not happen. Langley et al. explain it as follows:

> In all of these cases, the error arose from accepting "loose" fits of a law to data, and the later, correct formulation provided a law that fit the data much more closely. If we wished to simulate this phenomenon with BACON, we would only have to set the error allowance generously at the outset, then set stricter limits after an initial law had been found. (Langley et al. 1987, p. 224)

Such a claim implies that BACON—and the model of data-driven induction—not only can explain the successes of scientific discovery but the failures as well. This will be taken into consideration in a later section.

4.4 BACON's "Discoveries"

The sine law of refraction has been "rediscovered" by two versions of the BACON program. BACON.4 was given as input a table of 9 data lines which contains the sines of angles of incidence and refraction for three angles and combinations with three media (see Table 4.1).

Table 4.1 Input data for the BACON program

Medium1	Medium2	sin A_1	sin A_2	sin A_1/ sin A_2
Vacuum	Vacuum	0.500	0.500	1.00
Vacuum	Vacuum	0.707	0.707	1.00
Vacuum	Vacuum	0.866	0.866	1.00
Vacuum	Water	0.500	0.665	0.75
Vacuum	Water	0.707	0.940	0.75
Vacuum	Water	0.866	1.152	0.75
Vacuum	Oil	0.500	0.735	0.68
Vacuum	Oil	0.707	1.039	0.68
Vacuum	Oil	0.866	1.273	0.68

Table 4.2 Input data of BACON.1 for discovering Kepler's third law

Planet	Distance (D)	Period (P)	D/P	D^2/P	D^3/P^2
A	1.0	1.0	0.500	1.0	1.00
B	4.0	8.0	0.707	2.0	1.00
C	9.0	27.0	0.866	3.0	1.00

The program "discovers" that the ratio between the two sines is invariant for a given combination of media. Against the possible objection—which I will raise anyway in Sect. 4.9—that giving the sines as input to the program is giving away the discovery, Langley et al. write:

> BACON.4 does not have heuristics for considering trigonometric functions of variables directly. Thus, in the run described here we simply told the system to examine the sines. In the following chapter we will see how BACON can actually arrive at the sine term on its own in a rather subtle manner. (Langley et al. 1987, p. 142, note 5)

This subtle manner appears to be that instead of the sines of the angles BACON.5 is given the lengths of the hypotenuse and the opposite side of a rectangular triangle. But as we all know from secondary school, the ratio of these two parameters amounts to exactly the same.

The other discovery is Kepler's third law, which describes a relation between the average distance of a planet from the Sun and the time of its orbit. This was discovered by Kepler shortly before finishing his *Harmonices Mundi*, published in 1619 (see Kepler 1858–1871 for the complete works). Newton later proposed a more general form of the law applying to any two objects orbiting around a common center of mass. Langley et al. point out that in both cases the discovery involves a single step of data-driven induction: "The important point to notice here is that, in discovering the relation between acceleration and distance, the only step of induction is the inference of Kepler's third law from the observation—a data-driven induction" (Langley et al. 1987, p. 56). The "rediscovery" of the law was taken as a task for the first version of their program BACON.1. The input data consisted of three lines with the two relevant parameters (see Table 4.2). The program "discovers" that the expression D^3/P^2 is invariant for the input data.

The authors admit that the input data "were contrived to fit Kepler's law exactly" (Langley et al. 1987, p. 69). The problem of early modern astronomy was that observational data did provide only approximate measures of distance and period. If BACON has to deal with real-world data, this increases the complexity dramatically. The BACON.3 program seems to work with real-world data. The "discovery" of Kepler's third law by BACON.3 is achieved from the distance relative to the earth and a slope (see Table 4.3).

A slope together with an intercept defines a linear relation between the given parameters a and b as in $as + i = b$. A separate table lists the angle between the sun and the planet seen from a fixed point and the distance determines an invariant slope. As can be seen from the third line in Table 4.3, the result for D^3/s^2 should be 0.986 but it still fits to 0.971. This is due to the modifiable noise margin allowed when looking for invariants.

Table 4.3 Input data of BACON.3 for discovering Kepler's third law

Planet	Distance (D)	Slope (s)	Ds	D^2s	D^3s^2
Mercury	0.387	4.091	1.584	0.613	0.971
Venus	0.724	1.600	1.158	0.839	0.971
Earth	1.000	0.986	0.986	0.986	0.971
Mars	1.524	0.524	0.798	1.217	0.971
Jupiter	5.199	0.083	0.432	2.247	0.971
Saturn	9.539	0.033	0.319	3.044	0.971

4.5 The Historical Data

In order to assess the descriptive and explanatory value of BACON for the two selected historical cases, it is necessary to verify that the data given to BACON actually corresponds with what was known to Kepler and his contemporaries. As cited before, Langley et al. claims the sine law is one of the "instances where, in the actual history of the matter", BACON used the same data that was available to the discoverers of the law. However this is actually not the case. For many centuries the refraction tables from Ptolemy's *Optics* were the main authority.

Around c.160–168 AD, Ptolemy collected data by setting up carefully contrived experiments, using a bronze instrument. He reached quite accurate measures for the angles of incidence and corresponding angle of refraction between three types of media: air/water, air/glass and water/glass (Smith 1996; *Optics*, V, 7–11, 20–21, 31–35; see Table 4.4).

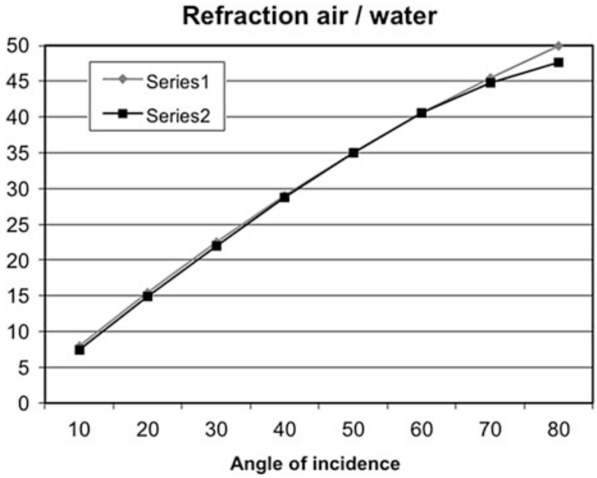
How do these tables compare with the data used by the BACON programs? For a start, notice that Table 4.1 in the Langley book lists the sines for incidence and refraction between vacuum and water. Vacuum did not exist as a concept at the time of Ptolemy and started to evolve only after the experiments by Robert Boyle and Robert Hooke after the 1650s. It was unfeasible to set up experiments with a vacuum in the first half of the seventeenth century. Apart from this historical blunder, also remark that the sines for the denser media, water or oil, are higher than those for vacuum. As was known already by Ptolemy, the angle of refraction is smaller than the angle of incidence when moving from a rare to a dense medium. The refractive index of water is 1.33, defined as the sine of the angle of incidence in vacuum, divided by the sine of the angle of refraction in water. From Table 4.1 we read that the refractive index for water is 0.75. BACON seems to discover the inverted sine law! A conclusion that cannot be avoided is that neither Table 4.1 nor the data used by BACON.5 correspond in any way with the historical data.

The story gets even worse. If BACON would have used the historical data for data-driven induction, it would have arrived at a different law! To demonstrate this, let us look again at Table 4.4. There is something peculiar about the angles. The angles of refraction all fit within a full or half degree. This could be explained as rounding errors caused by the measuring equipment. However, I tried to imagine what one would arrive at if one tried to induce a general quantitative law from the Ptolemy's observational data. For that purpose I used the polynomial curve fit

Table 4.4 The refraction tables from Ptolemy's *Optics*

Air/water		Air/glass		Water/glass	
Incidence	Refraction	Incidence	Refraction	Incidence	Refraction
10	8.0	10	7.0	10	9.5
20	15.5	20	13.5	20	18.5
30	22.5	30	19.5	30	27.0
40	29.0	40	25.0	40	35.0
50	35.0	50	30.0	50	42.5
60	40.5	60	34.5	60	49.5
70	45.5	70	38.5	70	56.0
80	50.0	80	42.0	80	62.0

Fig. 4.1 Comparing the quadratic relation against the sine law



function of the symbolic computation program *Mathematica*. To my great surprise Ptolemy's data perfectly match with three simple quadratic relations between angles of incidence (i) and angles of refraction (r) (Fig. 4.1):

$$r = \frac{330i - i^2}{400} \quad \text{for air/water,}$$

$$r = \frac{290i - i^2}{400} \quad \text{for air/glass,}$$

$$r = \frac{390i - i^2}{400} \quad \text{for water/glass.}$$

BACON.3 looks for a linear relation between the input data and would not find this relation. However, Langley et al. mention that an early version of BACON.5 used a polynomial curve fit module (Langley et al. 1987, p. 171, note 1). At the 1981 IJCAI conference Langley, Bradshaw and Simon reported on the discoveries of this

Table 4.5 The "computation" of Ptolemy's refraction tables

Incidence	Refraction	Arithmetical progression with decreasing increments
10	8.0	
20	15.5	8.0 + (8.0 − 0.5)
30	22.5	15.5 + (7.5 − 0.5)
40	29.0	22.5 + (7.0 − 0.5)
50	35.0	29.0 + (6.5 − 0.5)
60	40.5	35.0 + (6.0 − 0.5)
70	45.5	40.5 + (5.5 − 0.5)
80	50.0	45.5 + (5.0 − 0.5)

program, including the sine law of refraction (Langley et al. 1981). As is shown here, the early version of BACON.5 should have found the quadratic relations.

The graph in Fig. 4.1, the data from Ptolemy (series 1) fits the sine law (series 2) very closely for smaller angles of incidence. The difference becomes apparent only for degrees of incidence higher than 70°. This raises some questions. Did anyone notice this before? Apparently yes. Gilberto Govi, who published a Latin edition of Ptolemy's *Optics*, was the first to point out the existence of a quadratic relation: $r = ai - bi^2$ (Govi 1885, pp. XXII). Smith (1996) finds the relation

$$r = R - \frac{(n^2 d_2 - n d_2)}{2}.$$

A second question is why Ptolemy's data shows this remarkable regularity. Albert Lejeune who published a French translation of the *Optics* suspects that the figures may have been "adjusted" by previous translators ("Il n'est pas absolument exclu que les tables aient été régularisées par un interpolateur grec ou arabe", Lejeune 1956). However, a satisfactory explanation is found in Otto Neugebauer's treatment of Babylonian astronomy (Neugebauer 1957, p. 111). Since 500 BC it was the practice to constantly diminish increments for the construction of astronomical tables. The computation of the ephemerides was achieved by such method, which Neugebauer calls a zigzag function.

We can indeed reconstruct Ptolemy's table by means of such zigzag function, as shown in Table 4.5. Such customs raise serious objections about a simple view of scientific discovery as data-driven induction from observational data. The data used for astronomy and optics in antiquity do not consist of raw observational data but are already shaped by geometrical or arithmetical models, believed to govern the organisation of nature and the universe. Such an epistemological view is clearly present in Kepler. Kepler did not use Ptolemy's refraction tables but a slightly different one from Witelo. Witelo's *Perspectiva* is a seminal work on optics from the thirteenth century. It was known to Kepler through the Risner edition of 1572 (for a modern edition see Unguru 1977). Witelo's refraction tables differ from Ptolemy's only for the first line and only for water to air. The refracted rays appear under an angle of 7°55′ in the manuscripts and 7°45′ in the printed edition.

Table 4.6 Data for the mean radius compared

Planet	BACON	Tycho	Kepler 1618
Mercury	0.387	0.387	0.388
Venus	0.724	0.723	0.724
Earth	1.000	1.000	1.000
Mars	1.524	1.524	1.524
Jupiter	5.199	5.202	5.200
Saturn	9.539	9.539	9.510

Kepler published in 1604 an extensive work on optics as a critique on Witelo, titled *Ad Vitellionem paralipomena, quibus astronomiae pars optica traditur* (for a modern edition and English translation see Donahue 2000). Chapter 4 is completely devoted to refraction. Proposition 8 contains a passage which reveals Kepler's attitude towards observational data and Witelo's refraction tables in particular. Kepler decomposes the angle of incidence and refraction into two components and then uses an algebraic relation to calculate the composite again. He compares his results against Witelo's table:

> This tiny discrepancy should not move you; believe me: below such a degree of precision, experience does not go in this not very well fitted business. You see that there is a large inequality in the differences of my figures and Witelo's. But my refractions progress from uniformity and in order. Therefore, the fault lies in Witelo's refractions. You will believe this all the more, if you look to the increments of the increments in Witelo. For they increase through 30 min. It is therefore certain that Witelo laid his hand upon his refractions gathered from experience so as to bring them into order through an equality of the second increments. (Donahue 2000)

Here Kepler shows that he understands that the tables are adjusted into an order of decreasing increments. He also believes that there are some hidden relations between the angles and that observational data only approximates data arrived by calculations. His method, as revealed here, starts from a hypothetical relation for which he constructs an algebraic (as in Kepler 1604, Chap. 4, Proposition 8) or geometrical model (discussed below). From this he calculates the expected angles of refraction and checks them against observational data (Table 4.4).

For the second case of Kepler's third law, two parameters determine the law: the average distance of planets from the sun and the periods of their revolution around the sun. Table 4.6 lists the data used by the BACON program, data available to Kepler from Tycho Brahe and data used by Kepler when writing the *Harmonices Mundi*.

Given the modifiable noise margin of BACON.3 the data for the average distance of planets from the sun can be considered the same as the historical data. However, for the second parameter BACON.3 "is given the angle found by using the fixed star and the planet as the two endpoints and the primary body (the Sun) as the pivot point" (Langley et al. 1987, p. 99). This data allows one to calculate an invariant slope for each planet. Nowhere amongst the many tables in the *Harmonices Mundi* do we find a table like this.

In conclusion we may state that the first condition for an adequate explanation of historical cases, the use of the same historical data, is not met for the two cases discussed here. Despite the claims made by the BACON team their program is given data which is different from the actual data used at the time of the discoveries. This seriously undermines the descriptive and explanatory ambitions of the BACON program.

4.6 The Historical Methods

Having established that BACON's input data does not follow the historical sources we can deal with the second, more important question: does the method of data-driven induction fit the historical discoveries? Let us start with the sine law of refraction. As is now established, the sine law was discovered independently by Thomas Harriot around 1602, by Willebrord Snellius in 1621 and by René Descartes between 1626 and 1628 (Schuster 1978). Descartes was the first to publish the law in his *Dioptrique* of 1637. In some sense you can also say that Pierre Fermat, who tried to disprove Descartes, came to the same discovery in 1662. Kepler attempted to formulate a law in his *Paralipomena* of 1604. The fact that these natural philosophers came to the same discovery independently within a very short time period is an interesting phenomenon of scientific discovery. If accurate data were available since Ptolemy and the discovery is only a matter of data-driven induction, why did it take 15 centuries to come to the sine law? Furthermore, why did several individuals came to the discovery within a matter of a few decades? The answer is beyond the scope of the current paper but the contributions of the medieval perspectivist tradition should be mentioned here. These contributions include the decomposition of rays into orthogonal components and the idea of the conservation of one component during refraction (Heeffer 2006). By the beginning of the seventeenth century there was a consensus that there must exist some geometrical relation or an arithmetical proportion between the components of the rays of incidence and the refracted rays. This insight is very well reflected in Kepler's introduction to "the measurement of refraction":

> Since density is obviously a cause of refraction, and refraction itself appears to be a kind of compression of light (i.e., towards the perpendicular), it comes to mind to ask whether the ratio of the media in the case of densities is the same as the ratio of the bottom of the spaces that light has entered into and strikes, first in an empty vessel, and then one filled with water. (Kepler 1604, Chap. 4, §2); translation from (Donahue 2000, p. 102))

So, Kepler starts from the general hypothesis that in a geometrical model of refraction (shown in Fig. 4.2) some component of the ray of incidence is in a specific ratio to a corresponding component of the refracted ray. This ratio should be invariant with the angle of incidence. To able to check the invariances he draws three angles of incidence, AFC, AGC and ABC. The clever representation of Fig. 4.2 allows Kepler to reason about refraction using geometrical knowledge of rectangular

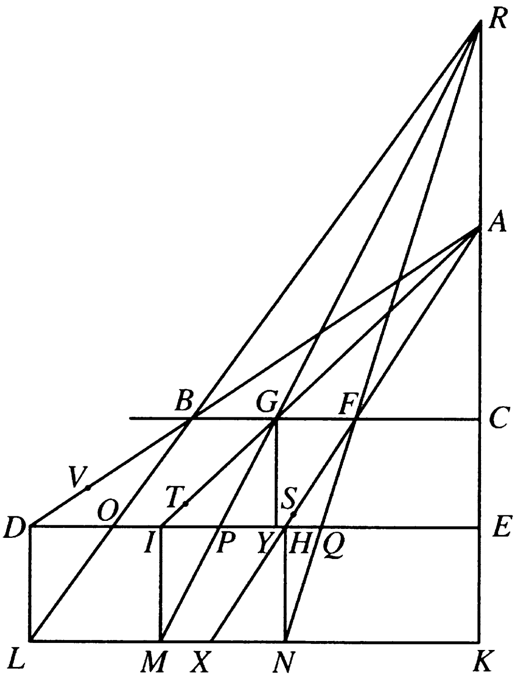


Fig. 4.2 Kepler's geometrical model for refraction

triangles. Based on the refraction tables he then draws three corresponding angles of refraction, RFC, RGC and RBC. He then considers the ratio's between components of the incidence rays and the refracted rays. The first one, which I will call H_1, looks at the ratio of the distance of the refracted ray at the bottle of the vessel with the distance of the unrefracted ray, e.g. EQ/EH. The next step is to either prove or disprove the hypotheses considered. Kepler's approach is to eliminate hypotheses by either deduction or observation. In order to be invariant the EQ/EH should be the same as EP/EI and EO/ED. Kepler dismissed this first hypothesis as it "is refuted by experience". He then goes on formulating alternative hypotheses. They are summarized in Table 4.7. Each of the formulated hypotheses gets refuted in the end.

With some disappointment Kepler moves on to a next section writing "Hitherto, we have followed an almost blind plan of enquiry, and have called upon luck". However, his plan was not so blind, but he was unlucky. In Sect. 4.7 I will demonstrate that his method was the right one to discover the sine law.

4.7 Kepler's Method for the Sine Law

We can design a model based on Kepler's general hypothesis. In fact, the hypothesis he uses is more general than in the quotation above, because more ratios are

Table 4.7 Kepler's hypotheses

Hypothesis	Relation
H_1	EQ/EH
H_2	FQ/FH
H_3	EQ^2/EH^2
H_4	$a \times EQ/b \times EH$
H_5	$FHEC/FQEC$
H_6	EQ^3/EH^3
H_7	IY/IP
H_8	GC/IE
H_9	CE/CK
H_{10}	FH/FX
H_{11}	CK/FX

investigated than those on the bottom of the vessel. A fair reformulation of Kepler's central hypothesis would be the following:

> The ratio of the optical densities of two media is proportional to some ratio of two line segments in the geometrical representation of a light ray traversing the two media, the first line segment related to the angle of incidence, the second to the angle of refraction.

When this central hypothesis is implemented in a computational model, it would generate all of Kepler's hypotheses H_1 to H_{11}, as well as many more (Heeffer 2003). As can be noticed from Table 4.8, with exception of H_8 and H_{11}, in all ratios the two line segments have one point in common. If a restriction is added to the starting hypothesis, that the line segments used in the ratios, should start in the same point, the model generates only nine (linear) instances. These ratios cover all of Kepler's except the special case H_7.

The next step is to either prove or disprove the generated hypotheses. Kepler's approach is to eliminate hypotheses by deduction or observation. Most hypotheses can be refuted deductively, by simple geometrical reasoning. For example, H_{11} states that the ratio CK/FX remains the same for varying angles of incidence. This is evidently not the case. As the angle of incidence increases, the length of FX will increase while CK remains the same. Therefore the ratio CK/FX decreases and hypothesis H_{11} is refuted. The tragedy is that Kepler succeeded in formulating both a suitable representation of the problem and the correct hypothesis that some geometrical proportion corresponds to the refraction index. He failed in identifying the correct ratio. Both FR/FA and FN/FH correspond to the ratio of optical densities of the two media. These ratios can be proved to be constant by geometrical reasoning.

Line segment FC allows establishing a relation between the angles of incidence and refraction as it is the same side of the right-angled triangles FRC and FAC. This unfortunate oversight was Kepler's failure in discovering the sine law. Both Snell and Descartes read the *Paralipomena* and undoubtedly found here their main inspiration for the sine law, Snell's formulation was based on a ratio of cosecants equaling FH/FN (Volgraff 1918, p. 21b).

$$
\begin{aligned}
FA \times \sin FAC &= FR \times \sin FRC \\
GA \times \sin GAC &= GR \times \sin GRC \\
\frac{\sin FAC}{\sin FRC} &= \frac{FR}{FA} = \frac{FN}{FH}
\end{aligned}
$$

Descartes never mentioned his sources and took care not to reveal his path of discovery of the sine law. Several authors have formulated hypotheses on how Descartes came to his discovery. In an early study by Kramer, it was suggested that Descartes hit upon the sine law through his study of conic sections, in particular the problem of the anaclastic curve (Kramer 1882, pp. 256–258 and note 39). Others such as William Shea believe that Descartes used the sine law for solving the anaclastic (Shea 1991). Shea argues that the demonstration of a refractometer, presented by Descartes in a letter to his lens cutter Ferrier, indicates the procedure leading to the sine law. Given Kepler's analysis as sketched above and the fact that Descartes called Kepler "my first teacher in optics" (in a letter from Descartes to Mersenne, 31 March 1638; Adam and Tannery 1986–1909, AT, II, 86). I consider it most likely that Descartes found in Kepler's drawing and his main hypothesis everything needed to deduce the sine law by pure geometrical reasoning (Heeffer 2006). None of the scholars who have worked on the discovery of refraction by Harriot, Snell or Descartes have reported any evidence pointing to the use of data-driven induction. Above is shown that Kepler's method is an adequate one in arriving at the sine law and that it is quite different from data-driven induction.

4.8 Kepler's Method for the Third Law

Kepler's discovery of his third law is not as well documented as his analysis of refraction. The first and second law were published in his most important work, the *Astronomia nova* of 1609. A quantitative relation between the average distance of planets to the sun and their period eluded him for many years. It was while writing the *Harmonices mundi* several years later that he hit upon the relation of $2/3$ of the powers. The law is buried within the text on p. 189: "The proportion between the periodic time of any two planets is precisely sesquialter $2/3$ the power of their mean distances", as if it were some side remark. But Kepler knew the significance of his discovery. He distinctively remembered the precise date of his discovery:

> it was conceived on March 8 of the year 1618, but unfelicitously submitted to calculation and rejected as false, and recalled only on May 15, when by a new onset it overcame by storm the darkness of my mind with such full agreement between this idea and my labor of 17 years on Brahe's observations that at first I believed I was dreaming and had presupposed my result in the first assumptions. (translation from Gingerich (1975, p. 595))

For 17 years he possessed the necessary data to discover the law. Why did he find it in 1618 and was it reached by data-driven induction? Kepler does not tell us and neither does he give a justification for the law. It appears in the book as an

Table 4.8 Comparing the orbital radii calculated by harmonic proportions with observation

Planet	Harmonic	*Mysterium*
Mercury	0.385	0.360
Venus	0.738	0.719
Earth	1.000	1.000
Mars	1.505	1.520
Jupiter	4.745	5.246
Saturn	8.837	9.164

auxiliary finding while dealing with the harmonies of the world. It “overcame” him as it were not the result of a conscious process of induction but as a simple ratio that did not occur to him before (reminiscent of Koestler 1959, who portrays Kepler as a sleepwalker). To understand his method it is necessary to place his discovery in the right context. Kepler had recorded an epiphany like this before on 19 July of 1595. He then realized that the respective radii of the five known planets measure in close correspondence with the radii of five Platonic solids nested in a specific order. This idea together with the famous depiction of the nested spheres was published in his first book, the *Mysterium Cosmographicum* or *Secret of the Cosmos* (Kepler 1596). Kepler was a deeply religious man who believed that God ordered the universe using principles of geometry. In order to understand the universe one has to look for geometrical relations which form the basis of the model that God had in mind. The *Harmonices mundi*, which contains the third law, is very much like his first book. Here Kepler comes back to his previous ideas on the relation between platonic solids and the orbits of planets but also adds and combines this with properties of regular polygons and harmonic principles of music, hence the title, *Harmonies of the World*. From the many principles he explores, one can discern a consistent pattern. The idea of a planned universe leads Kepler to look for principles of proportion in plane and solid geometry and in music. In each instance he tries to bring proportions from music or geometry in correspondence with observational data. In other words he fits a priori data available from mathematical theories with synthetic data arrived by observations. This is the opposite of data-driven induction, which formulates general laws in the language of mathematics from empirical data.

Table 4.8 shows an interesting example from lesser known work done by Kepler in 1599 (described by Stephenson in 1994, pp. 90–97). The first column lists the orbital radii of the planets calculated as musical proportions, starting with the ratio 3:4. These data are compared with observational data from the *Mysterium cosmographicum*. The *Harmonices Mundi* contains many such examples and the correspondence between the observed data and the calculated relations is often very close.

Concerning the third law, we do not know exactly how Kepler came to the idea. However, given the context of the book and his methodology summarized here we can assume that he fitted a mathematical proportion, “sesquialter the powers”, with observational data. As Kepler formulated sub-hypotheses using square and cubic terms as in H_6 and H_9 for his study of refraction, he may have proceeded along the same course for the ratio between the period and average distance of planets around the sun. In conclusion we may safely state that data-driven induction does not at all fit the process that led to the discovery of Kepler's third law.

4.9 The Value of the Data-Driven Induction Model

Drawing further on the analogy with chess programming I have demonstrated that BACON is as limited in explaining the mechanisms of discovery in the two historical cases discussed as a chess program is in explaining the problem solving capacities of chess grand masters. The next question that I will address is if BACON is as good in scientific discovery as current chess programs are in beating the world champion.

Reading through the many "discoveries" by BACON in the Langley book, one cannot escape the impression that data has been contrived and procedures have been arranged ad hoc to discover what is already known. The data for discovering the sine law (Table 4.1) gives away the discovery. The real discovery was made by Kepler, namely the hypothesis that some ratio between the components of the angles of incidence and refraction remains invariant with the degree of the angles. He failed to find the right line segments in his geometrical model but others such as Harriot, Descartes and Snell did, probably reasoning in the same way. Furthermore I have shown that when applying data-driven induction to the historical data this would lead to a quadratic relation between the angles of incidence and refraction which is different from the sine law. Actually, also the data used by BACON.1 for "discovering" Keplers's third law can lead to many, probably infinitely many, possible quantitative laws. The fact that BACON.1 finds precisely Kepler's third law is very suspicious as it involves a combination of a quadratic and a cubic term. The relation between the data for period and average distance in Table 4.2 can be expressed by one simple quadratic relation:

$$P = \frac{11}{60}D^2 + \frac{17}{12}D - \frac{6}{10}$$

It can easily be verified that for the input data $D = 1, 4$ and 9 this leads to $P = 1, 8$ and 27. Why did BACON.1 not find this simpler quadratic relation or any similar one? Because the production rules were set up to find Kepler's third law. In all the "discoveries" involving a first pass through the program for finding a slope, such as with Kepler's third law in BACON.3, the slope already represents the relevant relation between the parameters.

So, the claims by the BACON team that their program is constructive, that it did "rediscover" all these laws in physics and optics can only be met with scepticism. As data-driven induction is not adequate to describe the historical discoveries I have discussed, nor is it adequate as a general exclusive framework for doing discovery. Data-driven induction would produce an infinite number of laws that fit the input data but that are meaningless and break down for additional cases. As a matter of fact, Kepler himself was aware of this. As pointed out to me by an anonymous

Table 4.9 Data for the Titius-Bode law

Planet	k	Calculated	Observed	Error (%)
Mercury	0	0.4	0.39	2.56
Venus	1	0.7	0.72	2.78
Earth	2	1.0	1.00	0.00
Mars	4	1.6	1.52	5.26

referee of this paper, Kepler himself argued in Chap. 21 of the *Astronomia Nova* that a process of data fitting can emulate any varying magnitude to any degree of accuracy without giving the least insight into the reality behind the variation. He concludes the chapter with:

> And so this sly Jezebel [the fitted curve] cannot gloat over the dragging of truth (a most chaste maiden) into her bordello. Any honest woman following the lead of this prostitute would stay closely in her tracks owing to the narrowness of the streets and the press of the crowd, and the stupid, bleary-eyed professors of the subtleties of logic, who cannot tell a candid appearance from a shameless one, judge her to be the prostitute's maidservant. (quoted and translated by Goddu in Goddu (2010, p. 431))

It is further instructive to keep in mind the so-called Titius-Bode law. The law was formulated during the eighteenth century as an inductive generalization of the respective distances of planets around the sun. In its modern formulation, the law expresses a measure in astronomical units a of the semi-major axis, or the longest axis of the elliptical orbit, such that

$$a = 0,4 + 0,3k \quad \text{where } k = 0, 1, 2, 4, 8, 16, 32, 64, 128.$$

Table 4.9 shows that the law fits the observational data well for the first four planets. Then there is a gap for $k = 8$ which could be filled up by the dwarf planet Ceres. It further continues to go well up to Uranus but the law breaks down for Neptune and Pluto.

Many possible mathematical relations can be found by induction without being laws of science. A physical law is more than a generalization of data. A physical law is part of a broader theory and should fit with other laws in a mathematical meaningful way. A good example is the reformulation of Kepler's third law in Newton's theory of universal gravitation. Newton established that two orbiting bodies with mass m_1 and m_2 with period P having respective distances to a centre of mass d_1 and d_2 relate as:

$$(m_1 + m_2)P^2 = (d_1 + d_2)^3 = D^3$$

Kepler's third law is thus only approximately true. Because the mass of the sun is much larger than those of the planets the value of $m_1 + m_2$ approximates m_2. Given the mathematical connection between Kepler's third law with Newton's theory, after Newton the former can be understood as a derived law of physics rather than an inductive generalization from the data. The isolated context of data generalization thus raises questions on the usability of data-driven induction for physical laws.

4.10 Conclusion

I have presented an assessment of the BACON program which is very critical. There are serious methodological problems with their modelling of scientific discovery. They make claims about the explanatory power of historical cases which cannot be sustained by historical research. They present BACON as a *general* model for scientific discovery, which it is not. Data-driven induction may have some function in scientific discovery but it is certainly not the decisive creative step as presupposed by some AI approaches. However, this does not imply that scientific discovery is beyond rational explanation or models of artificial intelligence. On the contrary, the rational process of hypothesis formulation and testing—as I have proposed—does provide an alternative model for scientific discovery, and is one which can be supported by historical research. The state-space search model for discovery and problem solving, pioneered by Simon, allows for the modelling of scientific discovery without having to rely on one single, simple and ad-hoc method, such as data-driven induction.

Acknowledgements Apart from the LRR10 conference, this paper was presented at a research seminar at the Sydney Center for the Foundations of Science at Sydney University where the author was invited as a visiting fellow in 2011. The text benefited from valuable comments by two anonymous referees as well as comments by Stephen Gaukroger, Ofer Gal and John Schuster.

References

Adam, C., & Tannery, P. (Eds.). (1896–1909). Oeuvres de Descartes (vol. 11). Paris: Librairie Philosophique J. Vrin.

Donahue, W. (2000). *Kepler's optics*. Green Lion, Santa Fe.

Gingerich, O. (1975). The origins of Kepler's third law. In: A. Beer & P. Beer (Eds.), *Kepler: Four hundred years* (Vistas in astronomy, vol. 18, pp. 595–601). New York: Pergamon.

Goddu, A. (2010). Copernicus and the Aristotelian tradition: education, reading and philosophy in copernicus's path to heliocentrism. Leiden: Brill.

Govi, G. (1885). L'Ottica die Cl. Tolomeo da Eugenio. Torino.

Heeffer, A. (2003). Kepler's near discovery of the sine law: A qualitative computational model. In: C. Delrieux & J. Legris (Eds.), *Computer modeling of scientific reasoning* (pp. 93–102). Bahia Blanca: Universidad Nacional del Sur.

Heeffer, A. (2006). The logic of disguise: Descartes' discovery of the sine law. Historia scientiarum. *International Journal of the History of Science Society of Japan, 16*(2), 144–165.

Hempel, C. (1945). Studies in the logic of confirmation. *Mind, 54*(1–26), 97–121.

Kepler, J. (1596). Prodromus dissertationum cosmographicarum, continens mysterium cosmographicum, de admirabili proportione orbium coelestium ...: Demonstratum, per quinque regularia corpora geometrica, Gruppenbach, Tübingen. [A. M. Duncan, Trans. with notes by E. J. Aiton. (1986) *The Secret of the Universe*. New York, Abaris. Books.]

Kepler, J. (1604). *Ad Vitellionem paralipomena, quibus astronomiae pars optica traditur....* Claudius Marnius & heirs of Johann Aubrius: Frankfurt. See Donahue (2000). *Kepler's Optics*. Santa Fe: Green Lion.

Kepler, J. (1619). *Harmonices Mundi Libri V* (Translated into English with an Introduction and Notes by Aiton, E. J., Duncan, A. M., & Field, J. V., 1997). Linz: Ioannes Plancus.

Kepler, J. (1858–1871). *Joannis Kelpleri Astronomi Opera Omnia*. Frankfurt: Heyder and Zimmer.

Koestler, A. (1959). *The sleepwalkers: A history of man's changing vision of the universe*. London: Hutchinson.

Kramer, P. M. (1882). Descartes und das Brechungsgesetz des Lichtes. Abhandlungen zur Geschichte der Mathematik. *Zeitschrift für Mathematik und Physik, XXVII*, 233–278.

Langley, P., Bradshaw, G. L., & Simon, H. A. (1981). BACON.5: The discovery of conservation laws. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 1*, 121–126.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT.

Lejeune, A. (1956). Ptolémée. L'optique de Claude Ptolémée dans la version latine d'après l'arabe de l'émir Eugène de Sicile. Edition critique et exégétique par Albert Lejeune. Publications universitaires de Louvain.

Neugebauer, O. (1957). *The exact sciences in antiquity* (2nd ed.). Reprinted by Dover 1969. Princeton: Princeton University Press.

Newell, A., Simon, H. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Nickles, T. (1994). Enlightenment versus romantic models of creativity in science — and beyond. *Creativity Research Journal, 7*, 277–314.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Book.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.

Risner, F. (Ed.) (1572). *Opticae Thesaurus*. Basel: Per Episcopios.

Schuster, J. A. (1978). Descartes and the scientific revolution, 1618–1634: An interpretation. PhD thesis, Princeton University, University Microfilms International, Ann Arbor.

Shea, W. R. (1991). *The magic of numbers and motion: The scientific career of René Descartes*. Cambridge: Science History Publications.

Smith, A. M. (1996). *Ptolemy's theory of visual perception: An English translation of the optics with introduction and commentary*. Philadelphia: American Philosophical Society.

Stephenson, B. (1994). *The music of the heavens: Kepler's harmonic astronomy*. Princeton: Princeton University Press.

Unguru, S. (Ed.). (1977). *Witelonis perspectiva, liber primus*. An English translation with introduction and commentary and Latin edition of the mathematical book of Witelo's Perspectiva, Studia Copernicana, XV. Warsaw: Polish Academy of Sciences.

Volgraff, J. A. (1918). *Risneri Opticum cum Annotationibus Willebrordi Snellii*. Gent: Aedibus Plantini.

Chapter 5
Dovetailing Belief Base Revision with (Basic) Truth Approximation

Theo A.F. Kuipers

5.1 Introduction

Recently, Roberto Festa as well as Gustavo Cevolani and Francesco Calandra have developed a qualitative and quantitative account of verisimilitude of 'conjunctive theories' of a finite propositional language (Festa 2007; Cevolani and Calandra 2009). The qualitative version of this 'conjunctive' account turned out to be formally equivalent to the definition of 'descriptive verisimilitude' in Kuipers (1982). Cevolani et al. (2011) managed to show that this account can be nicely linked to a variant of AGM belief *set* revision, viz. belief *base* revision, in the sense that the latter kind of revision is functional for truth approximation according to the conjunctive account. Following an idea of Festa, it is shown by Cevolani et al. (2013) that all this can be transformed for the purposes of basic theoretical or nomic verisimilitude, as initiated in Kuipers (1982) and elaborated in Kuipers (2000), and corresponding nomic truth approximation by belief base revision.

In the present paper I offer a generalization of these ideas to the case of approaching any divide of a (finite or infinite) universe allowing several interpretations, besides true (false) atomic propositions, notably nomic states (not) in equilibrium, nomic (im)possibilities, and (non-)instantiated 'Q-predicates' of a monadic language. To be sure, the paper is not very original relative to the one of Cevolani et al. (2011). They convincingly show that the restriction to conjunctive (propositional) theories not only enables the application of the belief *base* revision operations, but even in such a way that they, as mentioned, become functional for truth approximation (Cevolani et al. 2011, Sect. 4). This is a great improvement relative to the problematic relation between belief *set* revision and truth approximation, as already pointed out by Niiniluoto (1999). Moreover, they

T.A.F. Kuipers (✉)
University of Groningen, Groningen, The Netherlands
e-mail: T.A.F.Kuipers@rug.nl

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic,
Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_5,
© Springer Science+Business Media Dordrecht 2014

claim already in the final section that their "approach can be generalized to any language characterized by a suitable notion of constituent [...]" (Niiniluoto 1999, p. 198). In such languages, in fact, a c[onjunctive]-theory can be conveniently defined as a 'fragment' of a constituent and they refer more specifically to first order, nomic and even to statistical languages. The main purpose of the present paper is to offer a transparent formalization of such a general account, though restricted to non-statistical languages, which may well function as a start for extensions and refinements.

Point of departure is a, possibly infinite, universe of discourse (U), and a vocabulary (V) in which subsets of U, e.g. X^+, Y^-, R^+, are characterized. $\underline{\mathbf{T}} = \langle \mathbf{T}^+, \mathbf{T}^- \rangle$, note the bold characters, is a divide (formally: a bivalent partition) of U, that is, T^+ and T^- are non-overlapping subsets of U that together exhaust U. Crucial is the assumption that the divide is not (yet) given in terms of V. The target of research is supposed to be the identification, if possible, of the boundary of the two sets in V-terms. Such a characterization, if it exists, will be denoted by $\underline{T} = \langle T^+, T^- \rangle$, with non-bold $\underline{T}$, T^+ and T^-. This target will be more specifically called 'the true theory/boundary', or 'the truth' which is searched for.

In general, U amounts to a set of objects or items and $\mathbf{T}^+(\mathbf{T}^-)$ to the subset of objects to which a certain property P does (not) apply. In the propositional interpretation, U and V essentially coincide by being a (finite) set of atomic propositions and *(non-)P* corresponds to *true (false)*. In the 'partition-interpretation' (Kuipers 1982), U is a set of conceptually possible states generated by V and *(non-)P* corresponds to *(not) in a so-and-so state*, for example, *(not) in equilibrium* or *the bulb is (not) lighting*. As far as these interpretations are concerned both alternatives can be verified, they will be called symmetric interpretations, leading to symmetric sets of data. A typically asymmetric interpretation is the nomic interpretation, leading to asymmetric sets of data. Here U is a set of conceptual possibilities generated by V and *(non-)P* corresponds to *physically* or, more general, *nomically (im-)possible*. Whereas we may be able to verify by a single experiment that a conceptual possibility is in fact physically possible, it is not possible in some straightforward way to show that it is in fact not physically possible. It is easy to check that the same holds for the monadic-existential interpretation, in which U is a set of Q-predicates, i.e., a set of mutually exclusive and together exhaustive predicates generated by a vocabulary of primitive predicates. Here *(non-)P* corresponds to *(not) instantiated*. It is plausible to assume that there will be other essentially different interpretations of symmetric or asymmetric nature.

Our program in the paper is as follows. In Sect. 5.2 I will define a suitable notion of theories, their truth and falsity content, and the comparative expression "theory $\underline{Y}$ is at least as close to the truth/the true boundary as theory $\underline{X}$". In Sect. 5.3 I will tailor the belief (base) revision operations, viz., expansion, contraction, and revision, to the present context. More specifically, I will define the revision, $\underline{X} * \underline{A}$, of theory $\underline{X}$ in the light of an input theory $\underline{A}$. In Sect. 5.4 I will introduce the distinction between symmetric and asymmetric data and prove for (a) data (theory) $\underline{R}$ of either type that the revision ($\underline{X} * \underline{R}$) is closer to the truth than $\underline{X}$ if we assume that the data are true/correct. In Sect. 5.5 I will present quantitative versions of the relevant notions; notably, a quantitative measure for the closeness to the truth of a theory, i.e., its truthlikeness or verisimilitude (cf. Sect. 5.2) and the increase of verisimilitude by belief base revision by true data (cf. Sect. 5.4). I will end with some questions for further research and concluding remarks.

5.2 A Set-Theoretical Framework for Kinds of Basic Truth Approximation

In any context in which the target of research is supposed to be the identification of the boundary of a divide $\underline{\mathbf{T}} = \langle \mathbf{T}^+, \mathbf{T}^- \rangle$ of U in terms of a given vocabulary V, it is plausible to assume that a theory is a tentative proposal for objects or items that belong to one of the two subsets.

Formally this can be represented in set-theoretical fashion as follows. A theory $\underline{X}$ is a tuple $\langle X^+, X^- \rangle$ of non-overlapping subsets X^+, X^- of U, defined in terms of V, with the claim "$X^+ \subseteq \mathbf{T}^+$" associated to X^+ and the claim "$X^- \subseteq \mathbf{T}^-$" associated to X^-. Note that these claims are universally quantified conjunctive claims, that is:

$$X^+ \subseteq \mathbf{T}^+ \equiv \forall_{u \in X^+}\ u \in \mathbf{T}^+ \text{ and } X^- \subseteq \mathbf{T}^- \equiv \forall_{u \in X^-}\ u \in \mathbf{T}^-$$

Each conjunct will be called a *basic (b-)claim*. All possible 'positive' or 'negative' b-claims, for different u in U, are assumed to be logically independent. X^+ is called the *positive range* of $\underline{X}$, X^- its *negative range*, and $X^+ \cup X^-$ its *(total) range*. It is important to stress that $\langle X^+, X^- \rangle$ is supposed to represent theory $\underline{X}$, including the associated claims. Note that $\underline{\mathbf{T}} = \langle \mathbf{T}^+, \mathbf{T}^- \rangle$ is an improper theory: the constitutive sets are not defined in V-terms and its two claims are trivially true. However, the target $\underline{T} = \langle T^+, T^- \rangle$ is a proper theory.

It is not easy to evaluate the assumption of the logical independence of the b-claims. At first sight it might seem a very strong restriction. However, it should be realized that the assumption does not preclude that the formal characterization of a theory $\underline{X}$ fulfills criteria of being well-formed, such as formalizability in a certain way and forms of generality. Moreover, it may well be based on a fundamental idea. The more important restriction is that to conjunctive theories in the sense indicated by the universally quantified form of its claims. In Sect. 5.5 we will hint at the generalization of the present approach to theories in general, where the truth-value of individual b-claims may depend on that of other, not on logical grounds, but because the theory requires so.

In Table 5.1, we present both the logical and the set-theoretical representation of one example, first in general terms and then in terms of the four interpretations that were mentioned in the introduction.

Some general notions will be helpful below. Theory $\underline{X}$ is *maximal* (or a constituent theory) iff its range is maximal, i.e. $X^+ \cup X^- = U$. In this case

Table 5.1 Interpretations and representations of an example

Interpretation	Logical representation		Set-theoretical representation	
			$\underline{X} = \langle X^+$	$, X^-\rangle$
			"$X^+ \subseteq \mathbf{T}^+$"	"$X^- \subseteq \mathbf{T}^-$"
General	$Pa_1 \,\&\, Pa_2 \,\&\, \neg Pa_3$		$\langle\{a_1, a_2\}$	$, \{a_3\}\rangle$
	$(\neg)Pa_i$	P applies (does not apply) to a_i	"$\{a_1, a_2\} \subseteq \mathbf{T}^+$"	"$\{a_3\} \subseteq \mathbf{T}^-$"
Propositional	$p_1 \,\&\, p_2 \,\&\, \neg p_3$		$\langle\{p_1, p_2\}$	$, \{p_3\}\rangle$
	$(\neg)p_i$	p_i is true (false)	"$\{p_1, p_2\} \subseteq \mathbf{T}^+$"	"$\{p_3\} \subseteq \mathbf{T}^-$"
Partition	$Es_1 \,\&\, Es_2 \,\&\, \neg Es_3$		$\langle\{s_1, s_2\}$	$, \{s_3\}\rangle$
	$(\neg)Es_i$	s_i is a (non) equilibrium state	"$\{s_1, s_2\} \subseteq \mathbf{T}^+$"	"$\{s_3\} \subseteq \mathbf{T}^-$"
Nomic	$Nu_1 \,\&\, Nu_2 \,\&\, \neg Nu_3$		$\langle\{u_1, u_2\}$	$, \{u_3\}\rangle$
	$(\neg)Nu_i$	u_i is a nomic (im)possibility	"$\{u_1, u_2\} \subseteq \mathbf{T}^+$"	"$\{u_3\} \subseteq \mathbf{T}^-$"
Monadic-	$\exists x Q_1(x) \,\&\, \exists x Q_2(x) \,\&\, \neg\exists x Q_3(x)$		$\langle\{Q_1, Q_2\}$	$, \{Q_3\}\rangle$
Existential	$(\neg)\exists x Q_i(x)$	Q_i is (not) instantiated	"$\{Q_1, Q_2\} \subseteq \mathbf{T}^+$"	"$\{Q_3\} \subseteq \mathbf{T}^-$"

$X^- = \mathrm{c}X^+$ (where $\mathrm{c}X^+$ denotes the complement of X^+, i.e., $U - X^+$). Note that $\underline{\mathbf{T}} = \langle \mathbf{T}^+, \mathbf{T}^- \rangle$ is maximal (but, recall, an improper theory). Finally, for theory $\underline{X} = \langle X^+, X^- \rangle$, theory $\langle X^-, X^+ \rangle$ is called the *specular* of $\underline{X}$, indicated by $\mathrm{sp}(X)$. Hence the claims of $\mathrm{sp}(\underline{X})$ are "$X^- \subseteq \mathbf{T}^+$" and "$X^+ \subseteq \mathbf{T}^-$", respectively.

Now we can start to introduce the crucial notions of truth and falsity content and, in these terms, the notion of basic qualitative truthlikeness. The *positive truth content* of $\underline{X}$ (or the set of true positives) is $\mathrm{t}^+(\underline{X}) = X^+ \cap \mathbf{T}^+$. The *negative truth content* of $\underline{X}$ (or the set of true negatives) is $\mathrm{t}^-(\underline{X}) = X^- \cap \mathbf{T}^-$. The *truth (t-)content* of $\underline{X}$ is the union of these sets: $\mathrm{t}(\underline{X}) = \mathrm{t}^+(\underline{X}) \cup \mathrm{t}^-(\underline{X})$. Similarly, the *positive falsity content* of $\underline{X}$ (the set of false positives) is $\mathrm{f}^+(\underline{X}) = X^+ \cap \mathbf{T}^-(= X^+ - \mathbf{T}^+)$, the *negative falsity content* of $\underline{X}$ (the set of false negatives) is $\mathrm{f}^-(\underline{X}) = X^- \cap \mathbf{T}^+(= X^- - \mathbf{T}^-)$, and the *falsity (f-)content* of $\underline{X}$ is the union of both: $\mathrm{f}(\underline{X}) = \mathrm{f}^+(\underline{X}) \cup \mathrm{f}^-(\underline{X})$. It is easy to check that $\underline{\mathbf{T}}$ itself has extreme values for these six notions for trivial reasons: $\mathrm{t}^+(\underline{\mathbf{T}}) = \mathbf{T}^+, \mathrm{t}^-(\underline{\mathbf{T}}) = \mathbf{T}^-, \mathrm{t}(\underline{\mathbf{T}}) = U$, and $\mathrm{f}^+(\underline{\mathbf{T}}) = \mathrm{f}^-(\underline{\mathbf{T}}) = \mathrm{f}(\underline{\mathbf{T}}) = \emptyset$. However, the target $\underline{T} = \langle T^+, T^- \rangle$ has them non-trivially.

Note also that the union of the truth content and the falsity content of theory $\underline{X}$ exhausts its total range, formally: $\mathrm{t}(\underline{X}) \cup \mathrm{f}(\underline{X}) = X^+ \cup X^-$. Moreover, the union of the positive truth content and the positive falsity content $\underline{X}$ equals its positive range, formally: $\mathrm{t}^+(\underline{X}) \cup \mathrm{f}^+(\underline{X}) = X^+$. Similarly, the union of the negative truth content and the negative falsity content equals the negative range, formally, $\mathrm{t}^-(\underline{X}) \cup \mathrm{f}^-(\underline{X}) = X^-$. The four basic notions are depicted in Fig. 5.1, where $\mathbf{T}^-$, i.e. the complement of the indicated $\mathbf{T}^+$, is left implicit.

In my previous work on nomic truth approximation (Kuipers 2000), I have always restricted the attention to maximal theories. To connect the present unrestricted (and 'interpretation-free') analysis with it, I will from time to time present the results for maximal theories. To begin with, if $\underline{X}$ is maximal, the falsity content $\mathrm{f}(\underline{X})$ equals the symmetric difference of X^+ and $\mathbf{T}^+$, that is, $\Delta(X^+, \mathbf{T}^+) =_{\mathrm{df}} (X^+ - \mathbf{T}^+) \cup (\mathbf{T}^+ - X^+)$, which equals of course also the symmetric difference of their complements: $= \Delta(X^-, \mathbf{T}^-)$. Moreover, the truth content is the complement of the falsity content,

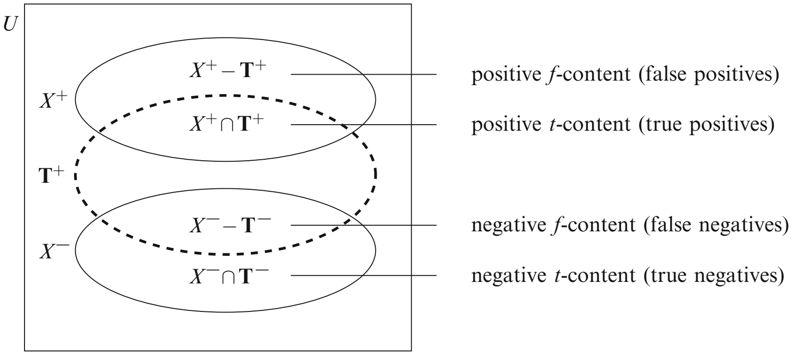

U
$X^+ - \mathbf{T}^+$
X^+
$X^+ \cap \mathbf{T}^+$
$\mathbf{T}^+$
$X^- - \mathbf{T}^-$
X^-
$X^- \cap \mathbf{T}^-$
positive *f*-content (false positives)
positive *t*-content (true positives)
negative *f*-content (false negatives)
negative *t*-content (true negatives)

Fig. 5.1 The four content parts of a theory

formally: $\mathrm{t}(\underline{X}) = \mathrm{c}(\mathrm{f}(\underline{X}))$. We can now also give plausible definitions of theories being true or false: $\underline{X}$ is *true* iff $\mathrm{f}(\underline{X}) = \emptyset$, and *false* otherwise. As an aside, $\underline{X}$ is said to be *completely false* iff $\mathrm{t}(\underline{X}) = \emptyset$, explicating the idea that the theory does not contain a grain of truth.

For the idea that there is something like 'the truth' to be approached, it is now crucial to note, and easy to check, that there is at most one true maximal (proper) theory, viz. $\underline{T} = \langle T^+, T^- \rangle$, the target of research! We can now state the definition of *basic truthlikeness* or *basic verisimilitude* in general terms.

Definition 1. Theory $\underline{Y}$ is (qualitatively) at least as close to theory $\underline{T}$, or to the truth, or verisimilar, as theory $\underline{X}$, $\underline{X} \leq_{\mathrm{vs}} \underline{Y}$, iff

$$\mathrm{t}(\underline{X}) \subseteq \mathrm{t}(\underline{Y}) \qquad \& \qquad \mathrm{f}(\underline{Y}) \subseteq \mathrm{f}(\underline{X})$$

or, in decomposed form:

$$\mathrm{t}^+(\underline{X}) \subseteq \mathrm{t}^+(\underline{Y}) \,\&\, \mathrm{t}^-(\underline{X}) \subseteq \mathrm{t}^-(\underline{Y}) \qquad \& \qquad \mathrm{f}^+(\underline{Y}) \subseteq \mathrm{f}^+(\underline{X}) \,\&\, \mathrm{f}^-(\underline{Y}) \subseteq \mathrm{f}^-(\underline{X})$$

By requiring at least once a *proper* subset relation we get the definition of "Theory $\underline{\mathrm{Y}}$ is (qualitatively) closer to theory $\underline{\mathrm{T}}$, or to the truth", but I will focus on the notion of 'at least as close to'. In the case of maximal theories the definition 'at least as close to' can be reduced due to the following theorem.

Theorem 1. *If theories* $\underline{X}$ *and* $\underline{Y}$ *are maximal,* $\mathrm{t}(\underline{X}) = \mathrm{c}(\mathrm{f}(\underline{X}))$ *and, vice versa,* $\mathrm{f}(\underline{X}) = \mathrm{c}(\mathrm{t}(\underline{X}))$*, and hence* $\mathrm{t}(\underline{X}) \subseteq \mathrm{t}(\underline{Y})$ *iff* $\mathrm{f}(\underline{Y}) \subseteq \mathrm{f}(\underline{X})$.

Consequently, for maximal theories one of the two conditions is enough for $\underline{Y}$ being at least as close to the truth as $\underline{X}$. As an aside, the 'f-condition', $\mathrm{f}(\underline{Y}) \subseteq \mathrm{f}(\underline{X})$, corresponds to the (basic) definition in the nomic interpretation, restricted to maximal theories, viz. $\Delta(Y^+, \mathbf{T}^+) \subseteq \Delta(X^+, \mathbf{T}^+)$ (Kuipers 2000).

Figure 5.2 depicts the situation in general that $\underline{Y}$ is at least as close to the truth as $\underline{X}$. The horizontally shaded areas are empty due to the fact that the f-content of $\underline{Y}$ is supposed to be a subset of the f-content of $\underline{X}$, whereas the vertically shaded areas are empty due to the fact that the t-content of $\underline{X}$ is supposed to be a subset of the t-content of $\underline{Y}$. Note that there are two double-shaded areas.

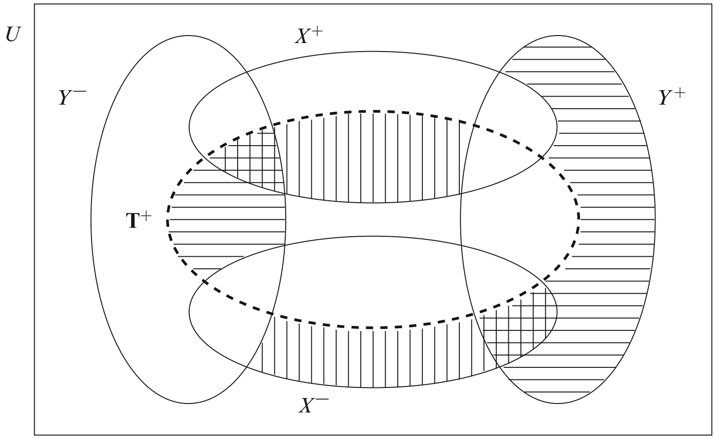


Fig. 5.2 Theory $\underline{Y}$ is at least as close to the truth as theory $\underline{X}$

5.3 Basic Belief Base Revision, Set-Theoretically Presented

As is well-known, *AGM-belief revision* centers around three (partially related) operations (Alchourrón et al. 1985), see also e.g. Hansson (1999, 2011) and Cevolani and Calandra (2009). A belief set, that is, a deductively closed set of sentences of a given language, is confronted with some 'input sentence' that, by minimal changes of the original belief set, either should become a consequence or no longer be a consequence of the revised belief set. For the first case, it makes an important difference whether or not the input sentence is compatible with the belief set. In the first subcase we get so-called *expansion*, viz., the belief set is strengthened to the set of consequences of the union of the belief set and the input sentence. Regarding the input sentence, assuming it was not already included, it leads from suspension of judgment about that sentence to its acceptance. In the second subcase the belief set has to be adapted in a more complicated way, satisfying certain axioms (see e.g. Alchourrón et al. 1985, p. 513; Hansson 1999, Chap. 3+). This operation is called *revision* (in the narrow sense). Regarding the input sentence revision leads from its rejection to its acceptance, except when the input sentence is inconsistent. Finally, in the second main case, the input sentence is supposed to belong to the belief set, but should no longer belong to the revised set. Hence, now the belief set has to be weakened in a minimal way, again in line with some axioms (see e.g. Alchourrón et al. 1985, p. 513; Hansson 1999, Chap. 2+). This operation is called *contraction*. Regarding the input sentence it leads from its acceptance to suspension of judgment, except when the input sentence is logically true, in which case it remains accepted after contraction. The three operations are related by the so-called *Levi-identity*, according to which the result of the revision of a belief set by an input sentence is or should be identical to the contraction of the belief set by the negation of the input sentence, followed by the expansion of the resulting belief set by the input sentence. The focus in the belief revision program has been on the link between, on the one hand, axiomatic characterizations of revision and contraction, and on the other hand, specific definitions of these operations, sometimes resulting in representation theorems.

As a variant of belief *set* revision, Hansson (1999, 2011) has developed so-called belief *base* revision, in which the operations are primarily defined for and applied to a base of sentences, taking its consequences into account when relevant. This variant fits best with our set-theoretical set-up. The (positive/negative) base of a theory $\underline{X}$ corresponds to its set of b-claims (associated with the positive range X^+ and with the negative range X^-, respectively) . New input corresponds to the base of another theory $\underline{A}$, which may represent empirical data or some other information that is given priority over $\underline{X}$.

Before we introduce the relevant definitions it will be useful to show in Fig. 5.3 the overlapping (o), conflicting (c) and excess (e) parts of theories $\underline{X}$ and $\underline{A}$, where three of them are named by way of example. That the pairs X^+ and X^- , and A^+ and A^- are non-overlapping is essential. However, that the elements of both pairs touch each other, even at the same place, is inessential and only for simplification of the pictures. The general picture is followed by Fig. 5.4 in which the following *propositional example* has been inserted. $\underline{X}$ corresponds to $(p_1 \,\&\, p_2 \,\&\, \neg p_3 \,\&\, p_4)$ and $\underline{A}$ to $(p_2 \,\&\, p_3 \,\&\, \neg p_5)$. Hence, we get $\underline{X} = \langle X^+, X^- \rangle = \langle \{p_1, p_2, p_4\}, \{p_3\} \rangle$ and $\underline{A} = \langle A^+, A^- \rangle = \langle \{p_2, p_3\}, \{p_5\} \rangle$. This example will be used for further illustrations.

It is of course possible to set-theoretically characterize all parts. The ordered pairs of corresponding positive and negative parts of the one relative to the other form theories themselves. The *overlapping* part of $\underline{X}$ relative to $\underline{A}$ is $\langle X^+ \cap A^+, X^- \cap A^- \rangle$, which amounts to $\langle \{p_2, \emptyset\} \rangle$ in the example. The *conflicting* part of $\underline{X}$ relative to $\underline{A}$ is $\langle X^+ \cap A^-, X^- \cap A^+ \rangle$, that is, $\langle \emptyset, \{p_3\} \rangle$ in the example. Finally, the *excess* part of $\underline{X}$ relative to $\underline{A}$ is $\langle X^+ - (A^+ \cup A^-), X^- - (A^+ \cup A^-) \rangle$, which can be rewritten as $\langle (X^+ - A^+) - A^-, (X^- - A^-) - A^+ \rangle$, that is $\langle \{p_1, p_4\}, \emptyset \rangle$ in the example. The overlapping part of $\underline{A}$ relative to $\underline{X}$ is of course the same as that of $\underline{X}$ relative to $\underline{A}$. The conflicting part of $\underline{A}$ relative to $\underline{X}$ is $\langle A^+ \cap X^-, A^- \cap X^+ \rangle$, which is the specular of that of $\underline{X}$ relative to $\underline{A}$, that is, $\mathrm{sp}(\langle X^- \cap A^+, X^+ \cap A^- \rangle)$. The excess part of $\underline{A}$ relative to $\underline{X}$ is $\langle A^+ - (X^+ \cup X^-), A^- - (X^+ \cup X^-) \rangle = \langle (A^+ - X^+) - X^-, (A^- - X^-) - X^+ \rangle$.

See Cevolani et al. (2011) for a detailed analysis of these parts, in the propositional interpretation, and for the way in which the three revision operations can be defined in terms of these parts. This can easily be generalized in set-theoretical terms. However, here I prefer another, straightforward, way of defining the operations.

We start with expansion, which is defined only for the case in which $\underline{X}$ and $\underline{A}$ are *compatible*, which is of course defined by the condition that the mutually conflicting parts are empty: $A^+ \cap X^- = \emptyset$ and $A^- \cap X^+ = \emptyset$, i.e. the c-areas in Fig. 5.3 are empty. This amounts to the assumption that none of the b-claims of $\underline{X}$ and $\underline{A}$ are in conflict. Note that $\underline{X}$ and $\underline{A}$ in the example are incompatible, hence, expansion cannot be illustrated with it.

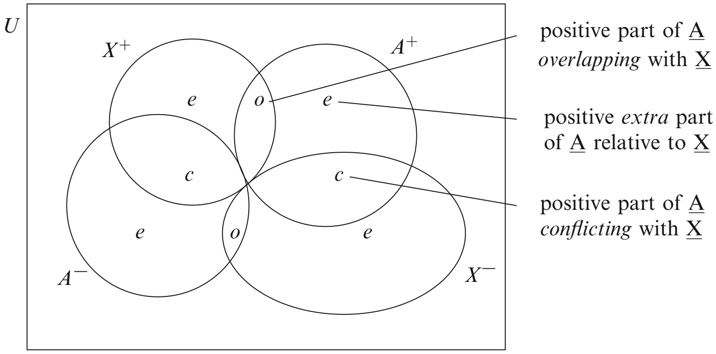
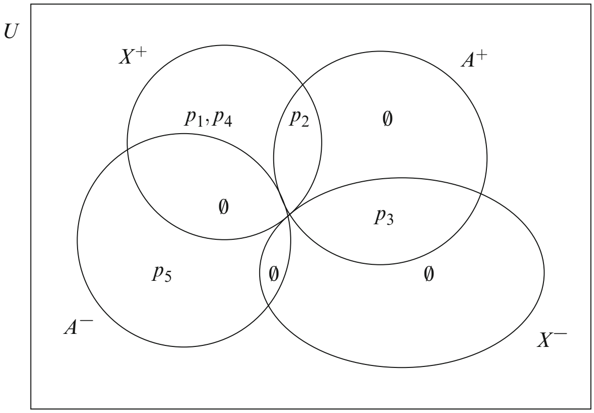

U
X^+
A^+
e
o
e
c
c
e
o
e
A^-
X^-
positive part of A overlapping with X
positive extra part of A relative to X
positive part of A conflicting with X

Fig. 5.3 The three positive parts of theory A relative to theory X

Fig. 5.4 All parts of $\underline{X} = \langle\{p1, p2, p4\}, \{p3\}\rangle$ relative to $\underline{A} = \langle\{p2, p3\}, \{p5\}\rangle$ and vice versa



In terms of b-claims, the idea of expansion of $\underline{X}$ by a compatible $\underline{A}$ is that the b-claims of A are added to those of X. Formally: expansion of X by a compatible A, denoted by $\underline{X} + \underline{A}$, is defined as $\langle X^+ \cup A^+, X^- \cup A^-\rangle$. It is easy to check that $\underline{X}$ + $\underline{A}$ is equal to the expansion of $\underline{X}$ by the excess part of $\underline{A}$ relative to $\underline{X}$. In Fig. 5.5, the positive and the negative range of the resulting theory have been indicated by plus- and minus signs, respectively, representing in total all b-claims of X and the excess b-claims of $\underline{A}$ relative to $\underline{X}$.

In terms of b-claims, the idea of contraction of $\underline{X}$ by $\underline{A}$ is that all relevant b-claims of $\underline{A}$ are deleted from those of $\underline{X}$. Formally: *contraction* of $\underline{X}$ by $\underline{A}$, denoted by $\underline{X} - \underline{A}$, is defined as $\langle X^+ - A^+, X^- - A^-\rangle$. It is now easy to check that $\underline{X} - \underline{A}$ is equal to the contraction of $\underline{X}$ by the overlapping part of $\underline{A}$ relative to $\underline{X}$. It is also equal to the expansion of the excess part of $\underline{X}$ relative to $\underline{A}$ by the conflicting part of $\underline{X}$ relative to $\underline{A}$. Again, in Fig. 5.6, the positive and the negative range of the resulting theory have been indicated, representing in total all conflicting b-claims of $\underline{X}$ relative to $\underline{A}$ and all excess b-claims of $\underline{X}$ relative to $\underline{A}$, that is, all b-claims of $\underline{X}$ that are non-overlapping with those of $\underline{A}$.

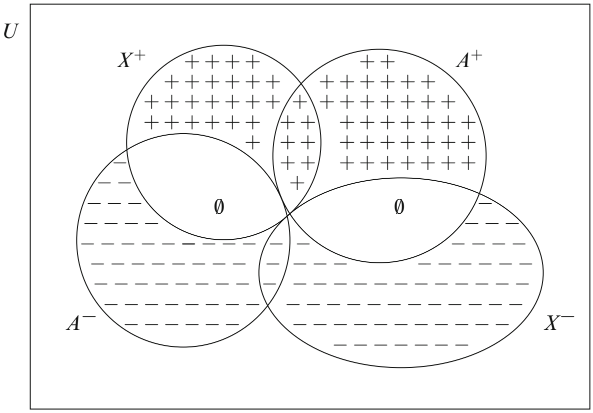


Fig. 5.5 The expansion of X by (compatible) A

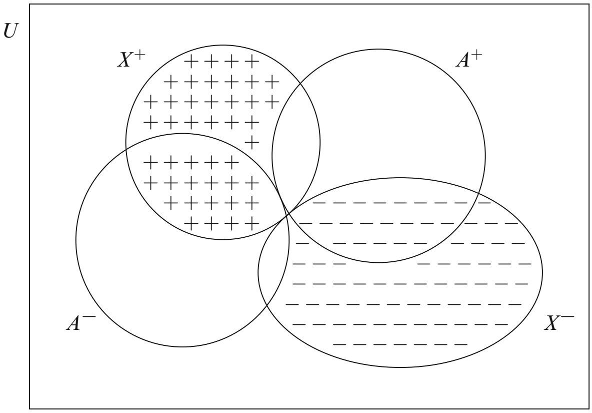


Fig. 5.6 The contraction of X by A

In the example $\underline{X} - \underline{A}$ amounts to $\langle\{p_1, p_4\}, \{p_3\}\rangle$. As an aside, I would like to note that from the AGM-perspective the present kind of contraction is rather elementary. AGM-contraction is based on the so-called *remainder set*. In terms of sentences, the remainder set of a belief set or belief base after contraction by a sentence is the set of maximal subsets that do not imply that sentence. Such subsets are called *remainders* (Hansson 1999, 2011). Note that for belief base contraction it is not sufficient that these subsets do not contain that sentence, for the they might still (jointly) imply it. Belief contraction becomes interesting (and complicated) when the remainder set contains more than one remainder. In this case the general form of so-called *partial meet contraction* is defined as the intersection of the sets in some subset of the remainder set, to be selected by some selection function. In our present context of contraction the relevant remainder set is already unique, due to the fact that we are assuming that all possible b-claims are logically independent and that we are only considering contraction by (sets of) b-claims.

Finally, we turn to (belief base) revision (in the narrow sense). As announced, we base our definition on the relevant form of Levi's identity. That is, in the revision of

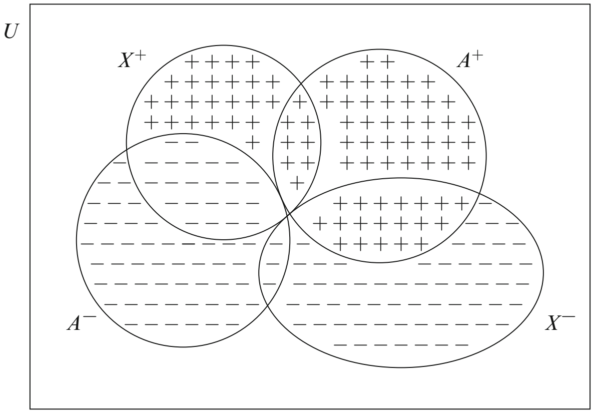


Fig. 5.7 The revision of $\underline{X}$ by $\underline{A}$

$\underline{X}$ by $\underline{A}$ there are two steps: first the b-claims of $\underline{X}$ that are in conflict with those of $\underline{A}$ are deleted, and second all b-claims of $\underline{A}$ are added. The first step amounts to the contraction of $\underline{X}$ by the specular of $\underline{A}$, sp($\underline{A}$), after which expansion of the result, $\underline{X}-\text{sp}(\underline{A})$, by $\underline{A}$ follows. Formally: r*evision* of $\underline{X}$ by $\underline{A}$, denoted by $\underline{X} * \underline{A}$, is defined by $[\underline{X} - \text{sp}(\underline{A})] + \underline{A}$. It is now easy to check that $\underline{X} * \underline{A}$ is equal to the expansion of the excess part of $\underline{X}$ relative to $\underline{A}$ by $\underline{A}$ or, equivalently, the expansion of $\underline{A}$ by the excess part of $\underline{X}$ relative to $\underline{A}$. Again, in Fig. 5.7, the positive and the negative range of the resulting theory have been indicated, representing in total all b-claims of $\underline{A}$ and all excess b-claims of $\underline{X}$ relative to $\underline{A}$.

Recall the example $\underline{X} = \langle\{p_1, p_2, p_4\}, \{p_3\}\rangle$ and $\underline{A} = \langle\{p_2, p_3\}, \{p_5\}\rangle$. Hence, $\text{sp}(\underline{A}) = \langle\{p_5\}, \{p_2, p_3\}\rangle$, which leads to $\underline{X} - \text{sp}(\underline{A}) = \langle\{p_1, p_2, p_4\}, \emptyset\rangle$ and hence to $[\underline{X} - \text{sp}(\underline{A})] + \underline{A} = \langle\{p_1, p_2, p_3, p_4\}, \{p_5\}\rangle$.

For the same reason as in the case of contraction, the present form of revision is a rather special case from the AGM-perspective. Due to the uniqueness of contraction and the rest of the Levi-identity form of our definition, revision is also unique. In line with calling truth approximation in terms of (comparing) sets of basic claims (Kuipers 2000), it is plausible to call belief base (expansion, contraction and) revision in terms of sets of basic claims also *basic*.

5.4 Truth Approximation by Data Based Theory Revision

Now we can start to dovetail theory revision with truth approximation. First we have to represent empirical data as they may come in by experiments or otherwise. It is important to distinguish two types of possible data, depending on the relevant interpretation as already alluded to in the introduction. In both cases we can represent data as theories.

Symmetric data can come in when, for example, in the propositional interpretation it is possible to verify whether some p is true or not, depending on what is in fact the case. A specific example of this arises in the partition interpretation, for example, when it can be verified whether a state is in equilibrium or not. In general, we will speak of (sets of) realized examples, denoted by R^+, and (sets of) realized non-examples, denoted by R^-. In combination we get a *symmetric data-theory* $\underline{R} = \langle R^+, R^- \rangle$, assuming that (examples and) non-examples can be realized. In the following we will assume that the data are correct. For later purposes we introduce the *Correct Data (CD-)hypothesis*, i.e. all b-claims of $\underline{R}$ are true or, simply, $\underline{R}$ is true, which amounts in the symmetric case to: $R^+ \subseteq \mathbf{T}^+, R^- \subseteq \mathbf{T}^-$, i.e. $R^+ \subseteq \mathbf{T}^+ \subseteq \mathrm{c}R^-$(recall: $\mathbf{T}^- = \mathrm{c}\mathbf{T}^+$).

Asymmetric data arise when non-examples cannot be realized in some direct sense by the nature of the interpretation, such as physical impossibilities, non-existence claims, etc. In this case we may get in due course a set of realized examples R and on their basis we may make inductive jumps to laws, claiming, for the time being, that certain conceptual possibilities cannot exist by the nature of physical reality or that certain types of individuals do not occur in the relevant universe of discourse. Note that it is plausible to assume that R arises, at least partly, by testing general hypotheses and hence that the inductive steps presuppose this kind of serious testing. Having induced laws implies having an indirectly induced strongest law, to be obtained by conjunction. It is easy to see that a law, and hence the strongest one, can set-theoretically be represented by a subset S of U with the claim that $\mathbf{T}^+$ is a subset of S or, equivalently, that $\mathrm{c}S$ is a subset of $\mathbf{T}^-$. In combination we now get an *asymmetric data-theory* $\underline{R} = \langle R, \mathrm{c}S \rangle$, in particular for nomic and monadic-existential interpretations. In the asymmetric case the Correct Data (CD-)hypothesis, i.e. $\underline{R}$ is true, amounts to: $R \subseteq \mathbf{T}^+, \mathrm{c}S \subseteq \mathbf{T}^-$, i.e. $R \subseteq \mathbf{T}^+ \subseteq S$. Of course, whereas symmetric data may contain descriptive mistakes, asymmetric data may contain descriptive as well as inductive mistakes. Hence, in the asymmetric case the CD-hypothesis is much more demanding than in the symmetric case.

We start with the revision of theories by symmetric data. The revision of theory $\underline{X}$ by symmetric data theory $\underline{R} = \langle R^+, R^- \rangle$, $\underline{X} * \underline{R}$, is by definition: $[\underline{X} - \mathrm{sp}(R)] + \underline{R}$, that is, $[\langle X^+, X^- \rangle - \langle R^-, R^+ \rangle] + \langle R^+, R^- \rangle$, which leads via $\langle X^+ - R^-, X^- - R^+ \rangle + \langle R^+, R^- \rangle$ to $\langle (X^+ - R^-) \cup R^+, (X^- - R^+) \cup R^- \rangle$. Now it is easy to prove the following:

Theorem 2. *Assuming the CD-hypothesis for symmetric data $\underline{R}$, $\underline{X} * \underline{R}$ is (qualitatively) at least as close to the truth as $\underline{X}$, i.e., $\underline{X} \leq_{vs} \underline{X} * \underline{R}$.*

For the proof we will first compare $\langle (X^+ - R^-) \cup R^+, (X^- - R^+) \cup R^- \rangle$ with $\langle X^+, X^- \rangle$. In this comparison it is easy to see that all established false b-claims of $\underline{X}$ are deleted and that the established (excess) true b-claims are added. In more detail, without 'established': the false positive b-claims of $\underline{X}$ are deleted and true positive b-claims of $\underline{R}$ are added and the false negative b-claims of $\underline{X}$ are deleted and true negative b-claims of $\underline{R}$ are added. For the proof in the strict sense it is only important to observe that only false b-claims of $\underline{X}$, if any, are deleted and that only true b-claims, if any, are added. In other words, that the falsity content of the result does not increase and the truth content of the result does not decrease, in accordance with the definition of 'at least as close to'.

Let us now turn to the revision of theories by asymmetric data. Of course, formally we can formulate and prove the relevant theorem by just replacing above ‘R^+’ by ‘R’ and ‘R^-’ by ‘cS’. However, for the sake of transparency, I will also present the resulting asymmetric story. The revision of theory $\underline{X}$ by asymmetric data theory $\underline{R} = \langle R, \mathrm{c}S \rangle$, $\underline{X} * \underline{R}$, is by definition: $[\underline{X} - \mathrm{sp}(\underline{R})] + \underline{R}$, that is, $[\langle X^+, X^- \rangle - \langle \mathrm{c}S, R \rangle] + \langle R, \mathrm{c}S \rangle$, which leads via $\langle X^+ - \mathrm{c}S, X^- - R \rangle + \langle R, \mathrm{c}S \rangle$ to $\langle (X^+ - \mathrm{c}S) \cup R, (X^- - R) \cup \mathrm{c}S \rangle$. Again it is easy to prove the following theorem:

Theorem 3. *Assuming the CD-hypothesis for asymmetric data $\underline{R}$, $\underline{X} * \underline{R}$ is (qualitatively) at least as close to the truth as $\underline{X}$, i.e., $\underline{X} \leq_{\mathrm{vs}} \underline{X} * \underline{R}$.*

By comparing $\langle (X^+ - \mathrm{c}S) \cup R, (X^- - R) \cup \mathrm{c}S \rangle$ with $\langle X^+, X^- \rangle$ it is again easy to see that only, even all, established false b-claims are deleted and that only, even all, established (excess) true b-claims are added, where the two times ‘only’, even if restricted to ‘only, if any’, are already sufficient to prove the ‘at least as close to’- claim. A slight transformation of the resulting revision, viz. $\langle (X^+ \cap S) \cup R, (X^- \cap \mathrm{c}R) \cup \mathrm{c}S \rangle$, enables to paraphrase the comparison with $\langle X^+, X^- \rangle$ by: the positive b-claims are restricted to those compatible with S and extended with those about R, and the negative b-claims are restricted to those compatible with the complement of R and extended with those implied by S, that is, those compatible with the complement of S. Note, that the resulting revision $\langle (X^+ \cap S) \cup R, (X^- \cap \mathrm{c}R) \cup \mathrm{c}S \rangle$ can be further transformed (using $R \subseteq S$) in terms of the laws claimed by $\underline{X}$, viz. “$\mathbf{T}^+ \subseteq \mathrm{c}X^-$”, leading to $\langle (X^+ \cap S) \cup R, (\mathrm{c}X^- \cap \mathrm{c}R) \cup \mathrm{c}S \rangle = \langle (X^+ \cap S) \cup R, \mathrm{c}(\mathrm{c}X - \cup R) \cup \mathrm{c}S \rangle = \langle (X^+ \cap S) \cup R, \mathrm{c}((\mathrm{c}X^- \cup R) \cap S) \rangle = \langle (X^+ \cup R) \cap S, \mathrm{c}((\mathrm{c}X^- \cap S) \cup R) \rangle$. Hence, the law-claim of the resulting revision amounts to “$\mathbf{T}^+ \subseteq (\mathrm{c}X^- \cap S) \cup R$”.

Finally, it is easy to see that the two theorems can be strengthened to the claims that the revised theory is even closer to the truth, if we assume that $\underline{R}$ is not just a subtheory of $\underline{X}$ in the sense that $R^+ \subseteq X^+$ and $R^- \subseteq X^-$ in the symmetric case and $R \subseteq X^-$ and $\mathrm{c}S \subseteq X^-$ in the asymmetric case, respectively. For by the contraction step some false b-claims of $\underline{X}$ will be deleted and/or by the final expansion the revised theory will get extra true b-claims relative to $\underline{X}$.

5.5 Generalized Feature-Contrast Measure of Verisimilitude

It is not difficult to present a quantitative version of the main points of Sect. 5.4. We will first define a measure for the verisimilitude of a theory and then show that, and to what extent, the verisimilitude of the revision by data, assumed to be true, has increased.

Let m be a so-called normalized countably additive real-valued measure function on the set of measurable subsets of U, that is, for all subsets V, W of U : $m(\emptyset) = 0 \leq m(V) \leq 1 = m(U)$ and $m(V \cup W) = m(V) + m(W) - m(V \cap W)$.

We define the following degrees of positive/negative truth/falsity content:

degree of positive truth content $\qquad \mathrm{cont}_{\mathrm{t}}^+(\underline{X}) = m(\mathrm{t}^+(X)) = m(X^+ \cap \mathbf{T}^+)$

degree of negative truth content $\mathrm{cont}_{\mathrm{t}}^{-}(\underline{X}) = m(\mathrm{t}^{-}(X)) = m(X^{-} \cap \mathbf{T}^{-})$
degree of positive falsity content $\mathrm{cont}_{\mathrm{f}}^{+}(\underline{X}) = m(\mathrm{f}^{+}(X)) = m(X^{+} \cap \mathbf{T}^{-})$
$= m(X^{+} - \mathbf{T}^{+})$
degree of negative falsity content $\mathrm{cont}_{\mathrm{f}}^{-}(\underline{X}) = m(\mathrm{f}^{-}(X)) = m(X^{-} \cap \mathbf{T}^{+})$
$= m(X^{-} - \mathbf{T}^{-})$

In order to take different absolute weights of truth and falsity and relative weights of true (false) positives and negatives into account, we introduce three positive real-valued parameters, viz. φ, γ, and δ, all positive, and define

Definition 2. The degree of verisimilitude of $\underline{X}$: $\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X})$

$$
\begin{aligned}
&=_{\mathrm{df}} [\mathrm{cont}_{\mathrm{t}}^{+}(\underline{X}) + \gamma \mathrm{cont}_{\mathrm{t}}^{-}(\underline{X})] - \varphi[\mathrm{cont}_{\mathrm{f}}^{+}(\underline{X}) + \delta \mathrm{cont}_{\mathrm{f}}^{-}(\underline{X})] \\
&= \quad [m(X^{+} \cap \mathbf{T}^{+}) + \gamma m(X^{-} \cap \mathbf{T}^{-})] - \varphi[m(X^{+} - \mathbf{T}^{+}) + \delta m(X^{-} - \mathbf{T}^{-})]
\end{aligned}
$$

Here φ represents the relative weight of truth and falsity and γ and δ represent the relative weight of true positives and negatives and false positives and negatives, respectively. The parameters γ, δ may or may not depend on $m(\mathbf{T}^{+})$ and $m(\mathbf{T}^{-})$, respectively, or they may be related in some other way. The reason to claim that they may depend on $m(\mathbf{T}^{+})$ and $m(\mathbf{T}^{-})$ is that, for example, if (one may assume that) $m(\mathbf{T}^{+})$ is much smaller than $m(\mathbf{T}^{-})$ it seems reasonable that true (false) positives are valued higher than true (false) negatives. For this reason the parameters could even both be set equal to $m(\mathbf{T}^{+})/m(\mathbf{T}^{-})$. To be sure, by making the weights dependent on $m(\mathbf{T}^{+})$ and $m(\mathbf{T}^{-})$, it becomes impossible to calculate the increase of verisimilitude of a revised theory (see below) in the standard situation in which we don't know $\underline{\mathbf{T}} = \langle \mathbf{T}^{+}, \mathbf{T}^{-} \rangle$.

Other kinds of special cases we get by putting $\varphi = 1$ or by putting $\gamma = \delta$ or even further $\gamma = \delta = 1$, the latter specification leading to

$$
\mathrm{Vs}_{\varphi}^{1,1}(\underline{X}) =_{\mathrm{df}} \mathrm{Vs}_{\varphi}(\underline{X}) = m(\mathrm{t}(\underline{X})) - \varphi m(\mathrm{f}(\underline{X}))
$$

which is the definition given in Cevolani et al. (2011), for which reason our general definition is called a 'generalized' measure. The authors call $\mathrm{Vs}_{\varphi}^{1,1}$ a 'feature-contrast' measure, because they relate it to a standard statistical practice of feature-contrast representation. Feature-contrast measures of verisimilitude have been introduced (without using this name) by Cevolani and Calandra (2009). See Festa and Cevolani for an overview of feature-contrast measures.

Of course, the definition of the quantitative measure would be inadequate when it would not cohere with the qualitative definition in Sect. 5.2. However, it is not difficult to prove the

Theorem 4. *If* $\underline{X} \leq_{VS} \underline{Y}$ *then for all* $\varphi, \gamma, \delta > 0 : \mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X}) \leq \mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{Y})$

This theorem automatically implies that the verisimilitude of the revision of $\underline{X}$ by true data $\underline{R}$, viz. $\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X} * \underline{R})$, is higher (or at least equal to) than that of $\underline{X}$, viz. $\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X})$, because the qualitative relation holds.

The increase of verisimilitude of the revision of $\underline{X}$ by true data $\underline{R}$ can also be expressed. Recall:

$$\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X}) = [m(X^+ \cap \mathbf{T}^+) + \gamma m(X^- \cap \mathbf{T}^-)] - \varphi[m(X^+ - \mathbf{T}^+) + \delta m(X^- - \mathbf{T}^-)]$$

Hence, in the symmetrical representation, $\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X} * \underline{R}) =$

$$\begin{aligned} &[m(((X^+ - R^-) \cup R^+) \cap \mathbf{T}^+) + \gamma m((X^- - R^+) \cup R^-) \cap \mathbf{T}^-)] \\ &\qquad - \varphi[m((X^+ - R^-) \cup R^+) - \mathbf{T}^+) + \delta m((X^- - R^+) \cup R^-) - \mathbf{T}^-)] \end{aligned}$$

It is not difficult to sort out the components of the four *m*easure expressions

$$\begin{aligned} m(((X^+ - R^-) \cup R^+) \cap \mathbf{T}^+) &= m(X^+ \cap \mathbf{T}^+) + m(R^+ - X^+) \\ m(((X^- - R^+) \cup R^-) \cap \mathbf{T}^-) &= m(X^- \cap \mathbf{T}^-) + m(R^- - X^-) \\ m(((X^+ - R^-) \cup R^+) \cap \mathbf{T}^+) &= m(X^+ \cap \mathbf{T}^+) + m(X^+ - R^-) \\ m(((X^- - R^+) \cup R^-) \cap \mathbf{T}^-) &= m(X^- \cap \mathbf{T}^-) + m(X^- - R^+) \end{aligned}$$

Now it is easy to express the increase of verisimilitude, in the symmetric as well as in the asymmetric representation (the latter again by just replacing 'R^+' by 'R' and 'R^-' by 'cS'), that is, it is easy to prove the following:

Theorem 5. $\mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X} * \underline{R}) - \mathrm{Vs}_{\varphi}^{\gamma,\delta}(\underline{X}) =$

$$\begin{aligned} \textit{symmetric} &: [m(R^+ \cap X^+) + \gamma m(R^- - X^-)] - \varphi[m(X^+ - R^-) + \delta m(X^- - R^+)] \\ \textit{asymmetric} &: [m(R \cap X^+) + \gamma m(\mathrm{c}S - X^-)] - \varphi[m(X^+ - \mathrm{c}S) + \delta m(X^- - R)] \end{aligned}$$

Assuming that the parameters do not depend on $\mathbf{T}$, it is interesting to note that the increase of verisimilitude can be calculated on the basis of $\underline{X}$ and $\underline{R}$ alone.

5.6 Concluding Remarks

We have presented a very general approach, with several interpretations, to the true boundary within some universe of discourse by belief base revision, guided by input data of symmetric or asymmetric nature. Section 5.2 provided the qualitative definition for the statement "theory $\underline{Y}$ is at least as close to the truth/the true boundary as theory $\underline{X}$", and Sect. 5.3 the definition of the statement "the revision $\underline{X} * \underline{A}$ of theory $\underline{X}$ in the light of input theory $\underline{A}$". In Sect. 5.4 followed the theorem for any input data (theory) $\underline{R}$ of symmetric or asymmetric nature: "if $\underline{R}$ is true/correct, then $\underline{X} * \underline{R}$ is closer to the truth than $\underline{X}$". In Sect. 5.5 the main lines of a quantitative version of this account were given.

There are some plausible challenges for further research. To begin with, the generalization from a divide, i.e. a bivalent partition, to a (finite) partition in general does not seem to be difficult in principle, although it seems technically rather complicated. More interesting is the following challenge. Festa and Cevolani, deal with quantitative measures of verisimilitude for constituent languages as well as for quantitative languages, where the latter incorporate quantitative, statistical and tendency hypotheses. Whereas our qualitative account essentially covers constituent languages of all kinds, the interesting question is whether this account, or its quantitative version, can be adapted to quantitative languages.

Another challenge to cope with is the problem in the basic qualitative account in Sect. 5.4 that, for example, in the nomic interpretation one or more of the established laws and hence the strongest law S, may not only be incompatible with X^{+} in the weak sense that X^{+} overlaps with cS , but even in the strong sense that S does not overlap with X^{+} at all, that is, $X^{+} \subseteq \mathrm{c}S$, in which case the revision of $\underline{X}$ by $\underline{R} = \langle R, \mathrm{c}S\rangle$, i.e. $\langle (X^{+} \cap S) \cup R, (X^{-} \cap \mathrm{c}R) \cup \mathrm{c}S\rangle$, is just $\langle R, (X^{-} - R) \cup \mathrm{c}S\rangle$. Hence, any feature or trace of X^{+} has disappeared. Formally similar, but in the nomic interpretation perhaps less frequently in practice, when $X^{-} \subseteq R$, the revision reduces to $\langle (X^{+} - \mathrm{c}S) \cup R, \mathrm{c}S\rangle$, and nothing is left of X^{-}. In a previous attempt (Kuipers 2011) to dovetail belief revision and truth approximation, restricted to the nomic interpretation and to maximal theories, I succeeded in overcoming the corresponding problem by taking refined forms of belief revision into account, notably partial meet revision. However, that dovetail attempt was unsatisfactory for other reasons already in its basic form. In a quite ad hoc way it dealt first with the (two-step) revision of the law or necessity claim ("$\mathbf{T}^{+} \subseteq X^{+}$" or, equivalently, "c$X+ \subseteq \mathbf{T}^{-}$") of a maximal theory $\underline{X}$ (hence with $X^{-} = \mathrm{c}X^{+}$), and just added at the end the relevant sufficiency claim ("$\ldots \subseteq \mathbf{T}^{+}$").[1] Our new 'two-sided' approach is, besides being much more general, not in need of this ad hoc move and may well enable a refined way of dealing with theory revision by strongly incompatible laws, in the first case, in a similar way as in the previous paper, which was in itself quite satisfactory. But I leave this for another occasion.

A further challenge is to generalize the present approach from conjunctive theories to theories in general, where the latter can be seen as disjunctions of constituents of the relevant kind, e.g. propositional constituents. In this sense, theories in general are disjunctive theories. As a matter of fact, conjunctive theories are disjunctive theories of a special kind. In terms of propositional constituents, for convenience's sake, a conjunctive theory is the disjunction of all constituents that fully agree for a certain subset of atomic propositions, but differ in all possible ways about all the other. In the generalization to (propositional) theories in general, the

[1] *Note added in proof.* Later I found out that this ad hoc move need not be made to enable the application of the basic and refined truth approximation analysis. The necessity claim is enough. See: Kuipers, T., "Empirical Progress and Truth Approximation Revisited", Proceedings Tilburg conference on *Progress in Science* (April, 2012), Studies in History and Philosophy of Science, special section *Progress in Science*, 46, 2014, 64–72.

truth-value of some atomic propositions may depend on that of other, not on logical grounds, but because the theory requires so. It may well be that such a generalization is only possible in quantitative terms.

I ended my previous paper on the subject with a number of debunking remarks (Kuipers 2011). Besides overcoming the indicated ad hocness charge,[2] I am afraid the other remarks remain valid. So I adapt them for the present paper and restrict them to the case of asymmetric data:

(a) The revision is rather diehard empiricist or instrumentalist. The ‘instrument’ $\underline{X}$ is precisely so adapted that it just saves the phenomena, not only with respect to individual cases (R), but also with respect to empirical generalizations (S).

(b) If there is something like well-formed theories, satisfying certain criteria, e.g. of formalizability and generality, there do not seem to be good reasons to expect that the revision of a theory that satisfies them, will also satisfy these criteria, even if R and S satisfy some derived criteria.

(c) Last, but not least, what remains of the idea behind $\underline{X}$? A proper theory, in some sophisticated sense even if without theoretical terms, is usually based on one or two ideas. It is difficult to imagine that such ideas do not become ‘mutilated’ by the revision.

Acknowledgements I am very grateful for the stimulating interaction with Gustavo Cevolani and Roberto Festa, and for the comments I got at conferences in Gent (Logic, Reasoning, and Rationality) and Padua (Science, Truth, and Ontology) in September 2010. Finally, I am grateful to the two anonymous referees for their scrupulous remarks and a number of interesting suggestions.

References

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change. *Journal of Symbolic Logic, 50*, 510–530.

Cevolani, G., & Calandra, F. (2009). Approaching the truth via belief change in propositional languages. In: M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSA epistemology and methodology of science*. Launch of the European philosophy of science association (vol. 1, pp. 47–62). Berlin: Springer.

Cevolani, G., Crupi, V., & Festa, R. (2011). Verisimilitude and belief change for conjunctive theories. *Erkenntnis, 75*(2), 183–222.

Cevolani, G., Festa, R., & Kuipers, T. (2013). Verisimilitude and belief change for nomic conjunctive theories. *Synthese, 190*, 3307–3324.

Festa, R. (2007). Verisimilitude, qualitative theories, and statistical inferences. In: S. Philström, P. Raatikainen, & M. M. Sintonen (Eds.), *Approaching truth: Essays in honour of Ilkka Niiniluoto* (pp. 143–178). London: College Publications.

Festa, R., & Cevolani, G. Features of verisimilitude. Manuscript.

Hansson, S. O. (1999). *A textbook of belief dynamics*. Dordrecht: Kluwer.

Hansson, S. O. (2011). Logic of belief revision. In: E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Http://plato.stanford.edu/

[2] *Note added in proof.* See the previous note.

Kuipers, T. (1982). Approaching descriptive and theoretical truth. *Erkenntnis, 18*(3), 343–378.

Kuipers, T. (2000). *From instrumentalism to constructive realism: On some relations between confirmation, empirical progress, and truth approximation* (Synthese library, vol. 287). Dordrecht: Kluwer.

Kuipers, T. (2011). Basic and refined nomic truth approximation by evidence-guided belief revision in AGM-terms. *Erkenntnis, 75*(2), 223–236.

Niiniluoto, I. (1999). Belief revision and truthlikeness. In: B. Hansson, S. Halldén, N. E. Sahlin, & W. Rabinowicz (Eds.), Internet Festschrift for Peter Gärdenfors. Lund: Department of Philosophy, Lund University. http://www.lucs.lu.se/spinning/

Chapter 6
A Method of Generating Modal Logics Defining Jaśkowski's Discussive $\mathbf{D_2}$ Consequence

Marek Nasieniewski and Andrzej Pietruszczak

6.1 Introduction

Jaśkowski's discussive logic $\mathbf{D_2}$ can be treated either as some set of discussive formulae or as a consequence relation on the set of all discussive formulae.

In the first case $\mathbf{D_2}$ is formulated with the help of the modal logic[1] **S5** as follows (see Jaśkowski 1999a,b): $A \in \mathbf{D_2}$ iff $\ulcorner \Diamond A^{\bullet} \urcorner \in \mathbf{S5}$, where $(-)^{\bullet}$ is a translation of discussive formulae into the modal language (see Sects. 6.2.1 and 6.2.2).

We say that a modal logic $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ iff $\mathbf{D_2} = \{A \in \mathrm{For^d} : \ulcorner \Diamond A^{\bullet} \urcorner \in \boldsymbol{L}\}$, where $\mathrm{For^d}$ is the set of all discussive formulae. The papers Nasieniewski and Pietruszczak (2008, 2009) and Perzanowski (1975) present respectively the weakest regular and the weakest normal logic which

has the same theses beginning with '$\Diamond$' as **S5**. (†)

In Nasieniewski and Pietruszczak (2011) it was shown that for all modal logics which are closed under replacement of tautological equivalents (rte-logics): defining $\mathbf{D_2}$ is equivalent to having the property (†). Moreover, in Nasieniewski and Pietruszczak (2011) a general method that produces the weakest logic having the property (†) for various classes of logics was given. For example, for rte-logics, congruential, monotonic, regular and normal we obtain the weakest in a given class logic defining $\mathbf{D_2}$.

[1] The term 'modal logic' is here always understood as a set of modal formulae. In Appendix we recall the standard notation, chosen basic facts and notions concerning modal logics.

M. Nasieniewski (✉) • A. Pietruszczak
Department of Logic, Nicolaus Copernicus University, Toruń, Poland
e-mail: mnasien@uni.torun.pl; pietrusz@uni.torun.pl

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_6,
© Springer Science+Business Media Dordrecht 2014

In the current paper we mainly consider the second way of understanding the logic $\mathbf{D_2}$, i.e., as a consequence relation between sets of discussive formulae and single discussive formulae (strictly speaking, as discussive systems based on this relation). Now we also use the modal logic **S5** and put: $\Pi \vdash_{\mathbf{D_2}} B$ iff $\{\ulcorner \Diamond A^{\bullet} \urcorner : A \in \Pi\} \vdash_{\mathbf{S5}} \Diamond B^{\bullet}$, where $\vdash_{\mathbf{S5}}$ is the pure modus-ponens-style inference relation based on **S5** (see Sect. 6.2.3). The basic features of discussive systems used by Jaśkowski are presented by himself in Jaśkowski (1948, p. 61) (see also Jaśkowski 1999a, p. 37–38). The consequence relation considered by Jaśkowski is called the $\mathbf{D_2}$-*consequence*. It has been elaborated in the paper Nasieniewski and Pietruszczak (2013).

Following Jaśkowski, we say that a subset $\boldsymbol{S}$ of $\mathrm{For^d}$ is a *discussive system* iff $\boldsymbol{S}$ is closed under the $\mathbf{D_2}$-consequence, i.e., for any $\Pi \subseteq \mathrm{For^d}$ and $B \in \mathrm{For^d}$, if $\Pi \vdash_{\mathbf{D_2}} B$ and $\Pi \subseteq \boldsymbol{S}$, then $B \in \boldsymbol{S}$. Of course, all discussive systems include the set $\mathbf{D_2}$ (see Sect. 6.2.4).

The main aim of the present paper is to find a general method of defining modal logics which also allow to define the $\mathbf{D_2}$-consequence. This paper is a continuation of the results from Nasieniewski and Pietruszczak (2011, 2013).[2] This perspective is important since it appears that properties of the general method are in a correlation with features of the very $\mathbf{D_2}$-consequence. The new results presented as well as these given in the mentioned publications, prove that the use of **S5** in the definition of the $\mathbf{D_2}$-consequence is inessential. We show that in this definition weaker logics can be used.

One can easily see that for a logic $\boldsymbol{L}$ to define the $\mathbf{D_2}$-consequence it is enough to have the following property:

$$\begin{gathered}\text{for any set } \Pi \text{ of modal formulae and any modal formula } B,\\ \Diamond\Pi \vdash_{\boldsymbol{L}} \Diamond B \text{ iff } \Diamond\, \Pi \vdash_{\mathbf{S5}} \Diamond B,\end{gathered} \tag{$\ddagger$}$$

where $\vdash_{\boldsymbol{L}}$ is the pure modus-ponens-style inference relation based on $\boldsymbol{L}$. Among others, we show in this paper that for all rte-logics: defining the $\mathbf{D_2}$-consequence is equivalent to having the property (‡).

We give a general method which, for some classes of modal logics determined by sets of joint axioms and rules, generates in the given class the weakest logic having the above property (‡). In particular, for the class of all modal logics we obtain the weakest modal logic which owns this property. Moreover, applying the method to various classes of modal logics: rte-logics, congruential, monotonic, regular and normal, we obtain the weakest in a given class logic defining the $\mathbf{D_2}$-consequence.

The presented method of generating of axiomatizations of modal logics applied to particular classes of logics, reveals a nature of the $\mathbf{D_2}$-consequence; applying the

[2] In Nasieniewski and Pietruszczak (2013) the $\mathbf{D_2}$-consequence was defined and the smallest normal and regular modal logics defining the $\mathbf{D_2}$-consequence were given (see Sect. 6.3). In Nasieniewski and Pietruszczak (2011) some general method of axiomatizing modal logics defining the logic $\mathbf{D_2}$ was proposed. In the current paper this method has been adopted for the case of the $\mathbf{D_2}$-consequence.

method to broader and broader classes containing weaker logics we obtain modal logics weaker than **S5**, which contain less and less of in a sense inessential theses but still define the $\mathbf{D_2}$-consequence.

6.2 Basic Notions

6.2.1 Discussive Formulae and a Transformation

Discussive formulae are formed in the standard way from propositional letters: 'p', 'q', 'p_0', 'p_1', 'p_2', $\ldots$; truth-value operators: '$\neg$' and '$\vee$' (negation and disjunction); discussive connectives: '$\wedge^{\mathrm{d}}$', '$\rightarrow^{\mathrm{d}}$', '$\leftrightarrow^{\mathrm{d}}$' (conjunction, implication and equivalence); and brackets. Let $\mathrm{For}^{\mathrm{d}}$ be the set of all these formulae.

Let $\mathrm{For}_{\mathrm{m}}$ be the set of all modal formulae (see Appendix) and let $(-)^{\bullet}$ be the translation from $\mathrm{For}^{\mathrm{d}}$ into $\mathrm{For}_{\mathrm{m}}$ such that:

1. $(a)^{\bullet} = a$, for any propositional letter a,
2. For any $A, B \in \mathrm{For}^{\mathrm{d}}$:

$$\begin{aligned}
&(\neg A)^{\bullet} = \ulcorner \neg A^{\bullet} \urcorner,\\
&(A \vee B)^{\bullet} = \ulcorner A^{\bullet} \vee B^{\bullet} \urcorner,\\
&(A \wedge^{\mathrm{d}} B)^{\bullet} = \ulcorner A^{\bullet} \wedge \Diamond B^{\bullet} \urcorner,\\
&(A \rightarrow^{\mathrm{d}} B)^{\bullet} = \ulcorner \Diamond A^{\bullet} \rightarrow B^{\bullet} \urcorner,\\
&(A \leftrightarrow^{\mathrm{d}} B)^{\bullet} = \ulcorner (\Diamond A^{\bullet} \rightarrow B^{\bullet}) \wedge \Diamond(\Diamond B^{\bullet} \rightarrow A^{\bullet}) \urcorner.
\end{aligned}$$

6.2.2 The Discussive Logic $\mathbf{D_2}$ *(as a Set of Discussive Formulae)*

Jaśkowski used the notation '$\mathbf{D_2}$' referring to a certain set of discussive formulae. This set is a logic in the first sense of the two explained in the introduction.

Definition 1. The logic $\mathbf{D_2}$ was formulated as follows:

$$\mathbf{D_2} := \{ A \in \mathrm{For}^{\mathrm{d}} : \ulcorner \Diamond A^{\bullet} \urcorner \in \mathbf{S5} \},$$

Notice that by Lemma 19 we obtain:

Fact 1. *The following formulae*

$$p \rightarrow^{\mathrm{d}} p \tag{6.1}$$

$$p \rightarrow^{\mathrm{d}} (q \rightarrow^{\mathrm{d}} (p \wedge^{\mathrm{d}} q)) \tag{6.2}$$

belong to $\mathbf{D_2}$ *and the set* $\mathbf{D_2}$ *is closed under* modus ponens *for '$\rightarrow^{\mathrm{d}}$' (i.e., for any $A, B \in \mathrm{For}^{\mathrm{d}}$, if $A, \ulcorner A \rightarrow^{\mathrm{d}} B \urcorner \in \mathbf{D_2}$ then $B \in \mathbf{D_2}$).*

Definition 2. A modal logic $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ iff $\mathbf{D_2} = \{A \in \mathrm{For}^{\mathrm{d}} : \ulcorner \Diamond A^{\bullet} \urcorner \in \boldsymbol{L}\}$.

It is known that also other modal logics than **S5** define $\mathbf{D_2}$. Now notice that:

Lemma 1. *If a modal logic* $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ *and* (C) $\in \boldsymbol{L}$*,*

$$\Diamond (\Diamond p \to q) \to (\Diamond p \to \Diamond q) \tag{C}$$

then the following formula

$$\Diamond p \to (\Diamond q \to \Diamond (p \wedge \Diamond q)) \tag{6.3}$$

belongs to $\boldsymbol{L}$*.*

Proof. Let $\boldsymbol{L}$ define $\mathbf{D_2}$ and (C) $\in \boldsymbol{L}$. Then, '$\Diamond(\Diamond p \to (\Diamond q \to (p \wedge \Diamond q)))$' belongs to $\boldsymbol{L}$, by Lemma 19. Thus, we use Lemma 20. □

While expressing the logic $\mathbf{D_2}$ we refer to modal logics which have the same as **S5** theses beginning with '$\Diamond$'.

Definition 3. Let $\mathsf{S5}_{\diamond}$ be the set of all modal logics having the property (†) given in the introduction, i.e.,

$$\boldsymbol{L} \in \mathsf{S5}_{\diamond} \overset{\mathrm{df}}{\iff} \forall_{A \in \mathrm{For_m}} (\ulcorner \Diamond A \urcorner \in \boldsymbol{L} \iff \ulcorner \Diamond A \urcorner \in \mathbf{S5}).$$

By definitions we have:

Fact 2.

(i) *Every logic from* $\mathsf{S5}_{\diamond}$ *defines* $\mathbf{D_2}$*.*
(ii) *If* $\boldsymbol{L} \subseteq \boldsymbol{L}' \subseteq \mathbf{S5}$ *and* $\boldsymbol{L}$ *defines* $\mathbf{D_2}$*, then* $\boldsymbol{L}'$ *defines* $\mathbf{D_2}$*.*
(iii) *If* $\boldsymbol{L} \subseteq \boldsymbol{L}' \subseteq \mathbf{S5}$ *and* $\boldsymbol{L} \in \mathsf{S5}_{\diamond}$*, then* $\boldsymbol{L}' \in \mathsf{S5}_{\diamond}$*.*

Moreover, we know that:

Fact 3 (Nasieniewski and Pietruszczak 2011).

(i) *For any rte-logic* $\boldsymbol{L}$: $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ *iff* $\boldsymbol{L} \in \mathsf{S5}_{\diamond}$*.*
(ii) *If* $\boldsymbol{L} \in \mathsf{S5}_{\diamond}$ *is a congruential logic, then* $\boldsymbol{L} \subseteq \mathbf{S5}$*.*
(iii) *If* $\boldsymbol{L}$ *is a congruential logic defining* $\mathbf{D_2}$*, then* $\boldsymbol{L} \subseteq \mathbf{S5}$*.*

Since **S5** defines $\mathbf{D_2}$ and any intersection of modal logics defining $\mathbf{D_2}$ is a modal logic defining $\mathbf{D_2}$, we have:

Fact 4. *There exists the smallest modal logic defining* $\mathbf{D_2}$*.*

Let **A** be the smallest modal logic defining $\mathbf{D_2}$ ("absolute" one). This logic is examined in Nasieniewski and Pietruszczak (2012). It is shown there that **A** does not belong to $\mathsf{S5}_{\diamond}$. This logic has no member of the form $\ulcorner \Diamond\Diamond\, A \urcorner$. Hence, for example, for any $T \in \mathbf{PL}$, $\ulcorner \Diamond\Diamond\, T \urcorner \notin \mathbf{A}$, but $\ulcorner \Diamond\Diamond\, T \urcorner \in \mathbf{S5}$.

6.2.3 The Discussive the $\mathbf{D_2}$*-Consequence*

In Jaśkowski's paper Jaśkowski (1948) the following definition of a discussive relation can be discovered:

Definition 4. For any $\Pi \subseteq \mathrm{For}^{\mathrm{d}}$ and $B \in \mathrm{For}^{\mathrm{d}}$:

$$\Pi \vdash_{\mathbf{D_2}} B \overset{\mathrm{df}}{\Longleftrightarrow} \{\Diamond A^{\bullet} : A \in \Pi\} \vdash_{\mathbf{S5}} \Diamond B^{\bullet},$$

where $\vdash_{\mathbf{S5}}$ is the pure modus-ponens-style inference relation based on **S5**, i.e., there exists a sequence $C_1, \ldots, C_n = \ulcorner \Diamond B^{\bullet} \urcorner$ in which for any $i \leqslant n$, either $C_i \in \mathbf{S5}$, or $C_i \in \{\Diamond A^{\bullet} : A \in \Pi\}$, or there are $j, k < i$ such that $C_k = \ulcorner C_j \to C_i \urcorner$.[3]

On the basis of $\mathbf{D_2}$ one can characterize the consequence relation for discussive systems in the following way:

Fact 5 (Nasieniewski and Pietruszczak 2013). *For any* $n \geqslant 0$, $A_1, \ldots, A_n, B \in \mathrm{For}^{\mathrm{d}}$:

$$\begin{aligned}
A_1, \ldots, A_n \vdash_{\mathbf{D_2}} B \quad &\textit{iff} \quad \ulcorner (\Diamond A_1^{\bullet} \wedge \cdots \wedge \Diamond A_n^{\bullet}) \to \Diamond B^{\bullet} \urcorner \in \mathbf{S5} \\
&\textit{iff} \quad \ulcorner \Diamond A_1^{\bullet} \to (\ldots \to (\Diamond A_n^{\bullet} \to \Diamond B^{\bullet}) \ldots) \urcorner \in \mathbf{S5} \\
&\textit{iff} \quad \ulcorner \Diamond (A_1 \to^{\mathrm{d}} (\ldots \to^{\mathrm{d}} (A_n \to^{\mathrm{d}} B) \ldots))^{\bullet} \urcorner \in \mathbf{S5} \\
&\textit{iff} \quad \ulcorner A_1 \to^{\mathrm{d}} (\ldots \to^{\mathrm{d}} (A_n \to^{\mathrm{d}} B) \ldots) \urcorner \in \mathbf{D_2} \\
&\textit{iff} \quad \ulcorner (A_1 \wedge^{\mathrm{d}} \cdots \wedge^{\mathrm{d}} A_n) \to^{\mathrm{d}} B \urcorner \in \mathbf{D_2}\,.
\end{aligned}$$

Since (6.1) $\in \mathbf{D_2}$ (or (C), (6.3) $\in$ **S5**), by Fact 5 we obtain the following lemma:

Lemma 2. *For any* $A, B \in \mathrm{For}^{\mathrm{d}}$:

$$A \to^{\mathrm{d}} B, A \vdash_{\mathbf{D_2}} B,$$

$$A, B \vdash_{\mathbf{D_2}} A \wedge^{\mathrm{d}} B.$$

Moreover, by Fact 5 the relation of the consequence $\vdash_{\mathbf{D_2}}$ can be characterized with the help of modus ponens for '$\to^{\mathrm{d}}$' as the only rule of inference, i.e., $\vdash_{\mathbf{D_2}}$ is the pure modus-ponens-style inference relation based on $\mathbf{D_2}$.

[3] As it is known $\{\Diamond A^{\bullet} : A \in \Pi\} \vdash_{\mathbf{S5}} \Diamond B^{\bullet}$ iff for some $A_1, \ldots, A_n \in \Pi$, $n \geqslant 0$, we have that $\ulcorner \Diamond A_1^{\bullet} \to (\ldots \to (\Diamond A_n^{\bullet} \to \Diamond B^{\bullet}) \ldots) \urcorner \in \mathbf{S5}$, or equivalently $\ulcorner (\Diamond A_1^{\bullet} \wedge \cdots \wedge \Diamond A_n^{\bullet}) \to \Diamond B^{\bullet} \urcorner \in \mathbf{S5}$ (see Lemma 22).

Fact 6 (Nasieniewski and Pietruszczak 2013). *For any* $\Pi \subseteq \mathrm{For}^{\mathrm{d}}$ *and* $B \in \mathrm{For}^{\mathrm{d}}$:

$\Pi \vdash_{\mathbf{D_2}} B$ *iff* *there exists a sequence* $A_1, \ldots, A_n = B$ *in which for any* $i \leqslant n$, *either* $A_i \in \Pi \cup \mathbf{D_2}$ *or there are* $j, k < i$ *such that* $A_k = \ulcorner A_j \to^{\mathrm{d}} A_i \urcorner$.

Definition 5. A modal logic $\boldsymbol{L}$ *defines* the $\mathbf{D_2}$-consequence iff for any set $\Pi \subseteq \mathrm{For}^{\mathrm{d}}$ and $B \in \mathrm{For}^{\mathrm{d}}$:

$$\Pi \vdash_{\mathbf{D_2}} B \quad \text{iff} \quad \{\Diamond A^{\bullet} : A \in \Pi\} \vdash_{\boldsymbol{L}} \Diamond B^{\bullet} \tag{$\star$}$$

where $\vdash_{\boldsymbol{L}}$ is the pure modus-ponens-style inference relation based on $\boldsymbol{L}$, i.e., there exists a sequence $C_1, \ldots, C_n = \ulcorner \Diamond B^{\bullet} \urcorner$ in which for any $i \leqslant n$, either $C_i \in \boldsymbol{L} \cup \{\Diamond A^{\bullet} : A \in \Pi\}$, or there are $j, k < i$ such that $C_k = \ulcorner C_j \to C_i \urcorner$ (see Lemma 22 and Footnote 3).

By Lemma 22 we obtain:

Lemma 3. *If* $\boldsymbol{L} \subseteq \boldsymbol{L}' \subseteq \mathbf{S5}$ *and* $\boldsymbol{L}$ *defines the* $\mathbf{D_2}$*-consequence, then also* $\boldsymbol{L}'$ *defines the* $\mathbf{D_2}$*-consequence.*

Remark 1. Let us consider the conditions which ought to be fulfilled by a modal logic $\boldsymbol{L}$, meant to define the $\mathbf{D_2}$-consequence. To this aim let us follow through the condition ($\star$).

Let us assume that $\Pi \vdash_{\mathbf{D_2}} B$. Then, by Fact 6, there exists a sequence $A_1, \ldots, A_n = B$ of formulae from $\mathrm{For}^{\mathrm{d}}$ in which for any $i \leqslant n$, either $A_i \in \Pi \cup \mathbf{D_2}$ or there are $j, k < i$ such that $A_k = \ulcorner A_j \to^{\mathrm{d}} A_i \urcorner$.

We transform the above sequence into the modal language. We obtain a sequence $\ulcorner \Diamond A_1^{\bullet} \urcorner, \ldots, \ulcorner \Diamond A_n^{\bullet} \urcorner = \ulcorner \Diamond B^{\bullet} \urcorner$ of formulae from $\mathrm{For_m}$, where either $A_i \in \Pi$, or $\ulcorner \Diamond A_i^{\bullet} \urcorner \in \mathbf{S5}$ (iff $A_i \in \mathbf{D_2}$), or for some $j, k < i$: $\ulcorner \Diamond A_k^{\bullet} \urcorner = \ulcorner \Diamond (A_j \to^{\mathrm{d}} A_i)^{\bullet} \urcorner = \ulcorner \Diamond (\Diamond A_j^{\bullet} \to A_i^{\bullet}) \urcorner$.

We see that the last sequence is not a pure modus-ponens-style inference based on $\boldsymbol{L}$. To obtain such an inference we have to appropriately supplement the sequence with:

(a) The adequate instances of the formula (C), i.e., some formulae of the form $\ulcorner \Diamond (\Diamond C \to D) \to (\Diamond C \to \Diamond D) \urcorner$;
(b) Formulae of the form $\ulcorner \Diamond C \to \Diamond D \urcorner$.

Thus, we see that given $\boldsymbol{L}$ that is meant to fulfil the "left-to-right" implication of ($\star$) it is enough to satisfy the following two conditions:

1. $\boldsymbol{L}$ defines $\mathbf{D_2}$ (for example, when $\boldsymbol{L} \in \mathsf{S5}_{\Diamond}$);
2. The formula (C) belongs to $\boldsymbol{L}$.

Further we will show that these conditions are necessary for $\boldsymbol{L}$ to define the $\mathbf{D_2}$-consequence.

Now notice that to fulfill the reverse implication in ($\star$), it is enough to have the condition that $\boldsymbol{L} \subseteq \mathbf{S5}$.[4] Indeed, let $\{\Diamond A^\bullet : A \in \Pi\} \vdash_{\boldsymbol{L}} \Diamond B^\bullet$, i.e. $\ulcorner(\Diamond A_1^\bullet \wedge \cdots \wedge \Diamond A_n^\bullet) \to \Diamond B^{\bullet}\urcorner \in \boldsymbol{L}$, for some $A_1, \ldots, A_n \in \Pi$. If $\boldsymbol{L} \subseteq \mathbf{S5}$, then also $\ulcorner(\Diamond A_1^\bullet \wedge \cdots \wedge \Diamond A_n^\bullet) \to \Diamond B^{\bullet}\urcorner \in \mathbf{S5}$, thus by Fact 5, $\Pi \vdash_{\mathbf{D_2}} B$. □

Theorem 1. *For any modal logic* $\boldsymbol{L}$:

(i) If $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ *and* (C) $\in \boldsymbol{L} \subseteq \mathbf{S5}$*, then* $\boldsymbol{L}$ *defines the* $\mathbf{D_2}$*-consequence.*
(ii) If $\boldsymbol{L}$ *defines the* $\mathbf{D_2}$*-consequence, then* $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ *and contains* (C).

Proof. (i) See the above remark.
(ii) Firstly, $\Diamond A^\bullet \in \boldsymbol{L}$ iff $\emptyset \vdash_{\boldsymbol{L}} \Diamond A^\bullet$ iff $\emptyset \vdash_{\mathbf{D_2}} A$ iff $A \in \mathbf{D_2}$, by Fact 6. Secondly, since $p \to^{\mathrm{d}} q, p \vdash_{\mathbf{D_2}} q$, so by the assumption we have that $\Diamond(\Diamond p \to q), \Diamond p \vdash_{\boldsymbol{L}} \Diamond q$. Hence (C) $\in \boldsymbol{L}$. □

For any modal logic $\boldsymbol{L}$ and any formula $A \in \mathrm{For_m}$, let $\boldsymbol{L}+A$ be the smallest modal logic which includes $\boldsymbol{L}$ and contains A.

By Lemma 19, Fact 2(ii), and Theorem 1(i) we obtain:

Corollary 1. *If* $\boldsymbol{L}$ *defines* $\mathbf{D_2}$ *and* $\boldsymbol{L} \subseteq \mathbf{S5}$*, then* $\boldsymbol{L}+$(C) *defines the* $\mathbf{D_2}$*-consequence.*

Moreover, we have that:

Theorem 2. *Let* $\boldsymbol{X}$ *be a set of modal logics such that both* $\mathbf{S5} \in \boldsymbol{X}$ *and there is in* $\boldsymbol{X}$ *the smallest logic defining* $\mathbf{D_2}$. *If* $\boldsymbol{L}$ *is this logic, then*:

(i) $\boldsymbol{L}+$(C) *defines the* $\mathbf{D_2}$*-consequence,*
(ii) $\boldsymbol{L}+$(C) *is included in all logics from* $\boldsymbol{X}$ *that define the* $\mathbf{D_2}$*-consequence.*

Proof. (i) Since $\mathbf{S5}$ defines $\mathbf{D_2}$, so $\boldsymbol{L} \subseteq \mathbf{S5}$. Thus, by Corollary 1, $\boldsymbol{L}+$(C) defines the $\mathbf{D_2}$-consequence.
(ii) Suppose that $\boldsymbol{L}' \in \boldsymbol{X}$ and $\boldsymbol{L}'$ defines the $\mathbf{D_2}$-consequence. Then $\boldsymbol{L}'$ defines $\mathbf{D_2}$ and (C) $\in \boldsymbol{L}'$, by Theorem 1(ii). Thus, $\boldsymbol{L} \subseteq \boldsymbol{L}'$ and $\boldsymbol{L}+$(C) $\subseteq \boldsymbol{L}'+$(C) $= \boldsymbol{L}'$. □

We put $\mathbf{A}_{\vdash} := \mathbf{A}+$(C). By Theorem 2 applied to the set of all modal logics we obtain:

Corollary 2. $\mathbf{A}_{\vdash}$ *is the smallest modal logic defining the* $\mathbf{D_2}$*-consequence.*

The logic $\mathbf{A}_{\vdash}$ is also examined in Nasieniewski and Pietruszczak (2012). It is shown there that $\mathbf{A}_{\vdash} \notin \mathbf{S5}_{\diamond}$. As in the case of $\mathbf{A}$, for any $T \in \mathbf{PL}$, $\ulcorner\Diamond\Diamond T\urcorner \notin \mathbf{A}_{\vdash}$.

We easily see:

Fact 7. *Let* $\boldsymbol{X}$ *be a set of modal logics which is closed under arbitrary intersections. If* $\boldsymbol{L} \in \boldsymbol{X}$, $A \in \mathrm{For_m}$ *and there is a logic in* $\boldsymbol{X}$ *including* $\boldsymbol{L} \cup \{A\}$*, then there is the smallest logic in* $\boldsymbol{X}$ *including* $\boldsymbol{L} \cup \{A\}$*. Let us denote this logic by* $\boldsymbol{L}+_{\boldsymbol{X}} A$.

[4] By Fact 3(iii) in the case of congruential logic this requirement follows from the condition that $\boldsymbol{L}$ defines $\mathbf{D_2}$. Moreover, as we will see in Theorem 8, all considered logics define $\mathbf{D_2}$ and are included in $\mathbf{S5}$ (also non-congruential ones).

Theorem 3. *Let $\boldsymbol{X}$ be a set of modal logics which contains* $\mathbf{S5}$ *and is closed under arbitrary intersections. Let $\boldsymbol{L}$ be the smallest logic in $\boldsymbol{X}$ defining* $\mathbf{D_2}$.[5] *Then $\boldsymbol{L}+_{\boldsymbol{X}}$(C) is the smallest logic in $\boldsymbol{X}$ defining the* $\mathbf{D_2}$*-consequence.*[6]

Proof. By Theorem 2(i), $\boldsymbol{L}+$(C) defines the $\mathbf{D_2}$-consequence. Moreover, $\boldsymbol{L}+(\mathrm{C}) \subseteq \boldsymbol{L}+_{\boldsymbol{X}}(\mathrm{C}) \subseteq \mathbf{S5}$. Hence $\boldsymbol{L}+_{\boldsymbol{X}}$(C) defines the $\mathbf{D_2}$-consequence, by Lemma 3. The rest follows as in the proof of Theorem 2(ii). □

Definition 6. Let $\mathsf{Cn}_{\Diamond}\mathsf{S5}$ be the set of modal logics which satisfies the condition (‡) given in the introduction, i.e.:

$$\boldsymbol{L} \in \mathsf{Cn}_{\Diamond}\mathsf{S5} \overset{\mathrm{df}}{\iff} \text{for any } \Pi \subseteq \mathrm{For_m} \text{ and } B \in \mathrm{For_m},$$
$$\Diamond\Pi \vdash_{\boldsymbol{L}} \Diamond B \quad \text{iff} \quad \Diamond\Pi \vdash_{\mathbf{S5}} \Diamond B.$$

It is obvious that

Fact 8.

(i) $\mathsf{Cn}_{\Diamond}\mathsf{S5} \subseteq \mathsf{S5}_{\Diamond}$.
(ii) Every logic from $\mathsf{Cn}_{\Diamond}\mathsf{S5}$ *defines the* $\mathbf{D_2}$*-consequence.*
(iii) If $\boldsymbol{L} \subseteq \boldsymbol{L}' \subseteq \mathbf{S5}$ *and* $\boldsymbol{L} \in \mathsf{Cn}_{\Diamond}\mathsf{S5}$, *then* $\boldsymbol{L}' \in \mathsf{Cn}_{\Diamond}\mathsf{S5}$.

Since $\mathbf{A}_{\vdash} \notin \mathsf{S5}_{\Diamond}$, so $\mathbf{A}_{\vdash} \notin \mathsf{Cn}_{\Diamond}\mathsf{S5}$. Thus, the smallest logic defining the $\mathbf{D_2}$-consequence does not belong to $\mathsf{Cn}_{\Diamond}\mathsf{S5}$.

Theorem 4. *For any modal logic $\boldsymbol{L}$:*

(i) If $\boldsymbol{L} \in \mathsf{S5}_{\Diamond}$, $\boldsymbol{L} \subseteq \boldsymbol{L}^{\star} \subseteq \mathbf{S5}$ *and* (C) $\in \boldsymbol{L}^{\star}$, *then* $\boldsymbol{L}^{\star} \in \mathsf{Cn}_{\Diamond}\mathsf{S5}$.
(ii) If $\boldsymbol{L} \in \mathsf{Cn}_{\Diamond}\mathsf{S5}$, *then* $\boldsymbol{L} \in \mathsf{S5}_{\Diamond}$ *and* (C) $\in \boldsymbol{L}$.

Proof. Let $\boldsymbol{L} \in \mathsf{S5}_{\Diamond}$, $\boldsymbol{L} \subseteq \boldsymbol{L}^{\star} \subseteq \mathbf{S5}$ and (C) $\in \boldsymbol{L}^{\star}$.

By Lemma 22, if $\Diamond\Pi \vdash_{\boldsymbol{L}^{\star}} \Diamond B$, then $\Diamond\Pi \vdash_{\mathbf{S5}} \Diamond B$, since $\boldsymbol{L}^{\star} \subseteq \mathbf{S5}$.

Reversely, suppose that $\Diamond\Pi \vdash_{\mathbf{S5}} \Diamond B$. Then, by Lemma 22, for some $A_1, \ldots, A_n \in \Pi$ we have that $\ulcorner\Diamond A_1 \to (\ldots(\Diamond A_n \to \Diamond B)\ldots)\urcorner \in \mathbf{S5}$. So, by Lemma 19, $\ulcorner\Diamond(\Diamond A_1 \to (\ldots(\Diamond A_n \to B)\ldots))\urcorner \in \mathbf{S5}$. Hence this formula belongs to $\boldsymbol{L}^{\star}$, since $\boldsymbol{L} \in \mathsf{S5}_{\Diamond}$ and $\boldsymbol{L} \subseteq \boldsymbol{L}^{\star}$. Thus, since (C) $\in \boldsymbol{L}^{\star}$, by Lemma 20, also $\ulcorner\Diamond A_1 \to (\Diamond A_2 \to \ldots(\Diamond A_n \to \Diamond B)\ldots)\urcorner \in \boldsymbol{L}^{\star}$. So $\Diamond\Pi \vdash_{\boldsymbol{L}^{\star}} \Diamond B$.

Firstly, we use Fact 8(i). Secondly, one can see that $\Diamond(\Diamond p \to q), \Diamond p \vdash_{\mathbf{S5}} \Diamond q$, so by the assumption we have that $\Diamond(\Diamond p \to q), \Diamond p \vdash_{\boldsymbol{L}} \Diamond q$. Hence (C) $\in \boldsymbol{L}$. □

Similarly as Fact 3(i) we obtain the following:

Theorem 5. *For any rte-logic $\boldsymbol{L}$: $\boldsymbol{L}$ defines the* $\mathbf{D_2}$*-consequence iff* $\boldsymbol{L} \in \mathsf{Cn}_{\Diamond}\mathsf{S5}$.

[5] Notice that this logic exists, because the set of logics in $\boldsymbol{X}$ defining $\mathbf{D_2}$ is closed under arbitrary intersections and $\mathbf{S5}$ belongs to this set.

[6] By Fact 7, the logic $\boldsymbol{L}+_{\boldsymbol{X}}$(C) exists, since $\boldsymbol{L}, \mathbf{S5} \in \boldsymbol{X}$ and $\boldsymbol{L} \cup \{(\mathrm{C})\} \subseteq \mathbf{S5}$.

Proof. "⇒" Let $\boldsymbol{L}$ be any rte-logic. We define a function $(-)^{\circ}$ from $\mathrm{For_m}$ into $\mathrm{For^d}$ which «un-modalizes» every modal formula:

1. $(a)^{\circ} = a$, for any propositional letter a,
2. For any $A, B \in \mathrm{For_m}$:

$$
\begin{aligned}
&(\neg A)^{\circ} = \ulcorner \neg A^{\circ} \urcorner,\\
&(A \vee B)^{\circ} = \ulcorner A^{\circ} \vee B^{\circ} \urcorner,\\
&(A \wedge B)^{\circ} = \ulcorner \neg(\neg A^{\circ} \vee \neg B^{\circ}) \urcorner,\\
&(A \to B)^{\circ} = \ulcorner \neg A^{\circ} \vee B^{\circ} \urcorner,\\
&(A \leftrightarrow B)^{\circ} = \ulcorner \neg(\neg(\neg A^{\circ} \vee B^{\circ}) \vee \neg(\neg B^{\circ} \vee A^{\circ})) \urcorner,\\
&(\Diamond A)^{\circ} = \ulcorner (p \vee \neg p) \wedge^{\mathrm{d}} A^{\circ} \urcorner,\\
&(\Box A)^{\circ} = \ulcorner \neg A^{\circ} \to^{\mathrm{d}} \neg(p \vee \neg p) \urcorner.
\end{aligned}
$$

Notice that for any $A, B \in \mathrm{For_m}$ we have the following equalities:

$$
\begin{aligned}
(\neg A)^{\circ\bullet} &= \ulcorner \neg A^{\circ\bullet} \urcorner,\\
(A \to B)^{\circ\bullet} &= \ulcorner \neg A^{\circ\bullet} \vee B^{\circ\bullet} \urcorner,\\
(\Diamond A)^{\circ\bullet} &= \ulcorner (p \vee \neg p) \wedge \Diamond A^{\circ\bullet} \urcorner.
\end{aligned}
\tag{$*$}
$$

In Nasieniewski and Pietruszczak (2011) it was proved that for any $A \in \mathrm{For_m}$:

$$\ulcorner (\Diamond A)^{\circ\bullet} \leftrightarrow \Diamond A^{\circ\bullet} \urcorner \in \mathbf{PL} \tag{$\bullet$}$$

$$A^{\circ\bullet} \in \boldsymbol{L} \ \text{ iff } \ A \in \boldsymbol{L}\,, \tag{$\bullet\bullet$}$$

$$\ulcorner A^{\circ\bullet} \leftrightarrow A \urcorner \in \mathbf{S5} \tag{$\bullet\bullet\bullet$}$$

Finally, suppose that $\boldsymbol{L}$ defines the $\mathbf{D_2}$-consequence. Then for any $A_1, \ldots, A_n, B \in \mathrm{For_m}$ we have: $\Diamond A_1, \ldots, \Diamond A_n \vdash_{\boldsymbol{L}} \Diamond B$ iff $\ulcorner \Diamond A_1 \to (\ldots(\Diamond A_n \to \Diamond B)\ldots) \urcorner \in \boldsymbol{L}$ iff (by ($\bullet\bullet$)) $\ulcorner (\Diamond A_1 \to (\ldots(\Diamond A_n \to \Diamond B)\ldots))^{\circ\bullet} \urcorner \in \boldsymbol{L}$ iff (by ($*$) and ($\bullet$)) $\ulcorner \Diamond A_1^{\circ\bullet} \to (\ldots(\Diamond A_n^{\circ\bullet} \to \Diamond B^{\circ\bullet})\ldots) \urcorner \in \boldsymbol{L}$ iff $\Diamond A_1^{\circ\bullet}, \ldots, \Diamond A_n^{\circ\bullet} \vdash_{\boldsymbol{L}} \Diamond B^{\circ\bullet}$ iff (since $\boldsymbol{L}$ defines the $\mathbf{D_2}$-consequence) $A_1^{\circ}, \ldots, A_n^{\circ} \vdash_{\mathbf{D_2}} B^{\circ}$ iff $\Diamond A_1^{\circ\bullet}, \ldots, \Diamond A_n^{\circ\bullet} \vdash_{\mathbf{S5}} \Diamond B^{\circ\bullet}$ iff (by ($\bullet\bullet\bullet$)) $\Diamond A_1, \ldots, \Diamond A_n \vdash_{\mathbf{S5}} \Diamond B$. So $\boldsymbol{L} \in \mathsf{Cn}_{\Diamond}\mathbf{S5}$.

"⇐" See Fact 8(ii). □

6.2.4 *Discussive Systems*

Jaśkowski's aim was to give a calculus which could be applied for inconsistent systems without leading them to overcompleteness i.e. systems whose set of theses is not equal to the set of all meaningful expressions of the language.

As a solution of the problem, Jaśkowski proposed a way to generate a *discussive system*, i.e. a system that is based on the situation of a discussion. A conclusive functor for Jaśkowski's choice was the fact that during a discussion some inconsistent statements can appear, but we are not inclined to accept every statement on that basis.

The statements explicitly expressed during a discussion can be treated as provided with a predicate 'according to one of participants of the discussion'. This can be written by the use of the phrase 'it is possible that'. If we take the point of view of an external observer (that is of someone who is not a participant of the discussion), then all statements are only possible. Also conclusions that follow from explicit statements are only possible and can be treated as implicit statements of the discussion. Thus, explicit and implicit statements of the discussion are treated equally as the theses of the discussive system.

Jaśkowski treats discussive systems as built up from sentences of a natural language with the use of the discussive and classical connectives. He used Greek letters '𝔓' and '𝔔' to denote sentences of a given system. Thus discussive systems do not consist of schemas. These are used by Jaśkowski to examine properties of discussive systems.

Summarizing, *a discussive system* is a set of sentences that fulfills the following two conclusions:

1. It is contained in some set of sentences that is closed under any operations of the classical as well as of the discussive connectives;
2. It is closed under the $\mathbf{D_2}$-consequence.

While examining formal properties of discussive systems we can consider these systems as subsets of $\mathrm{For}^{\mathrm{d}}$, but not as compound of sentences of the natural language. In this way we could think of sentential letters 'p', 'q', ... as referring to atomic sentences of the given discussive system.

Thus, applying the transition from the set of sentences of the natural language to the set $\mathrm{For}^{\mathrm{d}}$, we can say that a subset $\boldsymbol{S}$ of $\mathrm{For}^{\mathrm{d}}$ is a *discussive system* iff $\boldsymbol{S}$ is closed under the $\mathbf{D_2}$-consequence, i.e., for any $\Pi \subseteq \mathrm{For}^{\mathrm{d}}$ and $B \in \mathrm{For}^{\mathrm{d}}$, if $\Pi \vdash_{\mathbf{D_2}} B$ and $\Pi \subseteq \boldsymbol{S}$, then $B \in \boldsymbol{S}$.

Of course, all discussive systems contain the set $\mathbf{D_2}$. Moreover, as Jaśkowski observes himself (see Jaśkowski 1999a, p. 44 and also Jaśkowski 1948, p. 67), every discussive system is closed under the *modus pones* rule for '$\to^{\mathrm{d}}$'. Formally, by definitions and Lemma 2, we obtain:

Fact 9. *Let $\boldsymbol{S}$ be a discussive system. Then*:

(i) $\mathbf{D_2} \subseteq \boldsymbol{S}$,
(ii) For any $A, B \in \mathrm{For}^{\mathrm{d}}$: $A \in \boldsymbol{S}$ and $\ulcorner A \to^{\mathrm{d}} B \urcorner \in \boldsymbol{S}$, then $B \in \boldsymbol{S}$.

By the above fact and Fact 6 we have:

Fact 10. *A subset $\boldsymbol{S}$ of* $\mathrm{For}^{\mathrm{d}}$ *is a discussive system iff $\boldsymbol{S}$ satisfies the conditions (i) and (ii) given in Fact 9.*

A broader introduction to the theory of discussive systems can be found for example in Nasieniewski and Pietruszczak (2013).

6.3 The Smallest Normal and Regular Modal Logics Defining the $\mathbf{D_2}$-Consequence

In Perzanowski (1975) the smallest normal logic in $\mathsf{\mathbf{S5}}_\diamond$ denoted by $\mathbf{S5^M}$ was indicated. This logic was defined (see Perzanowski 1975) as the smallest normal logic containing ($\mathtt{P}$), $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$, and closed under the following rule:

$$\frac{\Diamond\Diamond\, A}{\Diamond\, A} \qquad (\mathrm{cut}^{\Diamond\Diamond}_{\Diamond})$$

Of course, since $\mathbf{S5^M} \in \mathsf{\mathbf{S5}}_\diamond$, so $\mathbf{S5^M}$ defines $\mathbf{D_2}$, by Fact 3(i). Summarizing:

Fact 11 (Nasieniewski and Pietruszczak 2008; Perzanowski 1975). $\mathbf{S5^M}$ *is the smallest normal logic in* $\mathsf{\mathbf{S5}}_\diamond$; *so* $\mathbf{S5^M}$ *is the smallest normal logic defining* $\mathbf{D_2}$.

In Nasieniewski and Pietruszczak (2008) $\mathbf{rS5^M}$—the smallest regular logic defining $\mathbf{D_2}$—was indicated. Its definition is as follows

$\mathbf{rS5^M}$:= the smallest regular logic containing $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and closed under ($\mathrm{cut}^{\Diamond\Diamond}_{\Diamond}$).

We recall the formula

$$\Box p \to \Diamond\Box\Box p \qquad (\mathtt{4}_{\mathtt{s}})$$

and the following facts.

Fact 12 (Nasieniewski and Pietruszczak 2008).

(i) *The logic* $\mathbf{rS5^M}$ *is not normal. Thus,* $\mathbf{rS5^M} \subsetneq \mathbf{S5^M}$.
(ii) ($\mathtt{D}$), ($\mathtt{P}$), $\ulcorner\Diamond\Box(\mathtt{5})\urcorner \in \mathbf{rS5^M}$, *so also* $\Diamond\mathbf{PL} \subseteq \mathbf{rS5^M}$.
(iii) $\mathbf{rS5^M}$ *is the smallest regular logic in* $\mathsf{\mathbf{S5}}_\diamond$; *so* $\mathbf{rS5^M}$ *is the smallest regular logic defining* $\mathbf{D_2}$.

Fact 13 (Nasieniewski and Pietruszczak 2009). $\mathbf{rS5^M}$ *is the smallest regular logic which*:

(i) *Contains* $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ *and* ($\mathtt{4}_{\mathtt{s}}$), *i.e.* $\mathbf{rS5^M} = \mathbf{C4_s} \oplus \Diamond\Box(\mathtt{T})$;
(ii) *Contains* ($\mathtt{5}_{\mathtt{c}}$) *and* ($\mathtt{4}_{\mathtt{s}}$), *i.e.* $\mathbf{rS5^M} = \mathbf{C5_c4_s}$;
(iii) *Contains* ($\mathtt{5}_{\mathtt{c}}$) *and is closed under* ($\mathrm{cut}^{\Diamond\Diamond}_{\Diamond}$).

Notice that the set of all normal (resp. regular) logics is closed under arbitrary intersections and **S5** is normal (resp. regular). Thus, by Facts 2(ii), 11 and 12(iii), and Theorem 3 we obtain.

Corollary 3. *(i)* $\mathbf{S5^M} \oplus (\mathtt{C})$ *is the smallest normal logic defining the* $\mathbf{D_2}$*-consequence.*
(ii) $\mathbf{rS5^M} \oplus (\mathtt{C})$ *is the smallest regular logic defining the* $\mathbf{D_2}$*-consequence.*

In Nasieniewski and Pietruszczak (2013)—using other methods—it was proved that:

Fact 14 (Nasieniewski and Pietruszczak 2013).

(i) $\mathbf{KD45}$ $(= \mathbf{K55_c})$ *is the smallest normal logic defining the* $\mathbf{D_2}$*-consequence.*
(ii) $\mathbf{CD45(1)}$ *is the smallest regular logic defining the* $\mathbf{D_2}$*-consequence.*[7]

Thus we have that

$$\mathbf{S5^M} \oplus (\mathtt{C}) = \mathbf{KD45} = \mathbf{K55_c}\,,$$

$$\mathbf{rS5^M} \oplus (\mathtt{C}) = \mathbf{CD45(1)} = \mathbf{CN^15_c5(1)}\,.$$

6.4 A General Method of Generating Logics Defining the $\mathbf{D_2}$-Consequence

Now we will recall some notions from Nasieniewski and Pietruszczak (2011). Thus, let $\boldsymbol{X}$ be a set of modal logics which are determined by some set $\mathscr{A}_{\boldsymbol{X}}$ of axioms and some set $\mathscr{R}_{\boldsymbol{X}}$ of rules, i.e. $\boldsymbol{X}$ is the set of all modal logics which include $\mathscr{A}_{\boldsymbol{X}}$ and are closed under all rules from $\mathscr{R}_{\boldsymbol{X}}$. In what follows we give a general method which generates the weakest element in $\mathsf{Cn}_\diamond\mathsf{S5} \cap \boldsymbol{X}$. Thus, for the set of all modal logics $(\mathscr{A}_{\boldsymbol{X}} = \emptyset = \mathscr{R}_{\boldsymbol{X}})$ we obtain the weakest logic in $\mathsf{Cn}_\diamond\mathsf{S5}$. Moreover, by Theorem 5, if $\boldsymbol{X}$ consists of rte-logics, then the obtained weakest logic in $\mathsf{Cn}_\diamond\mathsf{S5} \cap \boldsymbol{X}$ is also the weakest logic in $\boldsymbol{X}$ defining $\mathbf{D_2}$.

Similarly as in Nasieniewski and Pietruszczak (2011) we will apply the method for the case of chosen sets of modal logics. For any $\varPhi \subseteq \mathrm{For_m}$, let

$$\Diamond\Box\varPhi := \{\ulcorner\Diamond\Box A\urcorner : A \in \varPhi\},$$

and for any rule R on $\mathrm{For_m}$ we define the following rule $R^{\Diamond\Box}$ on $\mathrm{For_m}$:

$$R^{\Diamond\Box} := \{\langle\Diamond\Box A_1, \dots, \Diamond\Box A_n, \Diamond\Box B\rangle : \langle A_1, \dots, A_n, B\rangle \in R\}$$

For any set of rules $\mathscr{R}$ on $\mathrm{For_m}$ we put: $\mathscr{R}^{\Diamond\Box} := \{\, R^{\Diamond\Box} : R \in \mathscr{R}\,\}$.

Since $(\mathrm{sb})^{\Diamond\Box} \subseteq (\mathrm{sb})$ we obtain

Lemma 4 (Nasieniewski and Pietruszczak 2011). *All sets closed under* (sb) *are closed under* $(\mathrm{sb})^{\Diamond\Box}$*. So all modal logics are closed under* $(\mathrm{sb})^{\Diamond\Box}$*.*

Lemma 5 (Nasieniewski and Pietruszczak 2011). *Let R be a rule such that* $\mathbf{S5}$ *is closed under R. Then all logics from* $\mathsf{S5}_\diamond$ *are closed under the rule* $R^{\Diamond\Box}$*.*

[7] $\mathbf{CD45(1)}$ is the smallest regular logic which contains (D), (4), and (5(1)), i.e. $\ulcorner\Box\top \to (5)\urcorner$. Also $\mathbf{CD45(1)} = \mathbf{CN^15_c5(1)}$, where $(\mathtt{N}^1)$ is '$\Box(p \to p) \to \Box\Box(p \to p)$'.

Corollary 4 (Nasieniewski and Pietruszczak 2011). *For every logic* $\boldsymbol{L} \in \mathsf{\mathbf{S5}}_{\diamond}$*,* $\boldsymbol{L}$ *is closed under the rules* $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{mon})^{\Diamond\Box}$, $(\mathrm{nec})^{\Diamond\Box}$ *and* $(\mathrm{cut}^{\diamond\diamond}_{\diamond})^{\Diamond\Box}$*, i.e.,* $\boldsymbol{L}$ *is closed, respectively, under the following rules*:

$$\frac{\Diamond\Box A \qquad \Diamond\,\Box(A \to B)}{\Diamond\Box B} \qquad \frac{\Diamond\Box(A \to B)}{\Diamond\Box(\Box A \to \Box B)} \qquad \frac{\Diamond\Box A}{\Diamond\Box\Box A} \qquad \frac{\Diamond\Box\Diamond\Diamond\, A}{\Diamond\Box\,\Diamond\, A}$$

Let us recall

Lemma 6 (Nasieniewski and Pietruszczak 2011). *If* $\boldsymbol{L} \in \mathsf{\mathbf{S5}}_{\diamond}$ *and* $\boldsymbol{L} \subseteq \mathbf{S5}$*, then* $\Diamond\Box\boldsymbol{L} \subseteq \bigcap \mathsf{\mathbf{S5}}_{\diamond}$.

Now, since $\mathsf{\mathbf{Cn}}_{\diamond}\mathsf{\mathbf{S5}} \subseteq \mathsf{\mathbf{S5}}_{\diamond}$ we have a direct corollary from Lemma 6:

Corollary 5. *If* $\boldsymbol{L} \in \mathsf{\mathbf{Cn}}_{\diamond}\mathsf{\mathbf{S5}}$ *and* $\boldsymbol{L} \subseteq \mathbf{S5}$*, then* $\Diamond\Box\boldsymbol{L} \subseteq \bigcap \mathsf{\mathbf{Cn}}_{\diamond}\mathsf{\mathbf{S5}}$.

Let $\mathscr{A} \subseteq \mathrm{For_m}$ and $\mathscr{R}$ be a set of rules on $\mathrm{For_m}$. We say that the pair $\langle \mathscr{A}, \mathscr{R}\rangle$ is an *axiomatization* of a modal logic $\boldsymbol{L}$ iff $\boldsymbol{L}$ is the smallest set including $\mathscr{A}$, that is closed under all rules from $\mathscr{R}$. Then for any $A \in \mathrm{For_m}$: $A \in \boldsymbol{L}$ iff there exists a sequence $A_1, \dots, A_n = A$ in which for any $i \leqslant n$, either $A_i \in \mathscr{A}$, or there are $R \in \mathscr{R}$, $m < n$, $j_1, \dots, j_m < i$ such that $\langle A_{j_1}, \dots, A_{j_m}, A_i\rangle \in R$.

Below we recall Lemma 2.5 from Nasieniewski and Pietruszczak (2011):

Lemma 7. *Let* $\langle \mathscr{A}, \mathscr{R}\rangle$ *be any axiomatization of a modal logic* $\boldsymbol{L}$*. Let* $\boldsymbol{L}^*$ *be any modal logic such that* $\Diamond\Box\mathscr{A} \subseteq \boldsymbol{L}^*$ *and* $\boldsymbol{L}^*$ *is closed under all rules from* $\mathscr{R}^{\Diamond\Box}$*. Then* $\Diamond\Box\boldsymbol{L} \subseteq \boldsymbol{L}^*$*, i.e. for any* $A \in \boldsymbol{L}$: $\ulcorner\Diamond\Box A\urcorner \in \boldsymbol{L}^*$.

We will use Theorem 2.6 also from Nasieniewski and Pietruszczak (2011):

Theorem 6. *Let* $\boldsymbol{L} \in \mathsf{\mathbf{S5}}_{\diamond}$ *and* $\langle \mathscr{A}, \mathscr{R}\rangle$ *be an axiomatization of* $\boldsymbol{L}$ *such that* $\mathscr{A} \subseteq \mathbf{S5}$ *and* $\mathbf{S5}$ *is closed under all rules from* $\mathscr{R}$*. Let* $\boldsymbol{L}^{\Diamond\Box}$ *be the smallest modal logic including* $\Diamond\Box\mathscr{A}$*, closed under all rules from* $\mathscr{R}^{\Diamond\Box}$ *and*

$$\frac{\Diamond\Box\,\Diamond\, A}{\Diamond A} \qquad (\mathrm{cut}^{\diamond\Box\diamond}_{\diamond})$$

Then

(i) $\boldsymbol{L}^{\Diamond\Box} \in \mathsf{\mathbf{S5}}_{\diamond}$*,*
(ii) $\boldsymbol{L}^{\Diamond\Box} = \bigcap \mathsf{\mathbf{S5}}_{\diamond}$*; so* $\boldsymbol{L}^{\Diamond\Box}$ *is the smallest logic in* $\mathsf{\mathbf{S5}}_{\diamond}$*.*[8]

Using Theorem 6 we can take for example any $\boldsymbol{L}$ such that $\mathbf{rS5^M} \subseteq \boldsymbol{L} \subseteq \mathbf{S5}$, and any of its axiomatizations.[9] In each case we obtain the smallest logic in $\mathsf{\mathbf{S5}}_{\diamond}$. Let us denote this logic by $\mathbf{aS5^M}$ ("absolute" one).

[8] Notice that for any $\mathscr{A}$ and $\mathscr{R}$, the logic $\boldsymbol{L}^{\Diamond\Box}$ exists.

[9] For any such logic which is additionally closed under (rep), one can take in its axiomatization just the axiom (`df` $\Diamond$) instead of the set of formulae (`rep`$^{\Box}$).

Further on by the standard axiomatization of the logic **S5** we mean the setting consists of axioms **Taut**, ($\mathtt{df}\,\Diamond$), ($\mathtt{K}$), ($\mathtt{T}$) and ($\mathtt{5}$), and rules (mp), (sb) and (nec). Thus—selecting for example the standard axiomatization of **S5**—we obtain that $\mathbf{aS5^M}$ is the smallest modal logic which

- Includes the set $\Diamond\Box$**Taut**,
- Contains the formulae $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner$, $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$,
- And is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{nec})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Notice that $\mathbf{A} \subsetneq \mathbf{aS5^M}$ (see p. 98 and Fact 2(i)). So the smallest logic defining $\mathbf{D_2}$ is weaker than the smallest logic in $\mathsf{S5}_\diamond$.

We will refer to a logic $\boldsymbol{L}^{\Diamond\Box}_{\boldsymbol{X}}$ described in the following extension of Theorem 6:

Theorem 7 (Nasieniewski and Pietruszczak 2011). *Let $\boldsymbol{L} \in \mathsf{S5}_\diamond$ and $\boldsymbol{L} \in \boldsymbol{X}$, where $\boldsymbol{X}$ is a set of all modal logics including a given set of formulae $\mathscr{A}_{\boldsymbol{X}}$ and closed under all rules from some set $\mathscr{R}_{\boldsymbol{X}}$. Let $\langle\mathscr{A},\mathscr{R}\rangle$ be an axiomatization of $\boldsymbol{L}$ such that $\mathscr{A} \subseteq$ **S5** and **S5** is closed under all rules from $\mathscr{R}$. Let $\boldsymbol{L}^{\Diamond\Box}_{\boldsymbol{X}}$ be the smallest modal logic including $\mathscr{A}_{\boldsymbol{X}} \cup \Diamond\Box\mathscr{A}$ and closed under all rules from $\mathscr{R}_{\boldsymbol{X}} \cup \mathscr{R}^{\Diamond\Box} \cup \{(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})\}$. Then*

(i) $\boldsymbol{L}^{\Diamond\Box}_{\boldsymbol{X}} \in \mathsf{S5}_\diamond \cap \boldsymbol{X}$,
(ii) $\boldsymbol{L}^{\Diamond\Box}_{\boldsymbol{X}} = \bigcap(\mathsf{S5}_\diamond \cap \boldsymbol{X})$; *so* $\boldsymbol{L}^{\Diamond\Box}_{\boldsymbol{X}}$ *is the smallest logic in* $\mathsf{S5}_\diamond \cap \boldsymbol{X}$.

Let us recall (see Nasieniewski and Pietruszczak 2011) that $\mathrm{rte}\mathbf{S5^M}$ (respectively $\mathrm{cm}\mathbf{S5^M}$, $\mathbf{eS5^M}$ and $\mathbf{mS5^M}$) is the smallest rte- (respectively cm-, congruential and monotonic) logic in $\mathsf{S5}_\diamond$. By Fact 3(i) this logic is also the smallest rte-, (respectively cm-, congruential and monotonic) logic defining $\mathbf{D_2}$.

By Lemma 22, since the intersection of a non-empty family of modal logics is a modal logic, and thanks to **PL** we have

Fact 15. *If $\emptyset \neq \boldsymbol{X} \subseteq \mathsf{Cn}_\diamond\mathsf{S5}$, then $\emptyset \neq \bigcap \boldsymbol{X} \in \mathsf{Cn}_\diamond\mathsf{S5}$.*

Now we prove

Theorem 8. *Let $\boldsymbol{L} \in \mathsf{Cn}_\diamond\mathsf{S5}$ and $\langle\mathscr{A},\mathscr{R}\rangle$ be an axiomatization of $\boldsymbol{L}$ such that $\mathscr{A} \subseteq$ **S5** and **S5** is closed under all rules from $\mathscr{R}$. Let $\boldsymbol{L}^{\Diamond\Box\star}$ be the smallest modal logic including $\Diamond\Box\mathscr{A}$ and ($\mathtt{C}$), which is closed under all rules from $\mathscr{R}^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$. Then*

(i) $\boldsymbol{L}^{\Diamond\Box\star} \in \mathsf{Cn}_\diamond\mathsf{S5}$,
(ii) $\boldsymbol{L}^{\Diamond\Box\star} = \bigcap \mathsf{Cn}_\diamond\mathsf{S5}$; *so* $\boldsymbol{L}^{\Diamond\Box\star}$ *is the smallest logic in* $\mathsf{Cn}_\diamond\mathsf{S5}$.[10]

Proof. First, we prove that $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \bigcap \mathsf{Cn}_\diamond\mathsf{S5}$. Notice that $\bigcap \mathsf{Cn}_\diamond\mathsf{S5} \subseteq$ **S5**, since **S5** $\in \mathsf{Cn}_\diamond\mathsf{S5}$. Moreover, since whenever $\boldsymbol{L}' \in \mathsf{Cn}_\diamond\mathsf{S5}$, by Fact 15, $\boldsymbol{L}' \cap \mathbf{S5} \in \{\boldsymbol{L}'' \in \mathsf{Cn}_\diamond\mathsf{S5} : \boldsymbol{L}'' \subseteq \mathbf{S5}\}$, thus again by Fact 15 and the standard properties of intersections of families we have that $\emptyset \neq \bigcap \mathsf{Cn}_\diamond\mathsf{S5} = \bigcap\{\boldsymbol{L}' \in \mathsf{Cn}_\diamond\mathsf{S5} : \boldsymbol{L}' \subseteq \mathbf{S5}\} \in \mathsf{Cn}_\diamond\mathsf{S5}$. Applying Corollary 5 to the logic $\boldsymbol{L}$ given in the assumptions of

[10] Again notice that for any $\mathscr{A}$ and $\mathscr{R}$, the logic $\boldsymbol{L}^{\Diamond\Box\star}$ exists.

the present lemma, we get $\Diamond\Box\mathscr{A} \subseteq \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5}$. Hence, for any $\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5}$ we have $\Diamond\Box\mathscr{A} \subseteq \boldsymbol{L}'$. Besides, by Fact 8(i) and Lemma 5, $\boldsymbol{L}'$ is closed under all rules from $\mathscr{R}^{\Diamond\Box}$. Additionally $\boldsymbol{L}'$ is closed under $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$, since by Lemma 15, **S5** is closed under this rule and due to Fact 8(i), $\boldsymbol{L}' \in \mathsf{S5}_{\diamond}$. Of course, also $\mathbf{PL}\cup(\mathtt{rep}^{\Box}) \subseteq \boldsymbol{L}'$ and $\boldsymbol{L}'$ is closed under the rules (mp) and (sb). Finally, by Theorem 4(ii), $(\mathtt{C}) \in \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5} \subseteq \boldsymbol{L}'$. So $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \boldsymbol{L}'$, for any $\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5}$. Thus, $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \bigcap\{\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5} : \boldsymbol{L}' \subseteq \mathbf{S5}\}$. As it was observed $\bigcap \mathsf{Cn}_{\diamond}\mathsf{S5} = \bigcap\{\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5} : \boldsymbol{L}' \subseteq \mathbf{S5}\}$, therefore $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5}$.

(i) Let $\boldsymbol{L}^{\Diamond\Box}$ be defined as in Theorem 6. So, $\boldsymbol{L}^{\Diamond\Box} \in \mathsf{S5}_{\diamond}$ and $\boldsymbol{L}^{\Diamond\Box} \subseteq \boldsymbol{L}^{\Diamond\Box\star} \subseteq \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5} \subseteq \mathbf{S5}$. Since as it was mentioned $\boldsymbol{L}^{\Diamond\Box} \in \mathsf{S5}_{\diamond}$, so in Theorem 4(i) we put $\boldsymbol{L} := \boldsymbol{L}^{\Diamond\Box}$ and $\boldsymbol{L}^{\star} := \boldsymbol{L}^{\Diamond\Box\star}$. We obtain that $\boldsymbol{L}^{\Diamond\Box\star} \in \mathsf{Cn}_{\diamond}\mathsf{S5}$.

(ii) Taking into account the above results we have: $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5} \subseteq \boldsymbol{L}^{\Diamond\Box\star}$. □

The logic $\boldsymbol{L}^{\Diamond\Box\star}$ indicated in Theorem 8 is independent of the choice of $\boldsymbol{L}$ and of its axiomatization.

Using Theorem 8 we can take for example any $\boldsymbol{L}$ such that $\mathbf{CD45(1)} \subseteq \boldsymbol{L} \subseteq \mathbf{S5}$, and any of its axiomatizations.[11] In each case we obtain the smallest logic in $\mathsf{Cn}_{\diamond}\mathsf{S5}$. Let us denote this logic by $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$. Thus—selecting for example the standard axiomatization of **S5**—we obtain that $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$ is the smallest modal logic which

- Contains all formulae and is closed under all rules for $\mathbf{aS5}^{\mathbf{M}}$ from p. 108,
- Contains the formula $(\mathtt{C})$.

Thus, $\mathbf{aS5}^{\mathbf{M}} \subseteq \mathbf{aS5}^{\mathbf{M}}{+}(\mathtt{C}) \subseteq \mathbf{aS5}^{\mathbf{M}}_{\vdash}$. Moreover, $\mathbf{aS5}^{\mathbf{M}}{+}(\mathtt{C}) \in \mathsf{Cn}_{\diamond}\mathsf{S5}$, in view of Theorem 4(i). Thus

$$\mathbf{aS5}^{\mathbf{M}}_{\vdash} = \mathbf{aS5}^{\mathbf{M}}{+}(\mathtt{C})\,.$$

Fact 16.

(i) The logics $\mathbf{aS5}^{\mathbf{M}}$, rte$\mathbf{S5}^{\mathbf{M}}$, cm$\mathbf{S5}^{\mathbf{M}}$, $\mathbf{eS5}^{\mathbf{M}}$, $\mathbf{mS5}^{\mathbf{M}}$, $\mathbf{rS5}^{\mathbf{M}}$, $\mathbf{S5}^{\mathbf{M}}$ *do not include the logic* $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$.

(ii) $\mathbf{aS5}^{\mathbf{M}} \subsetneq \mathbf{aS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{CD45(1)} = \mathbf{rS5}^{\mathbf{M}} \oplus (\mathtt{C})$.

Proof. (i) The logics $\mathbf{aS5}^{\mathbf{M}}$, rte$\mathbf{S5}^{\mathbf{M}}$, cm$\mathbf{S5}^{\mathbf{M}}$, $\mathbf{eS5}^{\mathbf{M}}$, $\mathbf{mS5}^{\mathbf{M}}$, $\mathbf{rS5}^{\mathbf{M}}$ and $\mathbf{S5}^{\mathbf{M}}$ are included in **S4** (see Nasieniewski and Pietruszczak 2011), while $(\mathtt{C}) \notin \mathbf{S4}$.

(ii) The inclusion $\mathbf{aS5}^{\mathbf{M}} \subseteq \mathbf{aS5}^{\mathbf{M}}_{\vdash}$ is obvious. By (i), $\mathbf{aS5}^{\mathbf{M}} \neq \mathbf{aS5}^{\mathbf{M}}{+}(\mathtt{C})$.

By Nasieniewski and Pietruszczak (2013, Lemma 5.1), $(\mathtt{C}) \in \mathbf{CD45(1)}$. On the other hand, by Nasieniewski and Pietruszczak (2013, Fact 5.3), we know that $\mathbf{rS5}^{\mathbf{M}} \subseteq \mathbf{CD45(1)}$, and due to Nasieniewski and Pietruszczak (2011, Facts 3.1, 3.3, 3.4, 3.5), $\mathbf{aS5}^{\mathbf{M}} \subseteq \mathbf{rS5}^{\mathbf{M}}$. Therefore $\mathbf{aS5}^{\mathbf{M}}_{\vdash} = \mathbf{aS5}^{\mathbf{M}}{+}(\mathtt{C}) \subseteq \mathbf{rS5}^{\mathbf{M}}{+}(\mathtt{C}) \subseteq \mathbf{rS5}^{\mathbf{M}} \oplus (\mathtt{C}) = \mathbf{CD45(1)}$.

[11] See Footnote 9.

Taking a counterexample from Nasieniewski and Pietruszczak (2011, Fact 5.3)—a valuation v into $\{0,1\}$, where v preserves classical truth conditions for classical constants and for any $A \in \mathrm{For}_{\mathrm{m}}$:

$$v(\Diamond A) = \begin{cases} 1 & \text{if } \ulcorner \Diamond A \urcorner \in \mathbf{S5} \\ 0 & \text{otherwise.} \end{cases} \qquad v(\Box A) = \begin{cases} 0 & \text{if } \ulcorner \Diamond \neg A \urcorner \in \mathbf{S5} \\ 1 & \text{otherwise.} \end{cases}$$

one can see that for any formula $A \in \mathbf{aS5}^{\mathbf{M}}_{\vdash}$ we have that $v(A) = 1$, while $v((\Box p \wedge \Box \neg p) \to \Box(p \wedge \neg p)) = 0$. Thus $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$ is not regular, i.e., $\mathbf{aS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{CD45(1)}$. □

We also have the following widening of Theorem 8.

Theorem 9. *Let* $\boldsymbol{L} \in \mathsf{Cn}_{\diamond}\mathsf{S5}$ *and* $\boldsymbol{L} \in \boldsymbol{\mathsf{X}}$*, where* $\boldsymbol{\mathsf{X}}$ *is a set of all modal logics including a given set of formulae* $\mathscr{A}_{\boldsymbol{\mathsf{X}}}$ *and closed under all rules from some set* $\mathscr{R}_{\boldsymbol{\mathsf{X}}}$*. Let* $\langle \mathscr{A}, \mathscr{R} \rangle$ *be an axiomatization of* $\boldsymbol{L}$ *such that* $\mathscr{A} \subseteq \mathbf{S5}$ *and* $\mathbf{S5}$ *is closed under all rules from* $\mathscr{R}$*. Let* $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$ *be the smallest modal logic including* $\mathscr{A}_{\boldsymbol{\mathsf{X}}} \cup \Diamond\Box\mathscr{A} \cup \{(\mathtt{C})\}$*, and closed under all rules from* $\mathscr{R}_{\boldsymbol{\mathsf{X}}} \cup \mathscr{R}^{\Diamond\Box} \cup \{(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})\}$*. Then*

(i) $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$*,*
(ii) $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} = \bigcap(\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}})$*; so* $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$ *is the smallest logic in* $\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$*.*

Proof. By the assumptions $\boldsymbol{L} \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$ and $\boldsymbol{L} \subseteq \mathbf{S5}$, thus one can easily see that for any $\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$ it holds that $\boldsymbol{L} \cap \boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$ and $\boldsymbol{L} \cap \boldsymbol{L}' \subseteq \mathbf{S5}$, so $\bigcap(\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}) = \bigcap\{\boldsymbol{L}' \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}} : \boldsymbol{L}' \subseteq \mathbf{S5}\}$. By Corollary 5, for any $\boldsymbol{L}'$ from $\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$ such that $\boldsymbol{L}' \subseteq \mathbf{S5}$ we have $\Diamond\Box\mathscr{A} \subseteq \boldsymbol{L}'$, so also $\Diamond\Box\mathscr{A}_{\boldsymbol{\mathsf{X}}} \cup \mathbf{PL} \cup (\mathtt{rep}^{\Box}) \subseteq \boldsymbol{L}'$. Moreover, by the assumptions and Lemmas 15 and 5, $\boldsymbol{L}'$ is closed under all rules from $\mathscr{R}^{\Diamond\Box} \cup \mathscr{R}_{\boldsymbol{\mathsf{X}}} \cup \{(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond}), (\mathrm{mp}), (\mathrm{sb})\}$. So, by Theorem 4(ii), $(\mathtt{C}) \in \bigcap \mathsf{Cn}_{\diamond}\mathsf{S5} \subseteq \boldsymbol{L}'$. So $\boldsymbol{L}^{\Diamond\Box\star} \subseteq \boldsymbol{L}'$. Hence $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} \subseteq \bigcap(\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}})$.

(i) As in the proof of Theorem 8, we obtain that $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} \in \mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$.
(ii) We have that $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} \subseteq \bigcap(\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}) \subseteq \boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$. □

If in Theorem 9, $\boldsymbol{\mathsf{X}}$ is for example the set of all regular logics, we can take as $\boldsymbol{L}$ logics **S5**, **KD45** or **CD45(1)** together with their axiomatizations. In each case we obtain the smallest logic in $\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$.

Of course, if in Theorem 9, $\boldsymbol{\mathsf{X}}$ is the set of all modal logics ($\mathscr{A}_{\boldsymbol{\mathsf{X}}} = \emptyset = \mathscr{R}_{\boldsymbol{\mathsf{X}}}$), then $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} = \boldsymbol{L}^{\Diamond\Box\star} = \mathbf{aS5}^{\mathbf{M}}_{\vdash}$. Moreover, by Theorem 5, if all members of $\boldsymbol{\mathsf{X}}$ are rte-logics, then the obtained weakest logic in $\mathsf{Cn}_{\diamond}\mathsf{S5} \cap \boldsymbol{\mathsf{X}}$ is also the weakest logic in $\boldsymbol{\mathsf{X}}$ defining the $\mathbf{D_2}$-consequence. Thus, if $\boldsymbol{\mathsf{X}}$ is the set of all congruential (resp. monotonic) modal logics, then $\boldsymbol{L}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$ is the smallest congruential (resp. monotonic) logic defining the $\mathbf{D_2}$-consequence. Moreover, by Theorem 9 and Fact 14, if $\boldsymbol{\mathsf{X}}$ is the set of all normal (resp. regular) logics, then $\mathbf{S5}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$ is the smallest normal logic in $\mathsf{Cn}_{\diamond}\mathsf{S5}$, i.e., $\mathbf{S5}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} = \mathbf{KD45}$ (resp. $\mathbf{S5}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}}$ is the smallest regular logic in $\mathsf{Cn}_{\diamond}\mathsf{S5}$, i.e., $\mathbf{S5}^{\Diamond\Box\star}_{\boldsymbol{\mathsf{X}}} = \mathbf{CD45(1)}$).

By Theorems 7 and 9, analyzing the axiomatizations of logics $L^{\Diamond\Box}_{X}$ and $L^{\Diamond\Box\star}_{X}$ we observe an analogous fact to Theorem 3:

Corollary 6. *Let $L \in \mathsf{Cn}_{\Diamond}\mathbf{S5}$ and $L \in X$, where X is a set of all modal logics including a given set of formulae $\mathscr{A}_X$ and closed under all rules from some set $\mathscr{R}_X$ and let $\langle \mathscr{A}, \mathscr{R}\rangle$ be an axiomatization of L such that $\mathscr{A} \subseteq$ **S5** and **S5** is closed under all rules from $\mathscr{R}$. Then it holds that:*

$$L^{\Diamond\Box\star}_{X} = L^{\Diamond\Box}_{X} +_X (\mathtt{C})$$

Proof. By Theorem 7, $L^{\Diamond\Box}_{X}$ is the smallest logic in $\mathbf{S5}_{\Diamond} \cap X$. By Theorem 9, $L^{\Diamond\Box\star}_{X}$ is the smallest logic in $\mathsf{Cn}_{\Diamond}\mathbf{S5} \cap X$. In Theorem 4(i) we put $L := L^{\Diamond\Box}_{X}$ and $L^{\star} := L^{\Diamond\Box}_{X} +_X (\mathtt{C})$. We obtain that $L^{\Diamond\Box}_{X} +_X (\mathtt{C}) \in \mathsf{Cn}_{\Diamond}\mathbf{S5} \cap X$. Thus, $L^{\Diamond\Box\star}_{X} = L^{\Diamond\Box}_{X} +_X (\mathtt{C})$. □

6.5 Simplifications of Axiomatizations of Generated Logics

We apply Theorem 9 to give axiomatizations of the smallest logic defining the $\mathbf{D_2}$-consequence respectively in the set of rte-, cm-, congruential, monotonic, regular and normal logics. We will also show that one can drop from these axiomatizations some of axioms or rules.

6.5.1 The Case of rte-Logics

If in Theorem 9, X is the set of all rte-logics, then we can assume that $\mathscr{A}_X = (\mathtt{rep}_{\mathbf{PL}})$ and $\mathscr{R}_X = \emptyset$. We can take as L logics **S5**, **KD45** and **CD45(1)**, and any of their axiomatizations. In each case we obtain the smallest rte-logic in $\mathsf{Cn}_{\Diamond}\mathbf{S5}$, so by Theorem 5 also the smallest rte-logic defining $\mathbf{D_2}$. Let us denote it by rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Now applying the standard axiomatization of **S5** we obtain that rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$ is the smallest rte-logic which

- Contains $(\mathtt{C})$, $(\mathtt{PN})$, $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner$, $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$,
- And closed under the rules $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{nec})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Indeed, let L be the smallest rte-logic fulfilling two above conditions. Since $(\mathtt{PN}) \in \Diamond\Box\mathbf{Taut}$, thus obviously by Theorem 9, $L \subseteq$ rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$. Besides, for any $A \in \mathbf{PL}$: $\ulcorner(\mathtt{PN}) \leftrightarrow \Diamond\Box A\urcorner \in L$, since $\ulcorner(p \to p) \leftrightarrow A\urcorner \in \mathbf{PL}$. So $\ulcorner\Diamond\Box A\urcorner \in L$. Thus, $\Diamond\Box\mathbf{PL} \subseteq L$. Therefore rte$\mathbf{S5}^{\mathbf{M}}_{\vdash} \subseteq L$. Summarizing, $L =$ rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Of course, taking other axiomatizations of **S5** we can obtain different axiomatizations of rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$. Moreover, using the logics **CD45(1)** and **KD45**, and their axiomatizations we can obtain further axiomatizations of rte$\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Observe that

Fact 17. *The formula*

$$\neg \Diamond \neg p \rightarrow \Box p \tag{$*$}$$

does not belong to $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$. *So* $\mathbf{aS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Proof. It appears that the counterexample used in Fact 3.1 from Nasieniewski and Pietruszczak (2011) for the case of respective logics from $\mathsf{S5}_{\diamond}$ also works in the case of $\mathsf{Cn}_{\diamond}\mathsf{S5}$. Thus, let v be a valuation from $\mathrm{For_m}$ into the set $\{0, 1\}$ such that: v preserves classical truth conditions for classical constants; $v(\Box p) = 0$; for any $A \in \mathrm{For_m}$: $v(\Diamond A) = 1$ iff $\ulcorner \Diamond A \urcorner \in \mathbf{S5}$; and for any $A, C \in \mathrm{For_m}$: $v(C) = v(C[{}^{\neg \Box \neg A}/_{\Diamond A}])$.[12]

One can see that $v(\Diamond(\Diamond p \rightarrow q)) = 0$, thus by Corollary 4 and Lemma 15, for any $A \in \mathbf{aS5}^{\mathbf{M}}_{\vdash}$ we have $v(A) = 1$. However $v(\Diamond \neg p) = 0$, thus ($*$) does not belong to $\mathbf{aS5}^{\mathbf{M}}_{\vdash}$. □

6.5.2 *The Case of cm-Logics*

If in Theorem 9, $\boldsymbol{X}$ is the set of all cm-logics, then we can assume that $\mathscr{A}_{\boldsymbol{X}} = (\mathtt{rep}_{\mathbf{PL}}) \cup \{(\mathtt{K}), (\mathtt{N})\}$ and $\mathscr{R}_{\boldsymbol{X}} = \emptyset$. We can take again the logics **S5** and **KD45** as $\boldsymbol{L}$, together with their different axiomatizations. In both cases we obtain the smallest cm-logic in $\mathsf{Cn}_{\diamond}\mathsf{S5}$, so also the smallest cm-logics defining the $\mathbf{D_2}$-consequence. We denote it by $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Notice that, selecting the standard axiomatization of **S5**, we have that $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ is the smallest cm-logic which

- Contains $(\mathtt{C})$, $(\mathtt{PN})$, $\ulcorner \Diamond\Box(\mathtt{df}\ \Diamond) \urcorner$, $\ulcorner \Diamond\Box(\mathtt{K}) \urcorner$, $\ulcorner \Diamond\Box(\mathtt{T}) \urcorner$ and $\ulcorner \Diamond\Box(\mathtt{5}) \urcorner$,
- And is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{nec})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Observe that

Fact 18. $(\mathtt{N}) \notin \mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash}$; *so* $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash} \subsetneq \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$.

Proof. Let v be a valuation from $\mathrm{For_m}$ into the set $\{0, 1\}$ such that v preserves classical truth conditions for classical constants and for any $A \in \mathrm{For_m}$: $v(\Diamond A) = 1$ and $v(\Box A) = 0$. For any $A \in \mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ we have $v(A) = 1$, while $v((\mathtt{N})) \neq 1$, thus $(\mathtt{N}) \notin \mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash}$. □

[12] For formulae having other forms one can take whatever as value of v. However, by the given restrictions we have that for any $A \in \mathrm{For_m}$: $v(\Box \neg A) = 1$ iff $\ulcorner \Diamond A \urcorner \notin \mathbf{S5}$.

6.5.3 The Case of Congruential Logics

If in Theorem 9 $\boldsymbol{X}$ is the set of all congruential logics, then we can assume that $\mathscr{A}_{\boldsymbol{X}} = \emptyset$ and $\mathscr{R}_{\boldsymbol{X}} = \{(\text{cgr})\}$. Once again applying as $\boldsymbol{L}$ different logics and their axiomatizations we obtain the smallest congruential logic in $\mathsf{Cn}_\diamond\mathsf{S5}$, i.e., the smallest congruential logic defining the $\mathbf{D_2}$-consequence. This time we denote it by $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$.

If we take the standard axiomatization of **S5** we obtain that $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ is the smallest congruential logic which

- Contains (C), $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$ and (PN) (or $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner$),
- And is closed under the rules $(\text{mp})^{\Diamond\Box}$ and $(\text{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Indeed, let $\boldsymbol{L}$ be the smallest congruential logic fulfilling the two above conditions. Since '$p \to p$', $(\mathtt{df}\,\Diamond)$ and (PN) (or $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner$) belong to $\boldsymbol{L}$, so by the Lemmas 11 and 12, $\boldsymbol{L}$ is closed under (poss-nec); so $\Diamond\Box\mathbf{PL} \subseteq \boldsymbol{L}$ and $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner \in \boldsymbol{L}$. Moreover, if $\ulcorner\Diamond\Box A\urcorner \in \boldsymbol{L}$, then $\ulcorner\Diamond\Box \Diamond \Box A\urcorner \in \boldsymbol{L}$, by (poss-nec). Hence, by $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$ and $(\text{mp})^{\Diamond\Box}$, we obtain that $\ulcorner\Diamond\Box\Box A\urcorner \in \boldsymbol{L}$. So $\boldsymbol{L}$ is closed under $(\text{nec})^{\Diamond\Box}$. Thus, $\boldsymbol{L} = \mathbf{eS5}^{\mathbf{M}}_{\vdash}$.

Fact 19.

(i) $(\mathtt{N}) \notin \mathbf{eS5}^{\mathbf{M}}_{\vdash}$. *Thus,* $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ *is not a cm-logic*; *so* $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash} \nsubseteq \mathbf{eS5}^{\mathbf{M}}_{\vdash}$.
(ii) *The formula*

$$\Box \Diamond \Box(p \to p) \leftrightarrow \Box(p \to p) \qquad (**)$$

does not belong to $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$. *Thus, the logics* $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ *and* $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ *are not congruential*; *so* $\mathbf{eS5}^{\mathbf{M}}_{\vdash} \nsubseteq \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ *and* $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{eS5}^{\mathbf{M}}_{\vdash}$.

Proof. (i) Since each congruential logic is regular, **CD45(1)** contains all specific axioms of $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ and is closed under the same rules as $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$, thus $\mathbf{eS5}^{\mathbf{M}}_{\vdash} \subseteq$ **CD45(1)**, but $(\mathtt{N}) \notin$ **CD45(1)**, so $(\mathtt{N}) \notin \mathbf{eS5}^{\mathbf{M}}_{\vdash}$. Therefore $\mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash} \nsubseteq \mathbf{eS5}^{\mathbf{M}}_{\vdash}$.

(ii) Consider a valuation v into $\{0, 1\}$, where v preserves the classical truth conditions for the classical constants and for any $A \in \mathrm{For_m}$:

$$v(\Box A) = \begin{cases} 1 & \text{if } A \in \mathbf{PL} \\ 0 & \text{otherwise.} \end{cases} \qquad v(\Diamond A) = \begin{cases} 0 & \text{if } \ulcorner\neg A\urcorner \in \mathbf{PL} \\ 1 & \text{otherwise.} \end{cases}$$

For for any $A \in \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$ we have $v(A) = 1$, so $(**) \notin \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$, but $(**) \in \mathbf{eS5}^{\mathbf{M}}_{\vdash}$. Thus, $\mathbf{eS5}^{\mathbf{M}}_{\vdash} \nsubseteq \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$. By the definition $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash} \subseteq \mathbf{eS5}^{\mathbf{M}}_{\vdash}$, hence $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{eS5}^{\mathbf{M}}_{\vdash}$, because $\mathrm{rte}\mathbf{S5}^{\mathbf{M}}_{\vdash} \subseteq \mathrm{cm}\mathbf{S5}^{\mathbf{M}}_{\vdash}$. □

6.5.4 The Case of Monotonic Logics

If in Theorem 9, $\boldsymbol{X}$ is the set of all monotonic logics, then we can assume that $\mathscr{A}_{\boldsymbol{X}} = \emptyset$ and $\mathscr{R}_{\boldsymbol{X}} = \{(\mathrm{mon})\}$. Once again, taking as $\boldsymbol{L}$ different logics and their axiomatizations we always obtain the smallest monotonic logic in $\mathsf{\mathbf{Cn}}_{\diamond}\mathsf{\mathbf{S5}}$, i.e., the smallest monotonic logic defining the $\mathbf{D_2}$-consequence. We denote it by $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$.

We obtain that $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$ is the smallest monotonic logic which

- Contains $(\mathtt{C})$, $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$,
- And is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Indeed, let $\boldsymbol{L}$ be the smallest monotonic logic fulfilling the two above conditions. Then, by Lemma 16, we obtain that $(\mathtt{PN}) \in \boldsymbol{L}$. Hence, by Lemma 12, $\boldsymbol{L}$ is closed under (poss-nec); so $\Diamond\Box\mathbf{PL} \subseteq \boldsymbol{L}$ and $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner \in \boldsymbol{L}$. Besides, $\boldsymbol{L}$ is closed under $(\mathrm{nec})^{\Diamond\Box}$, as in the case of congruential logics. So $\boldsymbol{L} = \mathbf{mS5}^{\mathbf{M}}_{\vdash}$.

Taking different axiomatizations of **S5**, **CD45(1)** and **KD45** one can obtain other axiomatizations of $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$. For example, $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$ is the smallest monotonic logic which contains $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{4}_{\mathtt{s}})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5}_{\mathtt{c}})\urcorner$, and is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{mon})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Fact 20. *The logic* $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ *is not monotonic*; *so* $\mathbf{eS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{mS5}^{\mathbf{M}}_{\vdash}$.

Proof. As in Nasieniewski and Pietruszczak (2011, Fact 3.4) let $v\colon \mathrm{For_m} \longrightarrow \{0,1\}$ be a valuation which preserves classical truth conditions for classical constants and such that for any A in $\mathrm{For_m}$:

$$v(\Diamond A) = \begin{cases} 1 & \text{if } \ulcorner\Diamond A\urcorner \in \mathbf{S5} \\ 1 & \text{if } \ulcorner\neg A\urcorner \in \mathbf{S5} \\ 0 & \text{otherwise.} \end{cases} \qquad v(\Box A) = \begin{cases} 0 & \text{if } \ulcorner\Diamond\neg A\urcorner \in \mathbf{S5} \\ 0 & \text{if } \ulcorner A\urcorner \in \mathbf{S5} \\ 1 & \text{otherwise.} \end{cases}$$

We prove that $v(\mathbf{eS5}^{\mathbf{M}}) = \{1\}$.

We can consider $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ as being axiomatized by **PL**, all substitutions of formulae $(\mathtt{C})$, $(\mathtt{df}\,\Box)$, $(\mathtt{PN})$, $\ulcorner\Diamond\Box(\mathtt{K})\urcorner$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$, and the rules (mp), $(\mathrm{cgr}'_{\Diamond})$, $(\mathrm{mp})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$. We prove by induction on the length of the proof, relative to the chosen axiomatization, that all theses of $\mathbf{eS5}^{\mathbf{M}}_{\vdash}$ are mapped by v on 1. We use the following facts (see Lemmas 14 and 15): (i) for any axiom A, $v(A) = 1$; (ii) $\mathbf{eS5}^{\mathbf{M}}_{\vdash} \subseteq \mathbf{S5}$; (iii) **S5** is closed under $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{cgr}'_{\Diamond})$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$; (iv) for the rules (mp), $(\mathrm{mp})^{\Diamond\Box}$, $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$ and $(\mathrm{cgr}'_{\Diamond})$: if all premisses are mapped on 1, then a conclusion is mapped on 1, too.

Ad (i), let us only observe that $v(\Diamond(\Diamond p \to q)) = 0$, thus $v((\mathtt{C})) = 1$.

Ad (iv), the case of $(\mathrm{cgr}'_{\Diamond})$: suppose that in a given proof $C_1, \ldots, C_n$ (for the chosen axiomatization) we obtain $C_j = \ulcorner\Diamond A \to \Diamond B\urcorner$ from $C_i = \ulcorner A \leftrightarrow B\urcorner$ by $(\mathrm{cgr}'_{\Diamond})$, for some $i < j \leqslant n$ and $A, B \in \mathrm{For_m}$. Then $\ulcorner A \leftrightarrow B\urcorner, \ulcorner\Diamond A \to \Diamond B\urcorner \in \mathbf{eS5}^{\mathbf{M}}_{\vdash} \subseteq \mathbf{S5}$. We show that $v(\Diamond A \to \Diamond B) = 1$. Indeed, if $v(\Diamond A) = 1$, then either $\ulcorner\Diamond A\urcorner \in \mathbf{S5}$ or $\ulcorner\neg A\urcorner \in \mathbf{S5}$. In the first case, $\ulcorner\Diamond B\urcorner \in \mathbf{S5}$; so $v(\Diamond B) = 1$. In the second case, $\ulcorner\neg B\urcorner \in \mathbf{S5}$, since $\ulcorner\neg A \leftrightarrow \neg B\urcorner \in \mathbf{S5}$. So also $v(\Diamond B) = 1$.

Now notice that $\ulcorner(p \wedge \neg p) \to q\urcorner \in \mathbf{Taut} \subseteq \mathbf{eS5^M}$, $v(\Diamond(p \wedge \neg p)) = 1$ and $v(\Diamond q) = 0$. So $v(\Diamond(p \wedge \neg p) \to \Diamond q) = 0$ and $\ulcorner\Diamond(p \wedge \neg p) \to \Diamond q\urcorner \notin \mathbf{eS5^M}$. Thus, $\mathbf{eS5^M}$ is not monotonic. □

6.5.5 The Case of Regular Logics

If in Theorem 9 $\boldsymbol{X}$ is the set of all regular logics, $\mathscr{A}_{\boldsymbol{X}} = \{(\mathtt{K})\}$ and $\mathscr{R}_{\boldsymbol{X}} = \{(\mathrm{mon})\}$, then we obtain the smallest regular logic in $\mathbf{Cn}_{\Diamond}\mathbf{S5}$, i.e., the smallest regular logic defining the $\mathbf{D_2}$-consequence. We denote it by $\mathbf{rS5}^{\mathbf{M}}_{\vdash}$.

Notice that—for the standard axiomatization of **S5**—we obtain that $\mathbf{rS5}^{\mathbf{M}}_{\vdash}$ is the smallest regular logic which

- Contains $(\mathtt{C})$, $\ulcorner\Diamond\Box(\mathtt{T})\urcorner$ and $\ulcorner\Diamond\Box(\mathtt{5})\urcorner$,
- And is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

Indeed, let $\boldsymbol{L}$ be the smallest regular logic fulfilling the two above conditions. Then, by Lemma 16, we obtain that $(\mathtt{PN}) \in \boldsymbol{L}$. Hence, by Lemma 12, $\boldsymbol{L}$ is closed under (poss-nec); so $\Diamond\Box\mathbf{PL} \subseteq \boldsymbol{L}$ and $\ulcorner\Diamond\Box(\mathtt{df}\,\Diamond)\urcorner$, $\ulcorner\Diamond\Box(\mathtt{K})\urcorner \in \boldsymbol{L}$. Besides, similarly as in the case of congruential logics, $\boldsymbol{L}$ is closed under $(\mathrm{nec})^{\Diamond\Box}$. So $\boldsymbol{L} = \mathbf{rS5}^{\mathbf{M}}_{\vdash}$.

By Theorems 9 and 5, Lemma 10, Corollary 3(ii), and Fact 14(ii), we obtain that $\mathbf{rS5}^{\mathbf{M}}_{\vdash} = \mathbf{CD45(1)} = \mathbf{rS5^M} \oplus (\mathtt{C})$.[13]

Fact 21. *The formula*

$$\Diamond\,(p \vee \neg\, p) \to (\Diamond p \vee \Diamond \neg\, p) \tag{+}$$

does not belong to $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$. *Thus, the logic* $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$ *is not regular*; *so* $\mathbf{mS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{rS5}^{\mathbf{M}}_{\vdash}$.

Proof. We consider a valuation v into $\{0, 1\}$, where v preserves classical truth conditions for classical constants and for any $A \in \mathrm{For_m}$:

$$v(\Diamond A) = \begin{cases} 1 & \text{if } \ulcorner\Diamond A\urcorner \in \mathbf{S5} \\ 0 & \text{otherwise.} \end{cases} \qquad v(\Box A) = \begin{cases} 0 & \text{if } \ulcorner\Diamond \neg\, A\urcorner \in \mathbf{S5} \\ 1 & \text{otherwise.} \end{cases}$$

Considering the axiomatization of $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$ given in the Sect. 6.5.4 we again can prove by induction on the length of the proof, relative to this axiomatization, that for all theses A of $\mathbf{mS5}^{\mathbf{M}}_{\vdash}$, $v(A) = 1$. Let us examine (mp) and $(\mathtt{C})$. The other cases can be easily seen by Corollary 4 and Lemma 15. Now, suppose that $\ulcorner A \to B\urcorner \in \mathbf{mS5}^{\mathbf{M}}_{\vdash}$ and $v(\Diamond A) = 1$. Then $\ulcorner A \to B\urcorner, \ulcorner\Diamond A\urcorner \in \mathbf{S5}$. Therefore $\ulcorner\Diamond B\urcorner \in \mathbf{S5}$

[13] Notice that in Nasieniewski and Pietruszczak (2008, pp. 202–203) it was proved that $\ulcorner\Diamond\Box(\mathtt{5})\urcorner \in \mathbf{rS5^M}$ and $\mathbf{rS5^M}$ is closed under $(\mathrm{mp})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$; so $\mathbf{rS5}^{\mathbf{M}}_{\vdash} \subseteq \mathbf{CD45(1)}$. Of course, the reverse inclusion can be shown elementarily.

and $v(\Diamond B) = 1$. Again we observe that $v(\Diamond(\Diamond p \to q)) = 0$, thus $v((\mathtt{C})) = 1$. Summarizing, for any $A \in \mathbf{mS5}^{\mathbf{M}}_{\vdash}$ we have: $v(A) = 1$. Besides $v(\Diamond(p \vee \neg p)) = 1$ and $v(\Diamond p) = 0 = v(\Diamond \neg p)$. So $v((+)) = 0$. Thus, $(+) \notin \mathbf{mS5}^{\mathbf{M}}_{\vdash}$. □

6.5.6 The Case of Normal Logics

For $\boldsymbol{X}$ being the set of all normal logics, $\mathscr{A}_{\boldsymbol{X}} = \{(\mathtt{K})\}$ and $\mathscr{R}_{\boldsymbol{X}} = \{(\mathrm{nec})\}$, applying Theorem 9 we obtain $\mathbf{nS5}^{\mathbf{M}}_{\vdash}$—the smallest normal logic in $\mathbf{Cn}_{\Diamond}\mathbf{S5}$, i.e., the smallest normal logic defining the $\mathbf{D_2}$-consequence.

For the standard axiomatization of **S5**—as in the case of regular logics—we obtain that $\mathbf{nS5}^{\mathbf{M}}_{\vdash}$ is the smallest normal logic which

- Contains $(\mathtt{C})$, $\ulcorner \Diamond\Box(\mathtt{T}) \urcorner$ and $\ulcorner \Diamond\Box(\mathtt{5}) \urcorner$,
- And is closed under the rules $(\mathrm{mp})^{\Diamond\Box}$ and $(\mathrm{cut}^{\Diamond\Box\Diamond}_{\Diamond})$.

By Theorems 9 and 5, Lemma 10, Corollary 3(i), and Fact 14(i) we have that $\mathbf{nS5}^{\mathbf{M}}_{\vdash} = \mathbf{KD45} = \mathbf{S5}^{\mathbf{M}} \oplus (\mathtt{C})$.

As it is shown, the logic $\mathbf{rS5}^{\mathbf{M}}_{\vdash} = \mathbf{CD45(1)}$ is not normal; so $\mathbf{rS5}^{\mathbf{M}}_{\vdash} \subsetneq \mathbf{KD45}$.

Acknowledgements We would like to thank both anonymous referees for their valuable remarks on an earlier version of this paper.

Appendix: Some Facts from Modal Logic

As in Chellas (1980) modal formulae are formed in the standard way from propositional letters: 'p', 'q', 'p_0', 'p_1', 'p_2', ...; truth-value operators: '$\neg$', '$\vee$', '$\wedge$', '$\to$', and '$\leftrightarrow$' (connectives of negation, disjunction, conjunction, material implication, and material equivalence, respectively); modal operators: the necessity sign '$\Box$' and the possibility sign '$\Diamond$'; and brackets. By $\mathrm{For_m}$ we denote the set of modal formulae. Of course, the set $\mathrm{For_m}$ includes the set of all classical formulae (without '$\Box$' and '$\Diamond$'); let **Taut** be the set of all classical tautologies. Besides, for any $A, B, C \in \mathrm{For_m}$, let $C[{}^{A}/_{B}]$ be any formula that results from C by replacing one or more occurrences of A, in C, by B.

For any subset Φ of $\mathrm{For_m}$ we put $\Box\Phi := \{\ulcorner \Box A \urcorner : A \in \Phi\}$ and $\Diamond\Phi := \{\ulcorner \Diamond A \urcorner : A \in \Phi\}$.

Modal logics are certain sets of formulae. As in Bull and Segerberg (1984), we define a *modal logic* as a set $\boldsymbol{L}$ of modal formulae satisfying the following conditions:

- $\mathbf{Taut} \subseteq \boldsymbol{L}$,
- $\boldsymbol{L}$ includes the following set of formulae

$$\left\{ \ulcorner C[^{\neg\Box\neg A}/_{\Diamond A}] \leftrightarrow C\urcorner \;:\; A, C \in \mathrm{For_m} \right\}. \qquad (\mathtt{rep}^{\Box})$$

- $\boldsymbol{L}$ is closed under the following two rules: *modus ponens* for '→':

$$\frac{A \qquad A \to B}{B} \qquad (\mathrm{mp})$$

and *uniform substitution*:

$$\frac{A}{s\,A} \qquad (\mathrm{sb})$$

where $s\,A$ is the result of uniform substitution of formulae for propositional letters in A.

Of course, by (sb), every modal logic includes the set **PL** of modal formulae which are instances of classical tautologies (i.e. instances of elements of **Taut**).

All members of a logic are called its *theses*. By ($\mathtt{rep}^{\Box}$), every modal logic has the following thesis:

$$\Diamond\, p \leftrightarrow \neg\Box\neg\, p \qquad (\mathtt{df}\,\Diamond)$$

Remark 2. In Bull and Segerberg (1984) the symbol '◇' is only an abbreviation of '¬ □ ¬'. In the present paper '◇' is a primary symbol, thus we have to add the set of axioms ($\mathtt{rep}^{\Box}$). The use of this set corresponds to the applying of the formula ($\mathtt{df}\,\Diamond$) as a definition together with the definitional rule. Formulae from ($\mathtt{rep}^{\Box}$) allow to replace one or more occurrences of '¬ □ ¬' with '◇' and vice versa. □

Lemma 8. *A logic contains the formula*:

$$\Box p \to p \qquad (\mathtt{T})$$

iff it contains its dual version:

$$p \to \Diamond p \qquad (\mathtt{T}^{\Diamond})$$

We say that a modal logic $\boldsymbol{L}$ is *rte-logic* iff $\boldsymbol{L}$ is closed under replacement of tautological equivalents, i.e., for any $A, B, C \in \mathrm{For_m}$

$$\text{if } \ulcorner A \leftrightarrow B\urcorner \in \mathbf{PL} \text{ and } C \in \boldsymbol{L}, \text{ then } C[^A/_B] \in \boldsymbol{L}. \qquad (\mathrm{rte})$$

A modal logic is rte-logic iff it includes the following set

$$\left\{ \ulcorner C[^A/_B] \leftrightarrow C\urcorner : A, B, C \in \mathrm{For_m} \text{ and } \ulcorner A \leftrightarrow B\urcorner \in \mathbf{PL} \right\}. \qquad (\mathtt{rep}_{\mathbf{PL}})$$

In any thesis of any rte-logic we can replace one or more occurrences of '$\neg\Box\neg$' (resp. '$\Box\neg$', '$\neg\Box$', '$\neg\Diamond\neg$', '$\neg\Diamond$', '$\Diamond\neg$') by '$\Diamond$' (resp. '$\neg\Diamond$', '$\Diamond\neg$','$\Box$', '$\Box\neg$', '$\neg\Box$') and vice versa. Thus, every rte-logic has the following thesis

$$\Box p \leftrightarrow \neg\Diamond\neg p \tag{$\mathtt{df}\,\Box$}$$

Lemma 9. *An rte-logic contains, respectively, the following formulae*:

$$\Box(p \to q) \to (\Box p \to \Box q) \tag{$\mathtt{K}$}$$

$$\Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q) \tag{$\mathtt{R}$}$$

$$\Box p \to \Box\Box p \tag{$\mathtt{4}$}$$

$$\Diamond\Box p \to p \tag{$\mathtt{B}$}$$

$$\Diamond\Box p \to \Box p \tag{$\mathtt{5}$}$$

$$\Box p \to \Diamond\Box p \tag{$\mathtt{5_c}$}$$

iff it contains, respectively, theirs dual versions:

$$\Box(p \to q) \to (\Diamond p \to \Diamond q) \tag{$\mathtt{K}^\Diamond$}$$

$$\Diamond(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q) \tag{$\mathtt{R}^\Diamond$}$$

$$\Diamond\Diamond\, p \to \Diamond p \tag{$\mathtt{4}^\Diamond$}$$

$$p \to \Box\Diamond p \tag{$\mathtt{B}^\Diamond$}$$

$$\Diamond p \to \Box\Diamond p \tag{$\mathtt{5}^\Diamond$}$$

$$\Box\Diamond\, p \to \Diamond p \tag{$\mathtt{5}^\Diamond_\mathtt{c}$}$$

In Bull and Segerberg (1984) a modal logic is called *classical modal*[14] (*cm-logic* for short) iff it is rte-logic which contains (K) and

$$\Box(p \to p) \tag{$\mathtt{N}$}$$

Thus, all cm-logics include the set $\Box$**PL**.

We say that a modal logic is *congruential* iff it is closed under the congruence rule

$$\frac{A \leftrightarrow B}{\Box A \leftrightarrow \Box B} \tag{cgr}$$

[14] In Nasieniewski and Pietruszczak (2008, 2009, 2013), following a custom from Chellas (1980), the expression 'classical modal' was referred to congruential logics (please see further).

A modal logic is congruential iff it is closed under replacement:

$$\frac{A \leftrightarrow B}{C[{}^{A}/_{B}] \leftrightarrow C} \tag{rep}$$

iff it contains (df $\Box$) and is closed under the following rule

$$\frac{A \leftrightarrow B}{\Diamond A \to \Diamond B} \tag{cgr$'_\Diamond$}$$

Lemma 10. *Every modal logic closed under* (rep) *includes the set* ($\mathtt{rep}_{\mathbf{PL}}$); *so every congruential logic is an rte-logic.*

Besides (N), we use the following formulae

$$\Diamond(p \to p) \tag{P}$$

$$\Diamond\Box(p \to p) \tag{PN}$$

Lemma 11. *For any congruential logic* $\mathbf{L}$ *the following conditions are equivalent*:

(a) $\mathbf{L}$ *has a pair of theses of the form* B *and* $\ulcorner \Box B \urcorner$ (*resp.* $\ulcorner \Diamond B \urcorner$, $\ulcorner \Diamond\Box B \urcorner$),
(b) $\mathbf{L}$ *has a thesis of the form* $\ulcorner \Box T \urcorner$ (*resp.* $\ulcorner \Diamond T \urcorner$, $\ulcorner \Diamond\Box T \urcorner$), *where* $T \in \mathbf{PL}$,
(c) $\mathbf{L}$ *contains* (N) (*resp.* (P), (PN)).

Lemma 12. *If a congruential logic contains* (N) (*resp.* (P), (PN)), *then it is closed under the necessity (resp. possibility, possibility-necessity) rule*:

$$\frac{A}{\Box A} \tag{nec}$$

$$\frac{A}{\Diamond A} \tag{poss}$$

$$\frac{A}{\Diamond\Box A} \tag{poss-nec}$$

Lemma 13. *If a congruential logic has theses* (5) *and either*

- *A theses of the form* $\ulcorner \Diamond B \urcorner$, *or*
- (T),

then it is closed under the rule (nec).

It is known (cf. e.g. Chellas 1980) that while defining congruential logics one uses (df $\Diamond$) instead of ($\mathtt{rep}^{\Box}$), i.e., treats them as subsets of $\mathrm{For_m}$ which contain **Taut** and (df $\Diamond$), and are closed under the rules (mp), (sb) and (cgr).[15]

[15] We can also consider quite weak modal logics in which we use (df $\Diamond$) instead of ($\mathtt{rep}^{\Box}$). In some logics the symbol '$\Diamond$' has not to behave as an abbreviation of '$\neg\Box\neg$', although we can

Lemma 14. *Let $\boldsymbol{L}$ be any modal logic closed under* (nec)*, and containing* (T) *and* (5)*. Then for any $A \in \mathrm{For_m}$: $A \in \boldsymbol{L}$ iff $\ulcorner \Diamond\Box A \urcorner \in \boldsymbol{L}$.*

Lemma 15. *Let $\boldsymbol{L}$ be any modal logic containing* (T)*,* (5)*,* ($5^{\Diamond}$)*,* (4) *and* ($4^{\Diamond}$)*. Then for any sequences of modal operators $\mathscr{M}$ and $\mathscr{M}'$, and any $A \in \mathrm{For_m}$: $\ulcorner \mathscr{M} \Diamond A \urcorner \in \boldsymbol{L}$ iff $\ulcorner \Diamond A \urcorner \in \boldsymbol{L}$ iff $\ulcorner \mathscr{M}' \Diamond A \urcorner \in \boldsymbol{L}$.*

We say that a modal logic is *monotonic* iff it is closed under the monotonicity rule:

$$\frac{A \to B}{\Box A \to \Box B} \tag{mon}$$

Of course, every monotonic logic is closed under (rep) and under the dual form of (mon):

$$\frac{A \to B}{\Diamond A \to \Diamond B} \tag{mon$_\Diamond$}$$

Lemma 16. *For any monotonic logic the following conditions are equivalent*:

(a) It has at least one thesis of the form $\ulcorner \Box B \urcorner$ (resp. $\ulcorner \Diamond B \urcorner$, $\ulcorner \Diamond\Box B \urcorner$),
(b) It contains (N) *(resp.* (P)*,* (PN)*).*

We say that a modal logic is *regular* iff it is closed under the regularity rule:

$$\frac{A \wedge B \to C}{\Box A \wedge \Box B \to \Box C} \tag{reg}$$

A modal logic is regular iff it contains (K) and is closed under (mon) iff it contains (R) and is closed under (cgr). Every regular logic has the theses ($\mathrm{R}^{\Diamond}$) and

$$\Diamond\,(p \to q) \leftrightarrow (\Box p \to \Diamond q) \tag{R$^{\Diamond\Box}$}$$

By ($\mathrm{R}^{\Diamond\Box}$) we obtain.

Lemma 17. *For any regular logic the following conditions are equivalent*:

(a) It has at least one thesis of the form $\ulcorner \Diamond B \urcorner$,
(b) It contains (P)*,*
(c) It contains the following formula

$$\Box p \to \Diamond p \tag{D}$$

have there the thesis (df □) (cf. Remark 2). For example, the formula '$\Box\Diamond p \leftrightarrow \Box\neg\Box\neg p$' has not to be a thesis of such logics.

A modal logic is *normal* iff it contains (K) and is closed under (nec) iff it is regular and contains (N) iff it contains (N) and (K) and is closed under (cgr). Thus, all normal logics are cm-logics.

Let **K** (resp. **C2**) be the smallest normal (resp. regular) modal logic. Using names of the above formulae, to simplify notation of normal (resp. regular) logics we write the *Lemmon code* $\mathbf{KX}_1 \ldots \mathbf{X}_n$ (resp. $\mathbf{CX}_1 \ldots \mathbf{X}_n$) to denote the smallest normal (resp. regular) logic containing the formulae $(\mathbf{X}_1), \ldots, (\mathbf{X}_n)$ (see Bull and Segerberg 1984; Chellas 1980; Segerberg 1971). Besides, let for any formula $A \in \mathrm{For}_{\mathrm{m}}$, $\mathbf{KX}_1 \ldots \mathbf{X}_n \oplus A$ (resp. $\mathbf{CX}_1 \ldots \mathbf{X}_n \oplus A$) be the smallest normal (resp. regular) logic which includes $\mathbf{KX}_1 \ldots \mathbf{X}_n$ (resp. $\mathbf{CX}_1 \ldots \mathbf{X}_n$) and contains A.

Lemma 18. *(i)* (D) $\in \mathbf{C5_c} \subseteq \mathbf{K5_c}$.
(ii) (5_c) $\in \mathbf{CD4} \subseteq \mathbf{KD4}$.
(iii) $\mathbf{KD4} = \mathbf{K5_c4}$ *and* $\mathbf{CD4} = \mathbf{C5_c4}$.

Proof. (i) '$\Diamond(p \to \Box p)$' belongs to $\mathbf{C5_c}$, by ($\mathrm{R}^{\Diamond\Box}$). So, we use Lemma 17.
(ii) By (4), (sb), (D) and **PL** we obtain (5_c).
(iii) By (i) and (ii). □

The logic **CF**, called *falsum*, is the smallest regular logic containing the following formula

$$\Diamond\,(p \wedge \neg\, p) \tag{F}$$

We have $\ulcorner \Diamond A \urcorner \in \mathbf{CF}$, for any $A \in \mathrm{For}_{\mathrm{m}}$.

We standardly put $\mathbf{T} := \mathbf{KT}$, $\mathbf{S4} := \mathbf{KT4}$ and $\mathbf{S5} := \mathbf{KT5} = \mathbf{KT4B} = \mathbf{KD4B} = \mathbf{KD5B}$. As it is known, $\mathbf{T} \subsetneq \mathbf{S4} \subsetneq \mathbf{S5}$, $\mathbf{CD5} = \mathbf{KD5}$, $\mathbf{CD45} = \mathbf{KD45}$ and $\mathbf{CT5} = \mathbf{KT5} =: \mathbf{S5}$. Thus, to avoid "normalization" of regular logics one has to use some special formulae. We adopt a convention from Segerberg (1971, p. 206) and for any formula (X) we put (X(1)) $:= \ulcorner \Box(p \to p) \to (\mathrm{X}) \urcorner$. Notice that in all monotonic logics, any formula of the form $\ulcorner \Box A \to B \urcorner$ is equivalent to $\ulcorner \Box(p \to p) \to (\Box A \to B) \urcorner$. Thus, the formulae (T), (D), (4) and (5_c) are respectively equivalent to (T(1)), (D(1)), (4(1)) and (5_c(1)).

Lemma 19. *The following formulae*:

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

$$\Diamond(\Diamond p \to (\Diamond q \to (p \wedge \Diamond q)))$$

$$\Diamond(\Diamond p \to q) \leftrightarrow (\Diamond p \to \Diamond q)$$

belong to **S5**, *as well as for any* $n > 0$ *the following formula*:

$$\begin{aligned}\Diamond(\Diamond p_1 \to (\Diamond p_2 \to \ldots (\Diamond p_n \to q) \ldots))\\ \leftrightarrow (\Diamond p_1 \to (\Diamond p_2 \to \ldots (\Diamond p_n \to \Diamond q) \ldots))\end{aligned}$$

Lemma 20. *If* (C)*, i.e.* '$\Diamond(\Diamond p \to q) \to (\Diamond p \to \Diamond q)$'*, belongs to* $\boldsymbol{L}$*, then for any* $n > 0$ *the following formula belongs to* $\boldsymbol{L}$:

$$\Diamond(\Diamond p_1 \to (\Diamond p_2 \to \ldots (\Diamond p_n \to q)\ldots)) \\ \to (\Diamond p_1 \to (\Diamond p_2 \to \ldots (\Diamond p_n \to \Diamond q)\ldots))$$

Proof. By induction on the number of propositional letters in a formula. It the case of two letters we just have (C). For the inductive step, by (C) and (sb) we have also:

$$\Diamond(\Diamond p_1 \to (\Diamond p_2 \to \ldots (\Diamond p_n \to q)\ldots)) \\ \to (\Diamond p_1 \to \Diamond(\Diamond p_2 \to \ldots (\Diamond p_n \to q)\ldots))$$

By the inductive hypothesis we have

$$\Diamond(\Diamond p_2 \to \ldots (\Diamond p_n \to q)\ldots) \to (\Diamond p_2 \to \ldots (\Diamond p_n \to \Diamond q)\ldots)$$

The thesis follows by **PL**. □

Lemma 21 (Segerberg 1971, vol. II, Corollary 2.4). $\mathbf{CN^1X_1(1)\ldots X_{\mathit{n}}(1)} = \mathbf{CF} \cap \mathbf{KX_1\ldots X_{\mathit{n}}}$*, where*

$$\Box(p \to p) \to \Box\Box(p \to p) \tag{N^1}$$

In Segerberg (1971), Segerberg puts **E5** := $\mathbf{CN^1T4B(1)}$. So **E5** = **CF**∩**KT4B** = **CF** ∩ **S5**, by Lemma 21. Notice that **E5** = **CT4B(1)**, since (N^1) is an instance of (4). We have also **E5** = $\mathbf{CN^1T5(1)}$ and **E5** = **CF** ∩ **KT4B** = **CF** ∩ **KD4B** = **CD4B(1)**.[16] Moreover, notice that **CD45(1)** = $\mathbf{CN^1D45(1)}$ = **CF** ∩ **KD45**.

For any modal logic $\boldsymbol{L}$ we define the relation of consequence $\vdash_{\boldsymbol{L}}$ with the help of modus ponens for '→' as the only rule of inference, i.e., $\vdash_{\boldsymbol{L}}$ is the pure modus-ponens-style inference relation based on $\boldsymbol{L}$. For any $\Pi \subseteq \mathrm{For_m}$ and $B \in \mathrm{For_m}$:

$\Pi \vdash_{\boldsymbol{L}} B \overset{\mathrm{df}}{\iff}$ there exists a sequence $A_1, \ldots, A_n = B$ in which for any $i \leqslant n$, either $A_i \in \Pi$, or $A_i \in \boldsymbol{L}$, or there are $j, k < i$ such that $A_k = \ulcorner A_j \to A_i \urcorner$.

Lemma 22 (Lemmon and Scott 1977). $\Pi \vdash_{\boldsymbol{L}} B$ *iff for some* $n \geqslant 0$ *and for some* $A_1, \ldots, A_n \in \Pi$ *we have* $\ulcorner A_1 \to (\ldots \to (A_n \to B)\ldots)\urcorner \in \boldsymbol{L}$*, or equivalently* $\ulcorner(A_1 \wedge \cdots \wedge A_n) \to B\urcorner \in \boldsymbol{L}$.

[16] In Segerberg (1971), Segerberg also puts **D5** := $\mathbf{CN^1D4B(1)}$ = **CD4B(1)**. So we have **D5** = **E5**.

References

Bull, R. A., & Segerberg, K. (1984). Basic modal logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. II, pp. 1–88). Dordrecht: Reidel.

Chellas, B. F. (1980). *Modal logic: An introduction*. Cambridge: Cambridge University Press.

Jaśkowski, S. (1948). Rachunek zdań dla systemów dedukcyjnych sprzecznych. *Studia Societatis Scientiarum Torunensis*, Sect. A, *I*(5), 57–77. The first English version: Propositional calculus for contradictory deductive systems. *Studia Logica, 24*, 143–157 (1969).

Jaśkowski, S. (1999a). A propositional calculus for inconsistent deductive systems. *Logic and Logical Philosophy, 7*, 35–56. The second English version of Rachunek zdań dla systemów dedukcyjnych sprzecznych.

Jaśkowski, S. (1999b). On the discussive conjunction in the propositional calculus for inconsistent deductive systems. *Logic and Logical Philosophy, 7*, 57–59. The English version of O koniunkcji dyskusyjnej w rachunku zdań dla systemów dedukcyjnych sprzecznych, *Studia Societatis Scientiarum Torunensis*, Sect. A, Vol. I, no. 8, 171–172 (1949).

Lemmon, E. J., & Scott, D. (1977). *"Lemmon Notes": An introduction to modal logic* (American philosophical quarterly monograph series). Oxford: Basil Blackwell.

Nasieniewski, M., & Pietruszczak, A. (2008). The weakest regular modal logic defining Jaśkowski's logic D_2. *Bulletin of the Section of Logic, 37*(3/4), 197–210.

Nasieniewski, M., & Pietruszczak, A. (2009). New axiomatizations of the weakest regular modal logic defining Jaśkowski's logic D_2. *Bulletin of the Section of Logic, 38*(1/2), 45–50.

Nasieniewski, M., & Pietruszczak, A. (2011). A method of generating modal logics defining Jaśkowski's discussive logic D_2. *Studia Logica, 97*(1), 161–182.

Nasieniewski, M., & Pietruszczak, A. (2012). On the weakest modal logics defining Jaśkowski's logic D_2 and the D_2-consequence. *Bulletin of the Section of Logic, 41*(3/4), 215–232.

Nasieniewski, M., & Pietruszczak, A. (2013). On modal logics defining Jaśkowski's D_2-consequence. In K. Tanaka, F. Berto, E. Mares, & F. Paoli (Eds.), *Paraconsistency: Logic and applications* (Logic, epistemology and the unity of science, chap. 8, Vol. 26, pp. 141–161). Dordrecht/New York: Springer.

Perzanowski, J. (1975). On M-fragments and L-fragments of normal modal propositional logics. *Reports on Mathematical Logic, 5*, 63–72.

Segerberg, K. (1971). *An essay in classical modal logic* (Vols. 1 and 2). Uppsala: Department of Philosophy, Uppsala University.

Chapter 7
Frontier Theory of Inquiry: Apparent Conflicts Between the Ghent Logical Program and the “Darwinian” Selectionist Program

Thomas Nickles

7.1 (Auto)biographical Introduction

Diderik Batens and I have shared a strong interest in scientific discovery and invention from the beginning of our careers. We both believed (and continue to believe) that scientific practices at the frontiers of research are of great interest, not only to psychologists and historians but also to philosophers. Or at least they should be, for, as Karl Popper wrote, “the problem of the growth of knowledge” is the central problem of epistemology.[1] This means getting more from less, that is, parlaying what we already think we know, and know how to do, into something essentially new and original, without any help from transcendent “givens.” Yet Popper, together with the logical empiricists (with which he differed on other points), was among those who declared the “context of discovery” (better termed “context of innovation”[2]) off limits to philosophy. Meanwhile, both Batens and I

[1] See Popper (1963, 1972). Arguably, the problem goes all the way back to Plato’s Meno paradox of how inquiry is possible. Many prominent investigators, including Herbert Simon (1976), have recognized the importance of the paradox. For my take on the situation, see Nickles (2003a).

[2] Sociologists of science usually reject talk of “discovery” because, to them, it smacks of strong realism. I do not use the term in that way but only as an identifier for a subject area with a traditional label. (Even if our models, theories, and techniques are complete fabrications, it is still an interesting set of problems how scientists (and artists, etc.) construct them and get them to work as well as they sometimes do.) Given that sociologists of science regularly speak of ‘knowledge’ in a non-normative manner, philosophers need not apologize for using the term ‘discovery’ similarly (i.e., relative only to the norms of that day). Yet, to avoid controversy, I prefer the neutral terms ‘innovation’ and ‘context of innovation’. Ironically, it is the science studies disciplines other than philosophy that filled the void created by traditional philosophers of science, a void that includes detailed scientific practices generally, practices that earlier generations of philosophers tended to

T. Nickles (✉)
Department of Philosophy, University of Nevada, Reno, NV, USA
e-mail: nickles@unr.edu

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic,
Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_7,
© Springer Science+Business Media Dordrecht 2014

saw the emergence of promising new candidate knowledge structures and processes, rather than post hoc confirmation, as the epistemological mystery that most needed attention.[3] Both of us found what is today called analytic epistemology a rather dry subject that does not address how breakthroughs or even incremental progress at the frontiers of research are possible. However, neither of us believed in the existence of a formal logic of discovery. The idea that there is a known algorithm-like process, "the scientific method," that somehow harbors all future discoveries, is a strange notion indeed (Nickles 2009). For scientific work, like innovative work in the arts, is often genuinely creative. Both of us believed that this made the problem all the more interesting and important and that attention to flexible forms of reasoning was needed. Speaking of problems, we both take a problem-solving approach to science, and we resonate to similar views about how problems can be characterized in terms of the constraints on reaching a desired target.

At roughly the same time that Batens made his major discovery of adaptive logics (partly in order to do greater justice to contexts of innovation), I organized a conference on scientific discovery out of my small department in Reno. The two volumes that resulted (Nickles 1980b,c) soon caught Batens' attention. We met when I was invited to speak at the Ghent meeting of the International Union of History and Philosophy of Science in 1986. I have been fortunate to interact with Batens, Joke Meheus, Erik Weber, and other members of the Ghent logic group and the Philosophy Department on many occasions since then.[4]

Despite our common interests, it has at times appeared to Batens, Meheus, and me that our views had increasingly diverged over the years, especially since about 1997, when I became convinced that a selectionist paradigm such as that advanced by Donald T. Campbell was the only defensible way to understand innovation at all levels. For Batens and his group take a formal logical approach, whereas the selectionist program looks to many to be anti-logical and anti-methodological. Campbell himself seemed to think so. Skeptics of the selectionist paradigm are tempted to think of a thousand monkeys at typewriters!

My contribution to this volume is an attempt to show that the gap between these two broad approaches is not as wide as we once believed, that it can be closed to a

dismiss as "test tube washing." In excluding context of innovation from epistemology, philosophers simply "gave away" what should have been one of their central topics.

[3] I would now say that the problem of explaining the rapid diversification of the sciences is not unlike that of explaining evolutionary design, the problem raised in sharp form by William Paley (1802/2008), appreciated by Darwin, and eventually solved by Darwin in a manner that inverted Paley's solution of top-down design by the Great Artificer (who is credited with already having all possible designs at his disposal), namely, by explaining how to get more design than less instead of less from more (Nickles 2003a).

[4] I shall often refer to Batens and Meheus together, since they have jointly authored several important papers on our topic. A third member of the Ghent group is Erik Weber who has written several articles defending a "sophisticated pragmatism," e.g., Weber and Vanderbeeken (2001) and Van Bouwel and Weber (2008).

significant degree, although perhaps not completely.[5] We certainly share the same overall motivation. Both the Ghent program and the explicitly selectionist approach attempt to do justice to the uncertainties that everyone confronts at the frontiers of scientific research. We are all concerned with what might be termed "frontier epistemology" or (better) "frontier theory of inquiry," as opposed to completely routine problem solving. But there is more to say.

In Sect. 7.2 I shall characterize Batens' position as a variety of pragmatism, the philosophical "school" that has taken frontier inquiry more seriously than any other. Then, after providing an instructive bit of history from artificial intelligence in Sect. 7.3, I shall point out, in Sect. 7.4, that crucial variation-selection steps occur in the Batens logical program. Or so I shall claim. Section 7.5 ends the chapter by raising some questions for future exploration.

7.2 Locating the Batens Research Program: A Kind of Pragmatism

Whether or not Diderik Batens will accept my characterization of him as a pragmatist, I am not sure, but it seems to me that the designation fits well enough and helps us to appreciate the unity of his thought and work. (Keep in mind that pragmatism is a "low-church" philosophy, not an ideology, and thus welcomes a variety of thinkers.) For Batens the logician espouses a wonderful pragmatic flexibility and a sense of history. He opposes radical purity movements, whether doctrinal (as in epistemological foundationism), methodological (as in the idea of a quasi-algorithmic scientific method originally inspired by Bacon, Descartes, and Newton) or conceptual (as in operationism and the "clarity is enough" therapeutic approach), not to mention ideological political and religious movements. Although obviously fascinated by the properties of formal systems, his goal has always been to bring logic out of the clouds, closer to the way people actually reason, not only in ordinary conversation but also at the frontiers of scientific research and other creative venues. And he insists that the logics in question be understood naturalistically, as evolved tools for thinking and acting.[6] Gone are a priori justifications and transcendent appeals to certainty. Batens and his colleagues are strong fallibilists, a fallibilism that does not stop with descriptive belief commitments but extends to tools and norms, much as for the pragmatist John Dewey. Batens believes that an erotetic, problem-solving approach to theory of inquiry is the best way to go,[7]

[5] And it is thereby also an attempt to reconcile my own present with my past!

[6] In Meheus (1999b), the word 'tools' is used in its title: "Deductive and Ampliative Adaptive Logics as Tools in the Study of Creativity." Pragmatists employ the word 'tools' in a broad sense.

[7] Batens' erotetic approach is quite evident, e.g., in the recent (Batens 2007), where he also makes many references to the work of Andrzej Wiśniewski.

that methodological tools are context-relative, and that methods of inquiry should be prospective as well as retrospective, that they should provide some direction to future inquiry (e.g., Batens 1999, 2004, 2007, 2008). The context-relativity includes relativity to human purposes. There is a strong humanistic component to his research program.

Let us return to the problem situation in the decades following World War II. Given that the exciting new symbolic logic embraced by the logical empiricists (and somewhat by Popper) was what we today call classical logic (**CL**), and that they limited rational inquiry to what could be expressed, in principle, in terms of **CL**, it is not surprising that those founders of modern philosophy of science should have demoted "context of discovery" in the way that they did. Popper (and some of the logical empiricists) possessed a romantic conception of discovery as major breakthroughs that turn on opening up new worldviews that cannot be reached by reasoning logically or methodically from the old theory and its problems. A more critical take on this work is to see their denying the logical and epistemological interest of discovery contexts as an admission that their tools were not up to the task. To be fair to them, how leading a role logic plays in major breakthroughs remains an open question within the field of philosophy of science.

The emergence of the historical approach to philosophy of science in the 1960s took context of innovation more seriously, but, in so doing, it also demoted logic, calling into question whether there is even a logic of justification. And in *The Structure of Scientific Revolutions* (Kuhn 1962), Thomas Kuhn, in effect, promoted rhetoric to a position as important as logic, both in normal scientific practice and in revolutionary breakthroughs (Nickles 2003b), a point to which I shall return. In speaking of revolutions and incommensurability, both Kuhn and Paul Feyerabend were sometimes labeled irrationalists. Both men rejected traditional conceptions of rationality, supposedly crystallized in classical logic (**CL**), as too conservative to capture how scientists either do or ought to think. And both men totally rejected the idea of a monolithic scientific method. So, for Batens, a young logician interested in scientific discovery, it was hardly reassuring to hear that "justification" was now in the same boat as "discovery," given that the relevance of formal logic to both was in doubt. To be sure, Kuhn attempted to bring "context of discovery" back into philosophical discussion (Kuhn 1962). But while he did make progress in dealing with the tamed version of it in normal science, he treated the construction of, and conversion to, a new paradigm as still too much of an "aha" experience.

Batens' response to these challenges to logic was to argue that, even at the frontiers of inquiry, scientists and other creative individuals and communities are engaged in forms of reasoning that, to an interesting degree, can be formalized. Rather than establishing the unsuitability of formal logic, he believed, the historical turn challenged formal logicians to do more justice to actual reasoning practices to the reasoning involved in innovative scientific work as well as in everyday life. The trouble, he concluded, was not with formal logic per se but with trying to do everything in terms of classical logic, **CL**, as if there were one universal logic (and one already known to us!) that performs equally well in all contexts.[8] Much as Bacon and Descartes criticized the old syllogistic logic of their day, Batens and Meheus, in developing dynamic logics, have characterized **CL** as static and sterile.[9] It mainly helps us to organize what we already know.[10]

In those days many commentators (including myself in my introduction to Nickles 1980c) spoke of broadening the conception of rationality in order to deal with these problems in both context of discovery and context of justification, but Batens went further. He made it his research program to develop more liberal, pragmatic conceptions of logic itself, whereby logical systems are tools, alongside other tools, rather than a single, a priori system definitory of rational thinking and action. His concern was, and remains, to blur the distinction between formal logic and theory of argumentation, to show that we can go between the horns of the old debate between rigid formalists and equally rigid informalists. Batens (1996) and both Batens and Meheus in articles since then go even further in beginning to break down the traditional divide between logic and rhetoric to some degree. Here the reader will recall that skepticism of received dichotomies is a prominent characteristic of pragmatic approaches.

[8] Unlike prominent theorists such as Stephen Toulmin (1958) and Toulmin et al. (1979), Batens wanted to carry the program of formal logic as far as possible, thereby saving as many of the "phenomena" of informal reasoning as possible.

[9] The basic complaint of Bacon and Descartes was that the old logic merely organized what we already know rather than producing new knowledge. Descartes then turned to mathematics, in which field he himself was quite inventive (Heeffer 2008). Much the same can be said for **CL**. According to Meheus:

> *[C]lassical* logic does not provide insight in the reasoning involved in creative processes. It is not difficult to understand why. Problems that give rise to creativity are always *ill-defined*. This means that the information from which the solution should follow is incomplete or inconsistent. In such cases, classical logic does not render a sensible distinction between sound and unsound steps. (Meheus 1999b, p. 125) . . . Classical logic, as well as most other available logical systems, is *static: if a sentence is derived at some stage in a proof, it cannot be withdrawn at a later stage*. [pp. 325, 329, emphasis in original] (Meheus 1999b)

By the way, in complaining about the mere organization of knowledge, Bacon and Descartes went too far, for the organization of knowledge becomes increasingly important to ongoing inquiry as the stock of presumed knowledge grows. I term this the "knowledge pollution" problem. Derek Price noted, for example, that the size of the *Philosophical Transactions* of the Royal Society of London had grown exponentially over time (Price 1963). The same is true of later journals such as *The Physical Review*. No practitioner has time to scan all the articles in her field and neighboring fields, so selecting papers and information that are relevant to your project becomes a tricky business of expert judgment. Another sort of knowledge pollution arises from the fact that we expect current knowledge claims to harbor many errors, even inconsistencies. Many papers of Batens and Meheus are relevant to this problem.

[10] Of course, our situation today is vastly different from theirs. Given the explosion of results and techniques from dozens of different scientific specialties and subspecialties, the efficient organization and retrieval are important to ongoing research.

One central concern of the Ghent program is to provide logical tools for handling inconsistency. Inconsistencies frequently arise in our reasoning. According to **CL**, as everyone knows, we can immediately derive anything at all from an inconsistency. For this reason the logical empiricists and other traditional formalists imposed the rule that formal explications of fruitful reasoning must avoid inconsistency at all costs. For them inconsistency was the worst intellectual sin of all—as if inconsistency always blocks Charles Peirce's "road to inquiry" and stops rational debate of any kind. Yet, quite obviously, neither scientists nor ordinary people of good will actually do knowingly use an inconsistency as a license to conclude just anything they like. Moreover, at the frontiers of innovation it is hard to see how creative work that brings into existence new ideas by stretching and recombining old ones in unexpected ways can avoid sometimes producing inconsistencies with components of the old view that remain useful, at least for computational purposes. A point frequently made by writers such as Feyerabend and Kuhn in the 1960s was that it is impossible to break out of the old conceptual framework if one is logically and semantically confined within it. But the really arresting observation is that, when examined closely, several of the most fruitful historical scientific developments have been found to contain inconsistencies or incoherencies of various kinds (Meheus 1999a, 2002). The logical empiricist position has been empirically refuted.

There is an additional point of importance. Fairly recently we philosophers have come to recognize that even "normal" research typically involves the use of models (in the informal, scientific sense), and models are, almost by definition, defective in some way. This means that a given model is inconsistent with some other things the scientists think they know. And where multiple models are involved, they need not be fully consistent with one another (Shapere 1984, Chap. 17; Nickles 1980a, 2002b).

It is clear, then, that inconsistency does not preclude fertility. On the contrary, in scientific practice (and other practices at the frontiers of creativity) perceived fertility trumps inconsistency. An often-quoted remark of Einstein is appropriate here: "The scientist . . . must appear to the systematic epistemologist as an unscrupulous opportunist."[11] Pragmatists are opportunists in this sense. They are "unprincipled" in the sense of never letting an ideological principle get in the way of a promising advance. Or rather, "Do not block the way of inquiry" (Peirce 1899/1932) is the only principle to which they are sworn.[12]

[11] In Einstein's reply to Lenzen and Northrop, in Schilpp (1949, 684). Of course, the logical empiricists had their own opportunistic interests, as I tried to bring out in Nickles (2002a).

[12] The complete quote from Peirce's 1899 manuscript reads:

> Upon this first, and in one sense this sole, rule of reason, that in order to learn you must desire to learn, and in so desiring not be satisfied with what you already incline to think, there follows one corollary which itself deserves to be inscribed upon every wall of the city of philosophy: Do not block the way of inquiry.

In response, Batens and Meheus have developed dynamic, inconsistency-tolerant, adaptive logics. One application of the relativity-to-context of their logics is the communicative context. Some adaptive logics are sensitive to the cognitive and communicative needs of real people in conversation, thereby reflecting the rhetorical and communicative sensitivity long associated with Ghent and Brussels.

We see, then, that Batens' research program is pragmatic in several basic respects. I end this section by more fully describing six of them. First, Batens is a pluralist, having developed entire series of logics for various purposes. Although **CL** is sometimes used as a kind of base-line logic, he and others also employ paraconsistent logics as the lower limit logics. For Batens and his group (by contrast with Graham Priest: see below) there is no single logic that constitutes the normative essence of good reasoning-in-general. Rather, like all tools, a given logic will be better suited for some tasks than for others, where the task-targets themselves will be a function of human interests. Thus logics are task-relative. In unfamiliar territory, the choice of one logic over another is a fallible choice. I am not sure how far the Ghent group wishes to push this idea, and in Sect. 7.5 I shall leave as an open question how far it can coherently be pushed.

Second, Batens' approach to logic appears to be naturalistic, at least in general tendency (again, see my queries in Sect. 7.5). His logics are not a priori gifts of the gods. Rather, they are products of a sort of evolutionary selection process, another point to which I shall return. For now, suffice it to say that Batens and Meheus consider logics to be humanly constructed tools that emerge historically in much the way that other tools and norms of excellence emerge from craft traditions.[13] Thus there is a strong consequentialist component to the justification of logics: "Ye shall know them by their fruits," not by a priori intuitions.

Accordingly, I believe that Batens and Meheus would largely agree with the following statement from Dewey's *Logic: The Theory of Inquiry*, an otherwise rather quaint logical text from 1938 that regards logic and methods as evolved products of inquiry, subject to ongoing change:

> [A]ll logical forms . . . arise within the operation of inquiry and are concerned with control of inquiry so that it may yield warranted assertions. This conception implies much more than that logical forms are disclosed or come to light when we reflect upon processes of inquiry that are in use. Of course it means that; but it also means that the forms *originate* in operations of inquiry. To employ a convenient expression, it means that while inquiry into inquiry is the *causa cognoscendi* of logical forms, primary inquiry is itself *causa essendi* of the forms which inquiry into inquiry discloses. . . . (Dewey 1938, p. 3f)
>
> . . .[I]t is cause for surprise that writers who energetically reject the intervention of the supernatural or the non-natural in every other scientific field feel no hesitancy in invoking Reason and a priori Intuition in the domain of logical theory. It would seem to be more incumbent upon the logicians than upon others to make their position in logic coherent with their beliefs about other matters. (Dewey 1938, p. 25f)

Dewey sees norms and standards governing expertise of all kinds as arising out of craft traditions (building boats; constructing houses, developing cuisines and fine

[13] John Dewey developed such a viewpoint, e.g., in "The Construction of Good," (Dewey 1929, Chap. 10).

arts and viable forms of government, etc.) rather than dropping out of the sky. Reasoning expertise and logical norms are to be included among these crafts.

Now compare to the Dewey quotation this response from Batens, when Graham Priest asked him whether he was an instrumentalist, given his pluralism:

> A deductive logic fixes the meaning of a fragment of a language. Languages are not God-given but are complex social constructions. We (try to) modify them in view of what we (think to) learn about the world. Such conceptual changes occur frequently in the languages of the sciences and, with some delay, in natural languages as well. Which languages are most adequate to handle certain aspects of the world cannot be settled a priori. Few will balk at this for 'referring terms' such as "phlogiston" or "mass." I claim it also holds for logical terms . . .. So my view is this: logicians develop logics just like one invents instruments, but nature (as knowable by us) determines which are the good instruments. (Batens and Priest 2008)

Third, although Batens' position is more realist than Dewey's instrumentalism, Batens, too, adopts a problem-solving versus directly truth-seeking approach to inquiry (Batens and Meheus 1996, 2001). Since we cannot test the truth of our claims against reality by direct inspection, there is no test-for-truth tool in our methodological armory. As Peirce clearly notes, we can't do more than employ the usual devices of scientific problem solving (Peirce 1877/1935). Appeal to the truth gives us no additional purchase.

Fourth, in that same article, "The Fixation of Belief," Peirce also noted that "each chief step in science has been a lesson in logic." Batens applies the lesson quite locally, as is apparent in Batens (2007). The logic is not held fixed relative to content but, instead, adapts itself to it, with the result that both content and logic help to guide the reasoning.

Fifth, worth separate emphasis is that Batens wants logic to have a prospective role in inquiry at all levels, given that we live in an uncertain world and must always have an eye to the future. Flexible logics are survival tools, whether in life or in one's problem-solving profession. Scientific inquiry at the frontiers of a research institutionalizes the human predicament in a self-conscious manner and thus provides a particularly helpful locus for study. Dewey liked Kierkegaard's wise remark that "Life can only be understood backwards, but it must be lived forwards." "We live forward" could have been Dewey's motto.[14] Batens and his logic group can be viewed as carrying on a broadly Deweyan project of recovery in philosophy, ironically, a recovery in part from that great pragmatist, W. V. Quine himself!

Sixth, although Batens' views are pretty radical, he and his group often employ **CL** as the lower limit logic, not because they believe it is somehow basically correct, but because that enables them to relate their work to the standard logical tradition

[14] For example, Dewey writes:

> Since we live forward; since we live in a world where changes are going on whose issue means our weal or woe; since every act of ours modifies these changes and hence is fraught with promise, or charged with hostile energies—what should experience be but a future implicated in a present! (Dewey 1917, p. 9)

and, thereby, to have a common reference point both as an aid to their own thinking and as a help in communicating their innovative results to a wider audience.

In attacking what he calls "monologism," Batens extends his pragmatic, humanistic pluralism even to logic. When Quine was in his prime, Batens and a few others such as Graham Priest boldly challenged logical orthodoxy with their "deviant" logics, as Susan Haack dubbed them (Haack 1996). Batens and his colleagues responded that, although their logics are *deviant* relative to intellectual orthodoxy, they are still *decent*. As for Quine, Haack, herself no great fan of deviant logics, nicely brings out the central tension here in Quine's alleged pragmatism (Quine 1951). For how can a strong fallibilist maintain that nothing, not even logic, is "immune from revision, come what may" while holding fast to **CL** as canonical? Ironically, in this particular respect, Carnap, with his internal-external distinction, which allowed for a pragmatic choice among alternative logical languages (Carnap 1950), was more pragmatic than Quine.[15] I should add that Batens is more of a pragmatist in deviating from classical logic as "the one true logic" than is Priest and (apparently) even Quine in that Batens is a thoroughgoing pluralist. Priest and Quine still believe that there is one correct logic. Quine held that the correct logic may differ from classical logic, i.e., that classical logic is revisable in principle, while Priest is already sure that classical logic is wrong and needs replacing. By contrast, Batens denies that there exists one true logic, whether classical or not. Any particular logic, whether classical or not, may work very well in some contexts but not in others.

If this interpretation of Batens' program is correct, we can take another step or two, as follows. Quine held that it would take a major scientific revolution to dislodge **CL**, whereas Batens, Meheus, and other practitioners of deviant logics assert that it is necessary already to handle much ordinary reasoning and normal scientific work as well as radical thinking at extreme frontiers. From this point a corollary follows. It is commonly said that logical truths (and the logics that generate them) are "true in all possible worlds." But if Batens is correct, then **CL** is not even "true" in our world in the sense that it is not the most fruitful logic to apply in all reasoning contexts. But, again, rather than speak of truth in this stretched sense, I believe we should speak of fertility, including estimates of future fertility in scientific decision-making (and similar choices in other creative enterprises), or what I have termed "heuristic appraisal."[16] Putting the point of Batens' work in this way highlights both the magnitude of his accomplishment and the reasons why critics might resist accepting it as an accomplishment.

[15] Here I am indebted to Alan Richardson.

[16] I believe that William James got himself into trouble with his so-called pragmatic theory of truth precisely because he often meant something like "positive heuristic appraisal" rather than "truth" (Nickles 2006; Nickles and McCollum-Nickles 2002).

7.3 Campbell vs. Simon? BVSR vs. Logic?

Diderik Batens and I are in full agreement that we need a lot more attention to frontier theory of inquiry. However, he and I seem to take quite different approaches to the subject. As an admirer of Batens' work, I have always wanted to see if we could achieve some integration of our two approaches. My contribution to this volume takes some steps in that direction.

The primary obstacle is that I embrace Donald Campbell's broadly Darwinian position that selectionist processes underlie all forms of design innovation (creativity that survives and replicates), not only in the biological world but also in the world of human invention, that is, in social evolution (Campbell 1960, 1974; see also Dennett 1995). There are, of course, important differences between human creativity, and the non-human biological forms, but, at bottom, all such processes are minimally selectionist. Or so I maintain.[17] And, superficially at least, Campbell's "blind variation plus selective retention" (BVSR) mechanisms seem strikingly different from a rational, logical procedure.[18]

Two stories will help to bring out the apparent opposition here and hence to present the problem of integration. The first story is personal. I was once an opponent of Campbell. I agreed that there was something to what he said, but I thought he went too far in seeming to deny that there is any method to inquiry, or any rationality left at all in saying that it finally comes down to blind variation. I believed that rational constraints, including general heuristics of the sort that Allen Newell and Herbert Simon advocated (Newell and Simon 1972), would do the trick when supplemented by domain-specific content (as in knowledge-based computation). So around the time, in the late 1970s, when Batens was discovering adaptive logics, I, too, had concluded that we must expand the idea of rationality beyond classical logic and traditional conceptions of method to include frontier inquiry—context of discovery. It seemed to me then that Campbell's model was insufficiently rational even in this broadened sense.

I was not alone. The purpose of Newell and Simon's problem-solving strategies was to employ logic and heuristics to cut the search space down to manageable size. They seemingly agreed that most, if not all, research is problem solving and that problem solving requires search; but at that time, as implemented in their

[17] Mine is a minimal selectionism. It is not in itself a method or even a general strategy for creativity. To methodize it in particular cases requires the input of domain-specific targets and constraints. As Simon, Newell, and Shaw note, sometimes the difficulty lies more on the side of "solution generating processes," sometimes on the "verifying" side, e.g., in choosing a move in chess, where a legal move is easy to generate (except, of course, in a mate situation), while verifying that a move is a good one can be extremely difficult (Simon et al. 1962).

[18] The label "selectionist" (as opposed to providential and instructionist) should not lead us to think that the alternatives, the variants, are easy to come by, and that the problem is simply to select the best (or a satisfactory) one.

General Problem Solver, they regarded the laws of logic plus the general, content-neutral heuristics as sufficient, in fact as psychological analogues to Newton's laws. Classical logic did not take us very far into context of innovation, but adding heuristic rules sounded promising, and, indeed, it was an important step in the right direction. Adding content-laden constraints was a further step in the right direction, in my opinion. By contrast, Campbell's approach seemed wildly inefficient as a way of solving problems.

In the early 1990s Campbell scolded me gently for misrepresenting his position, which led me to reconsider it. Then, around 1997, I read Dan Dennett's *Darwin's Dangerous Idea* (Dennett 1995) and Gary Cziko's *Without Miracles* (Cziko 1995), and I finally saw the light—unfortunately after Campbell's death. I had picked up the wrong end of the stick.

Unfortunately, the term 'blind variation' has led to much misunderstanding as well as criticism, so much so that, late in life, Campbell preferred other expressions such as 'undirected variation' or 'unjustified variation'. 'Blind' implies neither indeterministic nor unconstrained, let alone unmotivated by personal "reasons." Typically in research and other creative endeavors, one has reached a point of relatively firm understanding or ability and wishes to push beyond that point by exploring the space of possibilities around it without knowing precisely what is there. Here the exploration or search is constrained by the reference point. Completely "off the wall" suggestions are not what is wanted. As for motivation, when trying to solve problems, people have all kinds of "reasons" for trying this or that, without knowing in advance which trial, if any, will be successful.

Two points to remember here are that at frontiers of creativity no one knows much of the domain structure beyond the frontier, so we have no alternative but trial and error. We want the variation process to be as constrained as possible by current knowledge[19]; but at the frontier, by definition, it will be even less constrained than for normal problem solving. 'Trial and error' does not, of course, mean that reasoning is absent. One may reason from intelligent guesses based on previous experience, for example. Abductive reasoning (to which the Ghent group has given much attention) involves just this sort of hypothesis formation. Notice, however, that even when the answer to the question or solution to the problem is *totally* constrained by current knowledge (not supposition), it still may take the greatest effort—extensive *search*—to find a path to the solution, since neither we nor our research technologies are omniscient. This was Kuhn's point about puzzle solving

[19] This is true of artificial intelligence implementations as well. E.g., in John Koza's version of genetic algorithms (Koza 1992), the initial population of variants may be computer programs for addition, multiplication, and other basic mathematical operations. The next generation is populated by variants probabilistically selected (and weighted according to their evaluation as problem solutions) from the preceding generation, with new variants produced by a probabilistic exchange ("crossing over") of subbranches of some of the program trees that remain in the competition. An anonymous referee reminds us of simulated annealing as another example of constrained variation.

in normal science (where the paradigm supposedly guarantees the solution in terms of application of available exemplars—see Kuhn 1962, Chap. III; Nickles 2012). Note that search is required to solve challenging problems even when we know the answer and also know that we possess the knowledge sufficient to get the answer but not "how to get there from here," as in the logic and mathematics problems that students are regularly asked to solve, where the answers are already given in the back of the textbook.

A second way of putting basically the same point is that all novel problem solving, whether in normal or revolutionary science or musical composition, involves search; and all genuine search is blind search to a greater or lesser degree, greater insofar as the constraints themselves are soft.[20] Even systematic, brute force search is blind in the sense that we don't know in advance where the answer lies and must search for it. The point is really a simple one but often misunderstood, even by the so-called "friends of discovery."

As I say, I was far from being Campbell's only opponent, and here is my second story. In 1959 Campbell presented his BVSR views at one of the Macy conferences, this one involving Simon, Newell, Clifford Shaw, Warren McCulloch, Frank Rosenblatt, W. K. Estes, Heinz von Foerster, Arthur Burks, a young Marvin Minsky, and others. Minsky led off the discussion of Campbell's paper with the following remark:

> I can't accept this feedback as being positive .... I would really like to know if you believe you are saying something constructive rather than anarchistic? Namely, it seems to me what you have said particularly in reference to Newell *et al.*, that you don't believe things are as bad as they make out in the British Museum algorithm. That is, the space isn't really so large, there aren't so many bad trials and generally speaking things are pretty good .... What is the constructive purpose in emphasizing the role of trial-and-error which is what we are trying to get rid of? (Yovits and Cameron 1959, p. 228)

That statement, together with the early work of Newell, Simon, Shaw, and others, is a good indication of the state of play at that time. To be sure, within a few years John Holland at Michigan would introduce an early form of evolutionary computation by means of genetic algorithms in which populations of digital strings representing candidate problem solutions were allowed to evolve (Holland 1975). But that effort, too, fell on deaf ears in the computer science and artificial

[20] The degree of blindness in search will depend on how constrained the search is, including the constraints that enable you to identify the goal of your search. Knowing that you had your keys when you entered the house, you are confident that they are somewhere in the house. At the other extreme, you are not sure whether your problem is solvable at all, whether the target of your search even exits, whether you are even asking the right question. This is the human situation. Another paper would be necessary to spell out the relevant similarities between and differences from biological evolution, in which there is no explicit formulation of problems or questions to be answered and no "targets" in the human sense. But even here variation is constrained. Two rabbits who mate produce other rabbits that closely resemble them in relevant respects, both genetically and phenotypically.

intelligence communities for many years thereafter. Nor did Campbell himself pay much attention to it, as far as I know. Today, of course, evolutionary computation and genetic algorithms are big things, indeed, entire fields that have produced families of BVSR mechanisms that we can rightly call problem-solving *methods*. As Campbell (1974) noted, hypothesis and test (including Popper's "conjectures and refutations") is simply a slow version of BVSR. Within the later framework of evolutionary computation, we can now regard evolutionary computation as a vastly scaled-up version of generate and test.

Now why do I think that Campbell's BVSR approach and the Ghent logical program are not incompatible? The answer is, roughly, for the same reason that I think Campbell's program and Simon's early AI work were not, after all, totally at odds. Simon and Newell certainly did not claim that they possessed an algorithm that dispensed with the need for trial and error. For example, at the core of their General Problem Solver was a generate-and-test heuristic that is indispensable in proposing new solution steps to evaluate. Although Newell, Simon, and their associates naturally wanted to employ every available means to cut down the search space as much as possible, as long as the problem remains unsolved some search space remains, no matter how large it was to start with. And search is search! Insofar as the constraints have exhausted their directive force, the search is blind and one can proceed only by trial and error, hoping to turn those results into further clues or constraints, whether hard or soft. Such search is not totally blind and "random," of course. It is directed toward the remaining search space. But within that space it is blind, undirected by knowledge already in our possession. We may have personal reasons for guessing one way or another, but this sort of direction is not epistemically justified. In this sense, the phrase 'blind search' is redundant. All search is 'blind' in this carefully restricted sense.

A later article by Simon and Glenn Lea makes the point quite explicitly.

> [F]rom a logical standpoint the processes involved in problem solving are inductive, not deductive.... To be sure, the proof of a theorem in a formal mathematical or logical system [such as Logic Theorist] is a deductive object; that is to say, the theorem stands in a deductive relation to its premises. But the problem-solving task is to *discover* this deduction, this proof; and the discovery process, which is the problem-solving process, is wholly inductive in nature. It is a search through a large space of logic expressions for the goal expression—the theorem. Hence, both a theory of problem solving and a theory of rule induction must explain inductive processes—a further reason for believing that these theories should have something in common. (Simon and Lea 1974, p. 330, their emphasis)

Notice that they are here talking about the need for search even in contexts in which the solution is totally constrained in the logical or mathematical sense but where no constructive proof or algorithm is available. There exists a major gap between the existence of constraints and our psychological ability to use them to construct a path to the solution (cf. Meheus 2004). Otherwise, we would possess logical omniscience and much of the best logic and mathematical work would be reduced to triviality. If search is needed to solve even routine problems, a fortiori it is needed all the more at frontiers. And as Simon et al. (1962) already pointed out, it will be less constrained there and hence often more creative. They go so far as to suggest that proximity to the limit of completely blind trial and error might provide a measure of creativity.[21]

7.4 Comparison with the Batens Program

There are parallels in Batens' program. He and Meheus are not providing algorithmic logics of discovery, and god knows that they deny omniscience! They have a lot to say about how their inconsistency-tolerant, dynamic logics can help narrow the suspicion concerning problematic premises as well as inconsistencies that arise in the course of reasoning. Yet this process requires a good deal of trial and error. To be sure, much of their discussion is couched in terms of abstract, logical examples, but Meheus also provides excellent historical case studies of Lavoisier's work on oxygen and of Clausius' reasoning to the laws of thermodynamics from the mutually incompatible work of Sadi Carnot and James Joule (Meheus 1993, 1999a, 2007). We need more case studies of this sort. (An interesting autobiographical and reflexive case study would be for Batens or Meheus to attempt to reconstruct the logics they employ in one of their own logical discoveries.) The volume *Inconsistency in Science* (Meheus 2002) is a good start. As noted above, adaptive logics may be especially valuable in understanding model-based reasoning, given that models are usually known to be defective in one or more ways from the beginning: they involve idealization, abstraction, simplification, and/or approximation (Shapere 1984; Nersessian 2002). In such cases there is no question of testing the model to determine whether or not it is true.

What follows are seven overlapping items of positive evidence that the Ghent approach is not really antagonistic to the selectionist model but, on the contrary, requires selectionist steps at several points.

1. Inconsistency adaptive logics are externally dynamic with respect to new information imported into a reasoning process from outside and internally dynamic in adjusting to inconsistencies that arise from the reasoning itself as new inferences are made. Response to both processes can involve a certain amount of trial and error in the sense that the reasoner, in the process of reasoning, as in ordinary logic, often has a choice of what step to try next, to see if it helps to lead to a desired conclusion.[22] In the case of inconsistencies arising as part

[21] I personally believe that such a criterion is problematic. Sometimes a problem is so over-constrained, where the constraints apparently conflict, that a solution appears to be impossible. Einstein's solution to the special relativity problem was close to that when he was able to show that the relativity principle and the constancy of the velocity of light in vacuo are not mutually incompatible.

[22] Again, goal-directed reasoning from premises and abnormalities to consequences that follow logically from them (according to the logic being used) is a quite different process than biological

of the internal dynamic, the logic can help locate where the trouble arises but cannot tell you which changes to make (as to which new information can be admitted as premises) or exactly how to remedy an internal contradiction (e.g., Batens 2006).

2. As a subcase of the above (in the sense of employing an adaptive logic for analogy as the mechanism to input new information), Meheus (2000) shows, by combining an adaptive logic for analogy with one for inconsistency, how adaptive logics can allow input from analogies, by treating analogical inputs to the reasoning process as premises that are abnormal and hence not to be trusted initially. I would say that using analogies in this way is just another way of producing variants. The logic does not, of course, generate the analogies to try as premises. And there remains a choice or selection on the part of the user, since the logic does not provide an automatic solution to the problem. Meanwhile, the inconsistency-adaptive logic acts as a selection mechanism.[23]
3. The overall strategy is consequentialist, whether or not the content-laden, premises-like items are literal imports, adapted rhetorically from another domain, or simply conjectured on the spot. You (or the logic) try a "premise" to see where it leads, often rejecting it or modifying it as the reasoning proceeds.
4. Batens (2008) notes that skilled problem solvers often do not start from scratch but instead search for solved problems that are similar Such case-based and model-based reasoning procedures employ analogical or metaphorical reasoning also, now at the level of entire problem complexes and solutions (or partial ones). This process involves BVSR twice over, first in the search for related problems and then in the further tinkering that is usually necessary in order to construct a sufficiently close match. One of Kuhn's most important insights in *Structure* was that most scientific problem solving is based on direct modeling rather than on rules. His most impressive example (in the "Postscript—1969") is the series of models employed in the pendulum-efflux case, by means of which Daniel Bernoulli was finally able to solve the problem of the flow of fluid from an orifice.

When possible, case-based and model-based reasoning are far more efficient than "wilder" or "blinder" forms of variation and selection, since one restricts the variation to the region around a solution already in existence and known to work, at least by analogy. A good deal of material context comes with the exemplar. The model provides an intuitive cluster of constraints on the variations that prevents one from having to start from scratch. In automatically connecting new work with previous work, it orients both the innovators and their audience. In effect, such

evolution. My point here is not to equate them but only to indicate that, unlike following the steps of a constructive proof, the reasoning agent must search through a constrained space of possibilities in order to figure out how to reach a particular conclusion. With Campbell, I hold that human innovative design also involves a BVSR process, but, again, I cannot here engage the similarities and dissimilarities with biological evolution.

[23] Thanks to the referees for clarification on these points.

work explores the region around previous results and thus remains grounded in research reality and therefore intelligible. This is very much in line with what I call Dewey's "historical nominalism": that we avoid the dangers of unnecessary blindness and/or of trying to be so revolutionary that we wander off too far into possibility space to have any chance of success.

5. Batens and his research group are especially interested in problem posing and problem solving at frontiers where (they rightly say) all problems are ill structured. As noted already by Simon (1973), if even well-structured problems require BVSR, ill-structured ones will require all the more. Here direct modeling is sometimes still possible, but the lack of domain knowledge means that the problem solving is no longer routine. In fact, we typically get variation even at the level of problem formulation.
6. The selectionist paradigm gives us a non-traditional conception of the economy of research in which apparently wasted effort becomes essential to innovation. That is surely one reason why selectionists still so often hear the "monkeys at typewriters" objection, with selectionism sometimes dubbed "the British museum algorithm." From the selectionist point of view, at least some of what was formerly considered wasteful and irritating "noise" in the system (accidental departures from the typological ideal) can now be seen as variation that is essential to evolutionary development. In biology it was breeders and then Darwin as theorist who first recognized the creative potential of variation, as essential to the evolutionary process. This is not, of course, to say that all variation processes are equally efficient. In any given context, many are clearly useless. Variation is necessary to innovation, although far from sufficient.

   One source of variation in the sciences and the arts is the "noise" of meaning variance within and across speakers and contexts, misinterpretation, even misprints. Batens, early and late, stresses the fluidity of language, e.g., in critiquing incommensurability claims and claims for rigid hierarchy (Batens 1983, 1992, 2001). Even in ordinary conversation and our class lectures there are variations in the ways that we and our students say and understand even familiar things, variations that sometimes provide creative sparks and the occasional epiphany. Purists who insist that all concepts be rigorously defined operationally before the research is fully underway fail to articulate correctly the nature of research, as even the logical empiricists eventually recognized (Hempel 1950). Operationists such as P. W. Bridgman, who wanted to permanently fix all relevant concepts even before theorizing or model-building began (Bridgman 1927), remind us of biologists before Darwin, for whom variations in the representative organisms of fixed species were just "noise" in the system and hence a nuisance rather than the raw material for innovation.
7. The entire Batens research program of constructing and testing variant logics—for fertility in various domains as well as for completeness and consistency—is naturalistic-evolutionary and hence consequentialist—a BVSR process. According to this program we develop new tools by selectionist processes. They come to us neither providentially (as a gift of the gods) nor by direct instruction from

> nature (Cziko 1995; Nickles 2003a). Thus the Ghent program can be regarded as a facet of evolutionary epistemology and evolutionary computation, where the populations in question are populations of logics.

The general point is that BVSR is everywhere. Douglas Lenat expressed the same idea in saying "Discovery is ubiquitous." (Lenat 1978). The selectionist's slogan might be: "You show me research, and I'll show you search—and hence BVSR." Search is necessary at all stages: choosing a problem, finding candidate solutions, often modifying the problem in the process, then searching for suitable ways to test the result in theory, suitable ways to realize the test in practice, to analyze the data that results, and so on. That variation and selection are necessary at all creative stages of research is evident already by thinking about the method of hypothesis.

Let me conclude this section by returning to the idea of a logic or method of discovery. While no one (in their right mind) claims that there exists an algorithmic logic of discovery, I see no reason why the Ghent work could not be incorporated as tools in computational modes of problem solving. In fact, it already has been.[24] While artificial intelligence has not made the rapid strides originally predicted for it, there has nevertheless been important progress in machine learning, causal Bayes network theory, and evolutionary computation.

Computer scientists are rapidly developing increasingly powerful problem-solving programs that solve difficult, computable problems that require increasingly less instruction of how to do it from the programmer. Thousands of technical papers published each year now make use of results obtained by one kind or another of evolutionary computation. The general approaches all involve BVSR in some form but, qua general approaches, they are not "good old" logics of discovery. Rather, they are problem-solving shells. In application to specific problems, the details remain to be specified. In fact, we are increasingly confident that there is no specific method, evolutionary or not, that is both powerful and universal or domain-neutral in application. As Simon and Newell eventually found out, programs with any power must incorporate domain knowledge, leading Edward Feigenbaum to note, already in 1968, "There is a kind of 'law of nature' operating that relates problem-solving generality (breadth of applicability) inversely to power (solution successes, efficiency, etc.) and power directly to specificity (task-specific information)."[25] And the "No Free Lunch" theorems of Wolpert and Macready, in denying that there is one best method for application to all possible universes, also cast doubt on the existence of a method that will be successful across all scientifically interesting domains of our universe (Wolpert 1997). There are many different specific ways

[24] An anonymous referee directed me to a page of the Ghent "logica" website that I had not previously known, where one can find several programs written by Batens, Giete Callaert, Alex Klijn, and Albrecht Heeffer. Go to http://logica/ugent.be/centrum/writings/programs.php.

[25] See Feigenbaum et al. (1971, p. 167) for quotation of this previously published remark of Feigenbaum.

to write and implement genetic algorithms. The good ones must be tuned to their domain of application.[26] So we have considerable pluralism within the selectionist camp as well.

Nonetheless, genetic algorithms and other forms of evolutionary computation do constitute methodological strategies of a fairly general sort that are proving their mettle. Consider the tradition from John Holland to John Koza. Koza wrote a series of four increasingly large tomes on the subject, each of which attempts to give the programs more autonomy in the sense of telling them what sort of problem to solve but not how to solve it. Clearly, the problem solutions produced are typically more powerful than any of the information that is input. In Koza (1992), the first volume, for example (which admittedly retains rather mechanistic modes of variant generation), the initial population may consist of computer programs for basic mathematical operations. After many iterations (including probabilistic selection of the more successful candidates at each stage and probabilistic mating by "crossing over" sub-branches of the programs) the solution emerges by a kind of evolutionary process. Just as in Darwinian evolution, we get more design out than we put in.

To the degree that it makes sense to speak of methods of discovery of this sort, it is ironic that Donald Campbell's best-known paper, "Evolutionary Epistemology" (Campbell 1974), appeared in the Popper volume in the Library of Living Philosophers, and was couched, in part, as an argument against discovery methods. (Popper's comment on the article was full of praise.) If I read him correctly, Campbell himself did not realize that scaling up BVSR in appropriate ways could turn it into a humanly usable problem-solving technology. So, after all, his case for BVSR can be converted into an argument for, rather than against, methodologies of discovery of a significant sort, although not ones that are pathbreaking, so far. To this limited extent, my early worries about Campbell's approach may be vindicated.

7.5 Some Concluding Questions for the Ghent Program

I have labeled Diderik Batens and his workgroup pragmatists, but what sort of pragmatists are they, hard pragmatists like Peirce, soft pragmatists of the Richard Rorty sort, or something in between? The answer is surely not Rorty (although his attack on analytic epistemology does score points), but something closer to Peirce (given his natural science orientation and his exploration of logics, including

[26] Dennett writes:

> any functioning structure carries implicit information about the environment in which its functioning "works." The wings of a seagull magnificently embody principles of aerodynamic design, and thereby also imply that the creature whose wings these are is excellently adapted for flight in a medium having the specific density and viscosity of the atmosphere within a thousand meters or so of the surface of the earth (Dennett 1995, p. 197).

abductive logics) and also to Dewey (in social constructivist respects), as already explained. In his day Peirce helped to reopen logical investigation as a genuine frontier for exploration, one that Quine in a sense attempted to shut down 50 years later by regarding **CL** as canonical. In his early jewel of an essay, "How to Make Our Ideas Clear," Peirce berated logicians for "slumbering through ages of intellectual activity, listlessly disregarding the enginery of modern thought [especially modern science], and never dreaming of applying its lessons to the improvement of logic" (Peirce1878/1935, Part 1). Such a remark would not be at all fair of twentieth-century logicians, and yet we can discern a surprising conservatism in some quarters!

First question. One question I have here is whether Batens and others in his research group adhere to the analytic-synthetic distinction, which, of course, Quine did reject on pragmatic grounds. I doubt whether they do, given their naturalistic, fallibilist, constructivist orientation, their explicit rejection of purely a priori principles, and especially their logical pluralism. Yet the Ghent researchers do sometimes speak of "strictly formal analysis," so I seek clarification on this point. Of course, "strictly formal analysis" remains possible within a well-defined, abstract system without commitment to analytic or synthetic a priori truths. In a way even Peirce anticipated the breakdown of this distinction in his explanation of apparently self-evident claims as having a cultural rather than an epistemological origin.

It would seem that reducing logics to investigative tools alongside other tools (tools that are themselves being tested indirectly for their fertility and subject to replacement by more effective tools as they are developed) would require a thoroughgoing abandonment of the analytic-synthetic distinction, at least beyond surface-level definitions. Once we go beyond making a sharp distinction between truth-bearers and non-truthbearers to speak pragmatically of whether something "works" well enough for the purpose at hand, once we make even logics context-dependent, then there no longer seems much point of saying that some statements are privileged (or not) because they are completely devoid of "empirical content." The point can be regarded as an extension to logic of the "No Free Lunch" theorems of Wolpert and Macready, themselves an attempt to formalize the insights of Hume and Wittgenstein.[27] The basic idea is that, coming fresh to a new domain (or a new "world"), we cannot know a priori, whether any method at all will help us determine the structure of that domain or which one will be more helpful than others. Wolpert and Macready argue that no method has an advantage when averaged over all possible domains. Like Batens, I am sympathetic with the idea that, on the conception of logics as domain-sensitive tools, there exists no general, context-free logic anymore than there is a general, context-free method of discovery in the old Baconian-Cartesian sense. But whether or not it is coherent to extend the "No Free Lunch" idea to logics surely remains an open question.

Second question. Are Batens and Meheus committed to the existence of logics of discovery after all? Now in their many papers on abductive logic, analogy, and such,

[27] See Wolpert (1996, 1997). I discuss the theorems in Nickles (2003a).

they shy away from the idea. But if context of discovery can be as rational as you please and if that rationality implies the existence of an underlying formal logic of some kind, then . . .?

The answer to this question is surely negative. The Ghent group do claim that inconsistency adaptive logics and analogy tolerant and abductive logics have roles to play in context of innovation, that at least some major innovations can be reconstructed as cogently reasoned within a decent formal system, but that such logics fall far short of providing deductive or inductive determination, let alone a constructive proof of the desired result. As Batens and Meheus explain in their papers, when an inconsistency arises within an adaptive logic, for example, the logic adapts by departing from **CL** in order to prevent disastrous inferences, but the logic itself does not tell the user which precise steps to retract. That's where BVSR comes in. Such a logic is not a strict logic of discovery.

Third question. Nonetheless, I wonder whether there is a worrisome regress lurking within the pluralistic pragmatism. (This is an old problem for all of us, one that affects Quine's pragmatism and especially Carnap's, indeed, one that affects everyone critical of the idea that we humans possess a faculty of universal reason.) As I understand it, the Ghent logic program presupposes that underlying all *rational* inference are one or more logics that can be formalized, logics that ultimately define or delimit what counts as rational behavior in their corresponding contexts of application. But then what rational basis is there for deciding which kind of context we are now in, and which logic to apply there? (This is the problem of the Big Switch, but in an especially virulent form, since it concerns the rational choice of logics themselves.) Insofar as *those* decisions are rational, they would seem to presuppose that there is a more general underlying formal logic, and so on down until we reach a bedrock logic that defines universal reason, so to speak. According to the Ghent program, can we get by without a logical counterpart to universal reason? If so, how exactly? Of course, the answer to the first of these questions could be the Deweyan one: We can, because we do! Their program is then an attempt to answer the second question.

Fourth question. My final two questions concern the Ghent group's claim that all creativity involves logical reasoning, or at least a reasoning process that can be modeled by logic. The Romantics reacted to the Enlightenment by making a strong distinction between imagination and reason. We need not accept the specifically Romantic conceptions of imagination, and we do not want to return to overly romantic conceptions of scientific discovery; but isn't there a grain of truth in the appeal to the creative power of imagination? Here is a passage from Rorty that expresses something of this view.

> What the Romantics expressed as the claim that imagination, rather than reason, is the central human faculty was the realization that a talent for speaking differently, rather than for arguing well, is the chief instrument of cultural change. (Rorty 1989, p. 7)
>
> The gradual trial-and-error creation of a new . . . vocabulary—the sort of vocabulary developed by people like Galileo, Hegel, or the later Yeats—is not a discovery about how old vocabularies fit together. That is why it cannot be reached by an inferential process—by starting with premises formulated in the old vocabularies. Such creations are not the result

> of successfully fitting together pieces of a puzzle. They are not discoveries of a reality behind the appearances, of an undistorted view of the whole picture with which to replace myopic views of its parts. The proper analogy is with the invention of new tools to take the place of old tools. To come up with such a vocabulary is more like discarding the lever and the chock because one has envisaged the pulley, or like discarding gesso and tempera because one has now figured out how to size canvas properly.
>
> This Wittgensteinian analogy between vocabularies and tools has one obvious drawback. The craftsman typically knows what job he needs to do before picking or inventing tools with which to do it. By contrast, someone like Galileo, Yeats, or Hegel (a "poet" in my wide sense of the term—the sense of "one who makes things new") is typically unable to make clear exactly what it is that he wants to do before developing the language in which he succeeds in doing it. His new vocabulary makes possible, for the first time, a formulation of its own purpose. It is a tool for doing something which could not have been envisaged prior [to that]. [ibid., pp. 12–13]

Here, of course, Rorty's use of 'vocabulary' is itself a metaphor for something more than words. Examples include even the development of new mathematical perspectives, as when Descartes realized that new mathematical modes of thinking were more powerful than the old logic, indeed, transformative (Heeffer 2008). Working toward a hoped-for conclusion already in hand (solving a well-defined problem) is normal science, whereas Rorty is interested in transformative change.

Talk of transformative change reminds me of the old pragmatic dictum that there are two ways to solve a problem: You can either get what you want (normal science) or you can want what you get (a radically prospective view that includes revolution), and that the justification of the latter, when successful, has the character of a bootstrap operation in being retrospectively self-supporting (Feyerabend 1989). While Rorty draws the distinction too sharply, in my opinion (again, read Meheus on Clausius, who, after all, was inventing modern thermodynamics in the process), I wonder how hard Batens and Meheus want to push the idea that innovation contexts are underlain by reasoning that can be reconstructed in formal logical terms. I am quite sure that they do not extend this claim to all such contexts. Most of them? Normal science versus revolutionary science? The thesis clearly fails for biological evolution itself, whereas I would claim that BVSR fits all cases, whether biological or intentional. (So there remains this much disconnect between my approach and Ghent's.) But let us here restrict the discussion to the creativity of humans and the rational beings that may exist on other planets.

I suspect that there is a broadly perceptual component to much creativity, including scientific problem solving, and that appeals to imagination attempt to capture this. Even Kuhn's account of normal science depends heavily on a kind of perceptual pattern matching. Experts who could not themselves produce the solution to a physics problem often quickly agree on whether a proposed solution is correct or not. It seems to be a matter of recognition, of paradigm-matching or exemplar-matching, over and above the specific tests that check the result. And, after all, recognizing a solution or an interesting novelty when one stumbles upon it is the key to solving the Meno problem of the very possibility of inquiry. Pattern matching also seems crucial to logical reasoning at one level. We are not born with an innate faculty of Reason, as the Ghent group well appreciates.[28] We typically reason well only in domains with which we are quite familiar. There is something "material," something more than formal, about much good reasoning, which is not to deny that formal similarities can be of great importance as well.

Fifth question. A related question concerns rhetorical tropes such as analogy, similarity, and metaphor. In his discussion of direct modeling of new puzzles on old exemplars by means of an "acquired similarity relation," Kuhn (1962) made rhetoric as important as logic in scientific cognition. Many philosophical writers since then have attached increasing importance to rhetoric, especially in contexts of conceptual growth. In the debate over whether or not analogies and other tropes employed by creative scientists are strictly syntactical matches of the equations, I am currently on the side of *not*! Expert knowledge of content and of the phenomena is necessary in many cases. For one thing, at the frontier the equations are not there in advance. One tries out an equation, one modeled on that applicable to an old problem, precisely because one thinks there is a kind of analogy between the phenomena themselves, or the underlying processes that produce them, a kind of heuristic "rhetorical realism," as we might term it.

Now some of this tropical thinking feels more sensory than symbolic. Here I side with people such as Mary Hesse, Douglas Hofstadter, Melanie Mitchell, and Lawrence Barsalou who stress content and the fluid visualization afforded by dynamic mental models. I do not believe that our highly developed visual system, for example, remains idle while we think.[29] When Meheus discusses the role of analogy in reasoning (Meheus 2000), she cites only Gentner (1989) and Holyoak and Thagard (1989), both of which use syntactical engines.[30] No one really knows the exact system of grammatical rules underlying any natural language, if, in fact, there is one; and, even if we did, in the flow of speaking, writing, even editing, we don't often appeal to rules. Rather, a fluent speaker knows that it looks right or sounds right, or not. Similarly, Kuhn says, for much expert problem solving.[31]

[28] Johnson-Laird bases the cognition of logical moves themselves on pattern matching (Johnson-Laird 1983).

[29] For current research support of this claim, an anonymous referee refers us to David Landy's list of publications at https//facultystaff.richmond.edu/~dlandy. Landy and colleagues argue that visual features are important to abstract processing, e.g., in solving algebraic equations. See also Changizi (2011).

[30] See Meheus (2000, p. 28). Compare Gentner et al. (1993), Bechtel (2008, p. 434), and Nersessian (2008, pp. 148–150). Someone may object that, on a computational model of the brain, the computations must ultimately be syntactical. My response is twofold. First, I'm here siding with those who reject a fully symbolic-computational model as opposed to important roles for association nets. But, second, I am talking about the level at which creative human beings think and act, not processes deep inside the brain. John Norton's defense of material induction, his claim that "There are no universal rules for induction," is relevant here (Norton 2003, 2010).

[31] I am something of a skeptic when it comes to claims such as Chomsky's that an elaborate, abstract system of rules underlies our linguistic practice, modulo the limitations of linguistic performance. It is notoriously difficult to see how such a holistic system could have evolved. The idea that everyone is born with an inbuilt language acquisition device that implicitly contains all possible

It remains an open question whether acquired intuitions of the kinds exhibited in linguistic behavior, in innovation contexts, and in expertise of any sort must be underlain by systems of rules.

In a way it does not matter to Meheus' purpose, in the analogy paper, where the analogical premise candidates come from. But I wonder whether she and Batens simply assume a version of the computational theory of mind, holding that all cognition is symbolic computation that can be captured by some logic or other. I suspect that other forms of mental representation are involved. In creative contexts, concepts tend to be especially fluid. Even within normal science we often get changes that are subtle yet large enough to introduce ambiguities into logical arguments. (Graduate students in history of science write dissertations claiming to locate previously unnoticed conceptual shifts.) What we might term "conceptual creep" thereby challenges the application of formal computational apparatus, or so it would seem.[32] The fluid view of concepts often goes hand-in-hand with the idea that concepts themselves are based on rhetorical relations (similarity relations) rather than sets of logically necessary and sufficient conditions. There is much more to be said along these lines. Psychological evidence is accumulating. Nancy Nersessian has pointed to much of it in her recent book (Nersessian 2008).

Acknowledgements Thanks to Joke Meheus and Erik Weber for their invitation to give a plenary lecture at the conference they organized to mark Diderik Batens' retirement. Thanks especially to Joke and to Werner Callebaut for comments on my contribution, and to exchanges with Bill Wimsatt in earlier times. Two anonymous referees were exceptionally helpful in correcting my misunderstandings of the Ghent program and in several other ways. Naturally, I am responsible for any deficiencies that remain.

References

Batens, D. (1983). Incommensurability is not a threat to the rationality of science or to the anti-dogmatic tradition. *Philosophica, 32*, 117–132.

Batens, D. (1992). Do we need a hierarchical model of science? In J. Earman (Ed.), *Inference, explanation, and other frustrations. Essays in the Philosophy of Science* (pp. 199–215). Berkeley: University of California Press.

human languages and only needs bits of empirical evidence to throw the right switches strikes me as almost as implausible as the idea of a scientific method that contains implicitly all humanly possible discoveries, to be dredged up, Meno fashion, by questioning it against smatterings of empirical claims. (Thanks to a remark by Gary Cziko for this insight.) By contrast, compare Donald Davidson's less systematic view, according to which a surprising amount of our linguistic give-and-take is made up on the fly in particular communicative contexts, and metaphors are not so different from literal discourse (Davidson 1978, 1986).

[32] An anonymous referee points out that this claim is challenged by recent work on mathematical discovery, e.g., Pease et al. (2010), who computationally implement Lakatos's ideas about mathematical reasoning (Lakatos 1976), including meaning shifts and heuristics.

Batens, D. (1996). Functioning and teachings of adaptive logics. In J. Van Benthem, F. H. Van Eemeren, R. Grootendorst, & F. Veltman (Eds.), *Logic and argumentation* (pp. 241–254). Amsterdam: North-Holland.

Batens, D. (1999). Contextual problem solving and adaptive logics in creative processes. *Philosophica, 64,* 7–31. Appeared 2001

Batens, D. (2001). Aspects of the dynamics of discussions and logics handling them. Ghent University Archives: http://archive.ugent.be/person/801000271859.

Batens, D. (2004). The need for adaptive logics in epistemology. In *Logic, epistemology and the unity of science* (pp. 459–485). Dordrecht: Kluwer Academic.

Batens, D. (2006). Narrowing down suspicion in inconsistent premise sets. In J. Malinowski & A. Pietruszczak (Eds.), *Essays in logic and ontology* (Poznań studies in the philosophy of the sciences and the humanities, Vol. 91, pp. 185–209). Amsterdam: Rodopi.

Batens, D. (2007). Content guidance in formal problem solving processes. In O. Pombo & A. Gerner (Eds.), *Abduction and the process of scientific discovery* (pp. 121–156). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Batens, D. (2008). The role of logic in philosophy of science. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 47–57). London/New York: Routledge.

Batens, D., & Meheus, J. (1996). In-world realism *vs.* reflective realism. In I. Douven & L. Horsten (Eds.), *Realism in the sciences* (pp. 35–53). Leuven: Universitaire Pers.

Batens, D., & Meheus, J. (2001). On the logic and pragmatics of the process of explanation. In M. Kiikeri & P. Ylikoski (Eds.), *Explanatory connections*. Electronic essays dedicated to Matti Sintonen. http://www.valt.helsinki.fi/kfil/matti/.

Batens, D., & Priest, G. (2008). Graham Priest and Diderik Batens interview each other. *The Reasoner, 2*(8), 2–4.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive science*. New York: Routledge.

Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan

Campbell, D. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*, 380–400.

Campbell, D. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl R. Popper* (pp. 413–463). LaSalle: Open Court.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue Internationale de Philosophie, 4*, 20–40.

Changizi, M. (2011). *Harnessed: How language and music mimicked nature and transformed ape to man*. Dallas: BenBella Books.

Cziko, G. (1995). *Without miracles: Universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT.

Davidson, D. (1978). What metaphors mean. *Critical Inquiry, 5*, 31–47. (Reprinted in Davidson's *Inquiries into truth and interpretation* 2nd ed., pp. 245–264, 2001, Oxford: Clarendon Press)

Davidson, D. (1986). A nice derangement of epitaphs. In E. LePore (Ed.), *Perspectives on the philosophy of Donald Davidson* (pp. 433–446). Oxford: Basil Blackwell. (Reprinted in Davidson's *Truth, Language and History: Philosophical Essays*, 2005, Oxford: Clarendon Press)

Dennett, D. (1995). *Darwin's dangerous idea*. New York: Simon and Schuster.

Dewey, J. (1917). The need for a recovery of philosophy. In J. Dewey (Ed.), *Creative intelligence: Essays in the pragmatic attitude* (pp. 3–69). New York: Henry Holt. (As reprinted in *John Dewey: Essays on philosophy and education 1916–1917 (the middle works, 1899–1924)*, Vol. 10, pp. 3–48, by J. A. Boydston, Ed., Carbondale: Southern ILlinois University Press)

Dewey, J. (1929). *The quest for certainty*. New York: Minton, Balch and Co. (Reprinted as Vol. 4 of *John Dewey: The later works, 1925–1953*, by Jo A. Boydston, Ed., Carbondale: Southern Illinois University Press)

Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Holt.

Feigenbaum, E. A., Buchanan, G., & Lederberg, J. (1971). On generality and problem solving: A case study using the DENDRAL program. *Machine Intelligence, 7*, 165–190.

Feyerabend, P. (1989). Realism and the historicity of knowledge. *Journal of Philosophy, 86*, 393–406.

Gentner, D. (1989). The mechanisms of analogical reasoning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge: Cambridge University Press.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*, 524–575.

Haack, S. (1996). *Deviant logic, fuzzy logic: Beyond the formalism*. Chicago: University of Chicago Press.

Heeffer, A. (2008). The emergence of symbolic algebra as a shift in predominant models. *Foundations of Science, 13*, 149–161.

Hempel, C. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie, 11*, 41–63. (Reprinted in significantly revised form with a postscript in Hempel's *Aspects of Scientific Explanation*, pp. 101–122, New York: Free Press)

Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Holyoak, K., & Thagard, P. (1989). *Mental leaps*. Cambridge, MA: MIT.

Johnson-Laird, P. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Koza, J. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT.

Kuhn, T. S. (1962). *The structure of scientific revolutions* (2nd ed., 1970), adds "Postscript–1969." Chicago: University of Chicago Press.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge: Cambridge University Press.

Lenat, D. (1978). The ubiquity of discovery. *Artificial Intelligence, 9*, 257–285

Meheus, J. (1993). Adaptive logic in scientific discovery: The case of Clausius. *Logique et Analyse, 143–144*, 359–391. Appeared 1996.

Meheus, J. (1999a). Clausius' discovery of the first two laws of thermodynamics. A paradigm of reasoning from inconsistencies. *Philosophica, 63*, 89–117. Appeared 2001.

Meheus, J. (1999b). Deductive and ampliative adaptive logics as tools in the study of creativity. *Foundations of Science, 4*, 325–336.

Meheus, J. (2000). Analogical reasoning in creative problem solving processes: Logico-philosophical perspectives. In F. Hallyn (Ed.), *Metaphor and analogy in the sciences* (pp. 17–34). Dordrecht: Kluwer.

Meheus, J. (Ed.). (2002). *Inconsistency in science*. Dordrecht: Kluwer.

Meheus, J. (2004). Adaptive logics and the integration of induction and deduction. In F. Stadler (Ed.), *Induction and deduction in the sciences* (pp. 93–120). Dordrecht/Boston: Kluwer

Meheus, J. (2007). Adaptive logics for abduction and the explication of explanation-seeking processes. In O. Pombo & A. Gerner (Eds.), *Abduction and the process of scientific discovery* (pp. 97–119). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Nersessian, N. (2002). Inconsistency, generic modeling, and conceptual change in science. In J. Meheus (Ed.), *Inconsistency in science* (pp. 197–211). Dordrecht: Kluwer.

Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Nickles, T. (1980a). Can scientific constraints be violated rationally? In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 285–315). Dordrecht: Reidel.

Nickles, T. (Ed.). (1980b). *Scientific discovery: Case studies*. Dordrecht: Reidel.

Nickles, T. (Ed.). (1980c). *Scientific discovery, logic, and rationality*. Dordrecht: Reidel.

Nickles, T. (2002a). The discovery-justification distinction and professional philosophy of science: Comments on the first day's five papers. In J. Schikore & F. Steinle (Eds.), *Revisiting discovery and justification* (pp. 67–78). Berlin: Max-Planck-Institut für Wissenschaftsgeschichte. Preprint 211.

Nickles, T. (2002b). From Copernicus to Ptolemy: Inconsistency and method. In J. Meheus (Ed.), *Inconsistency in science* (pp. 1–33). Dordrecht: Kluwer.

Nickles, T. (2003a). Evolutionary models of innovation and the Meno problem. In L. Shavinina (Ed.), *International handbook on innovation* (pp. 54–78). Amsterdam: Elsevier Scientific Publications.

Nickles, T. (2003b). Normal science: From logic of science to case-based and model-based reasoning. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 142–177). Cambridge: Cambridge University Press.

Nickles, T. (2006). Heuristic appraisal: Context of discovery or justification? In J. Schickore & F. Steinle (Eds.), *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction* (pp. 159–182). Dordrecht: Springer.

Nickles, T. (2009). The strange story of scientific method. In J. Meheus & T. Nickles (Eds.), *Models of discovery and creativity* (pp. 167–207). Dordrecht: Springer.

Nickles, T. (2012). Some puzzles about Kuhn's exemplars. In V. Kindi & T. Arabatzis (Eds.), *Kuhn's 'the structure of scientific revolutions' revisited* (pp. 112–133). London: Routledge.

Nickles, T., & McCollum-Nickles, G. (2002). James on bootstraps, evolution, and life. In B. Babich (Ed.), *Hermeneutic philosophy of science, Van Gogh's eyes, and god: Essays in honor of Patrick Heelan, S.J.* (pp. 361–376). Dordrecht: Kluwer.

Norton, J. (2003). A material theory of induction. *Philosophy of Science, 17*, 647–670.

Norton, J. (2010). There are no universal rules for induction. *Philosophy of Science, 77*, 765–777.

Paley, W. (2008). In M. D. Eddy & D. Knight (Eds.), *Natural theology: Or, evidences of the existence and attributes of the Deity* (1st edition published in 1802). Oxford: Oxford University Press.

Pease, A., Smaill, A., Colton, S., Ireland, A., Teresa Llan, M., Ramezani, R., Grov, G., & Gube, M. (2010). Applying Lakatos-style reasoning to AI problems. In J. Vallverdú (Ed.), *Thinking machines and the philosophy of computer science: Concepts and principles* (pp. 149–174). Hershey: IGI Global.

Peirce, C. (1877). The fixation of belief. *Popular Science Monthly, 12*, 1–15. (Reprinted in C. Hartshorne & P. Weiss (Eds.), *Collected Papers* 5.358–387. Cambridge, MA: Harvard University's Belknap Press.)

Peirce, C. (1878). How to make our ideas clear. *Popular Science Monthly, 5*, 388–410. (Reprinted in C. Hartshorne & P. Weiss (Eds.), *Collected Papers* 5.358–387. Cambridge, MA: Harvard University's Belknap Press.)

Peirce, 1899, F.R.L, Unpublished ms. (Printed in C. Hartshorne & P. Weiss (Eds.), *Collected papers*. Cambridge, MA: Harvard University's Belknap Press, 1932.)

Popper, K. (1963). *Conjectures and refutations*. New York: Basic Books.

Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon.

Price, D. (1963). *Little science, big science*. New York: Columbia University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20–43. (Reprinted in Quine's *From a logical point of view*, 1953, Cambridge, MA: Harvard University Press)

Rorty, R. (1989). *Contingency, irony and solidarity*. Cambridge: Cambridge University Press.

Schilpp, P. (Ed.). (1949). *Albert Einstein: Philosopher-scientist* (Library of Living Philosophers, Vol. VII). Open Court Publishing Company, LaSalle, IL.

Shapere, D. (1984). *Reason and the search for knowledge*. Dordrecht: Reidel.

Simon, H. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4*, 181–201.

Simon, H. (1976). Bradie on polanyi on the meno paradox. *Philosophy of Science, 43*, 147–151. (Reprinted as chapter 5.5 of Simon's *Models of Discovery*, pp. 338–341, 1977, Dordrecht: Reidel)

Simon, H., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. Gregg (Ed.), *Knowledge and cognition* (pp. 105–128). Potomac: Lawrence Erlbaum. (As reprinted in H. Simon, *Models of thought*, vol. 1, pp. 329–346, 1979, New Haven: Yale University Press)

Simon, H., Newell, A., & Shaw, J. C. (1962). The process of creative thinking. In H. Gruber, G. Terrell, & M. Wetheimer (Eds.), *Contemporary approaches to creative thinking* (pp. 63–119). New York: Lieber-Atherton. (As reprinted in H. Simon, *Models of thought*, vol. 1, pp., 144–174, 1979, New Haven: Yale University Press)

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1979). An introduction to reasoning. New York: Macmillan.

Van Bouwel, J., & Weber, E. (2008). A pragmatist defense of non-relativistic explanatory pluralism in history and social science. *History and Theory, 47*, 168–182.

Weber, E., & Vanderbeeken, R. (2001). A pragmatic approach to the explanation of actions. In J. Blasius, J. Hox, E. de Leeuw, & P. Schmidt (Eds.), Social science methodology in the new millenium. Berlin: Leske and Budrich. CD-Rom

Wolpert, D. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation, 8*, 1341–1390.

Wolpert, D., & Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*, 67–82. Condensed version of 1995 Santa Fe Institute Technical Report, SFI TR 95-02-010, “No Free Lunch Theorems for Search”.

Yovits, M., & Cameron, S. (Eds.). (1959). Self-organizing systems: Proceedings of an interdisciplinary conference. New York: Pergamon Press

Chapter 8
On the Propagation of Consistency in Some Systems of Paraconsistent Logic

Hitoshi Omori and Toshiharu Waragai

8.1 Introduction

One of the most well-known and widely studied classes of systems of paraconsistent logic is da Costa's sequence of systems $\mathrm{C}_n (1 \leq n \leq \omega)$ (cf. da Costa 1974). The distinctive characteristic of these systems is the notion of "behaving classically", which controls the explosion of contradictions. There are two interesting preceding works related to this notion. One is the work of Carnielli, Coniglio and Marcos (cf. Carnielli et al. 2007; Carnielli and Marcos 2002) which introduces a *primitive* notion called consistency (written as $\circ A$ for the consistency of a proposition A) playing the same role as that of behaving classically in da Costa's systems. Their systems, known as Logics of Formal Inconsistency (LFIs hereafter), are of great interest since they make the essence of da Costa's idea clear. Taking into account of their work, we shall use "behaving consistently" instead of "behaving classically" hereafter. The other is the work of Waragai and Shidori (cf. Waragai and Shidori 2007) which clarifies the relation between a restricted version of the so-called de Morgan's law and the propagation of behaving consistently. More concretely, they discovered that a restricted version of de Morgan's law enables us to derive the propagation of consistency in certain cases.

The importance of the propagation of consistency in LFIs can be explained briefly as follows. Take the formula $(C \supset (A \wedge B)) \supset (\neg(A \wedge B) \supset \neg C)$ as an example. This formula is certainly derivable in classical propositional calculus, though it is not a thesis of LFIs in general. However, in the spirit of LFIs we *do* consider

H. Omori (✉)
The Graduate Center, City University of New York, New York, NY, USA
e-mail: hitoshiomori@gmail.com

T. Waragai
Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Tokyo, Japan
e-mail: waragai.t.aa@m.titech.ac.jp

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_8,
© Springer Science+Business Media Dordrecht 2014

this formula to be derivable under a certain condition. Indeed, in the system **mbC**, which is the base system in Carnielli et al. (2007), the formula $\circ(A \wedge B) \supset ((C \supset (A \wedge B)) \supset (\neg(A \wedge B) \supset \neg C))$ is derivable. But then in this case, how can we relate $\circ A$ and $\circ B$ to $\circ(A \wedge B)$? It is the propagation of consistency that answers this question. However, this propagation of consistency was left mostly unanalyzed until the work of Waragai and Shidori. The results known before their work were the case for negation in the system C_n explored by Guillaume,[1] and also the case for negation in an LFI stronger than the system **Bk** of Avron, explored by Carnielli and Marcos. The situation at that time is summarized briefly by Carnielli and Marcos as follows:

> Now, what do we know about the propagation of consistency through other connectives besides negation? Not much, so far. (Carnielli and Marcos 2002, p. 63)

It was the work of Waragai and Shidori that paved the way for the exploration of the propagation of consistency. Immediately after getting acquainted with their results, Guillaume presented some further results in the area. In his work, the systems were similar to those of Carnielli and Marcos, as they also contained consistency as a primitive connective. The present paper proceeds along the lines of this research of Guillaume.

Based on these observations, the aim of the present paper is twofold. First, we provide a new characterization of consistency in the system **Bk**. This characterization makes explicit that the consistency operator reflects the way propositions A and $\neg A$ behave in a given theory. There exist characterization results for da Costa's systems C_n and for Waragai and Shidori's system PCL1C, but these systems all have the notion of consistency *defined* by other connectives employed in classical logic. The new result of the present paper is that we can also characterize the notion of consistency when it is taken as a *primitive* connective. The difficulty here lies in finding appropriate axioms for the consistency operator and it turns out that the axiom added by Avron to the system **mbC** plays an important role. Second, we examine the propagation of consistency in the system **Bk** and its first-order extension **Bk***, which has been left totally untouched so far. As is known, there are two main ways of formulating the propagation of consistency. One is that a compound formula behaves consistently if *all* of its components behave consistently, and the other is that a compound formula behaves consistently if *some* of its components behave consistently. Our main result shows that in both of these cases, formulas expressing the propagation of consistency for the classical connectives, that is, negation, implication, disjunction and conjunction, are equivalent to certain forms of de Morgan's laws not containing the consistency operator at all. We also examine the propagation condition for the case of the predicate calculus which was left open even in the work of Guillaume.

[1] As is noted in da Costa (1974, p. 500), Guillaume proved that the propagation condition for negation, which was originally stated as an axiom, can be proved in da Costa's systems $C_n(1 \leq n < \omega)$.

8.2 Preliminaries

The present section is devoted to preliminaries. We first revisit one of the most basic LFIs of Carnielli, Coniglio and Marcos known as **mbC**, and then recall its extension **Bk** which was introduced by Avron. Then, we show that many formulas provable in da Costa's systems are already provable in the system **Bk**. We begin with the definition of the system **mbC**.

Definition 1 (Carnielli & Coniglio & Marcos). The system **mbC** consists of the following axioms and rule of inference:

$$A \supset (B \supset A) \tag{A1}$$

$$(A \supset B) \supset ((A \supset (B \supset C)) \supset (A \supset C)) \tag{A2}$$

$$A \supset (B \supset (A \wedge B)) \tag{A3}$$

$$(A \wedge B) \supset A \tag{A4}$$

$$(A \wedge B) \supset B \tag{A5}$$

$$A \supset (A \vee B) \tag{A6}$$

$$B \supset (A \vee B) \tag{A7}$$

$$(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C)) \tag{A8}$$

$$A \vee (A \supset B) \tag{A9}$$

$$A \vee \neg A \tag{A10}$$

$$\circ A \supset (A \supset (\neg A \supset B)) \tag{A11}$$

$$\frac{A \quad A \supset B}{B} \tag{MP}$$

Following the usual convention, we define $A \equiv B$ as follows:

$$A \equiv B =_{\text{def.}} (A \supset B) \wedge (B \supset A)$$

Remark 1. As is mentioned in Carnielli et al. (2007, Remark 36), note that strong negation $\neg^*$ which is the classical negation can be defined in **mbC**. There are two keys for this fact. First, from (A11) it follows that $\circ A \wedge A \wedge \neg A$ is a bottom particle, i.e. $(\circ A \wedge A \wedge \neg A) \supset B$ is provable. Therefore we can introduce a bottom particle $\perp$ as follows:

$$\perp =_{\text{def.}} \circ X \wedge X \wedge \neg X$$

Second, the system **mbC** contains axioms from (A1) to (A9) with the rule (MP), that is, the positive fragment of classical propositional calculus. Combining these two facts, we can introduce the strong negation as follows[2]:

$$\neg^* A =_{\text{def.}} A \supset \bot$$

It should also be noted that the so-called Deduction Theorem holds in **mbC** and its extensions since Modus Ponens is the only rule of inference and we have both (A1) and (A2) as axiom schemata.

Remark 2. Many basic and interesting results for **mbC** such as the completeness result with respect to valuation semantics and possible-translation semantics are provided in Carnielli et al. (2007).

Remark 3. It should be noted that there is earlier work by Guillaume in which he introduces the system **mbC** (Guillaume 2007). The context in which he considers the system is different from that of Carnielli, Coniglio and Marcos in that Guillaume employs the intuitionistic positive calculus as his startingpoint. The system **mbC** is named $\mathbf{b}^{\mathbf{C}}\mathbf{C}$ in Guillaume (2007).

We next introduce the system **Bk** of Avron.

Definition 2 (Avron). The system **Bk** is obtained from **mbC** by adding the following formula:

$$\circ A \vee (A \wedge \neg A) \tag{k}$$

Remark 4. Note that Avron refers to the system **mbC** as simply **B** and adds some letters which are the names of formulas to be added. Also a completeness result for the system **Bk** with respect to non-deterministic semantics is provided in Avron (2009).

Remark 5. In fact, the above system introduced as **Bk** was independently introduced by the present authors under the name Σ. It should be noted that instead of adding the formula (k) to the system **mbC**, we added the following formula:

$$\neg^*(\circ A) \supset (A \wedge \neg A) \tag{$\star$}$$

Since ($\star$) is equivalent to (k) in **mbC**, **Bk** and Σ are equivalent.

In the remaining part of this section, we present some basic results regarding **Bk**. The following proposition confirms that the system **Bk** strictly extends the system **mbC**. This result was already known to Avron and is obtained as a corollary of the completeness theorem established by Avron, but we here give a direct proof of this result.

[2] We shall make full use of this strong negation in the appendix dedicated to proofs of and we write '(CN)' (Classical Negation) in the proof lines.

Proposition 1. *The formula (k) is not derivable in* ***mbC****, and therefore the system* ***Bk*** *strictly extends the system* ***mbC****.*

Proof. In order to prove the desired result, it would be sufficient to employ the following four-valued evaluation tables:

$\wedge$	1	i	j	0
1	1	i	j	0
i	i	i	0	0
j	j	0	j	0
0	0	0	0	0

$\vee$	1	i	j	0
1	1	1	1	1
i	1	i	1	i
j	1	1	j	j
0	1	i	j	0

$\supset$	1	i	j	0
1	1	i	j	0
i	1	1	j	j
j	1	i	1	i
0	1	1	1	1

	$\neg$	$\circ$
1	0	1
i	1	0
j	1	0
0	1	1

Note here that 1 is the only designated value. Then it is straightforward to check that all the axioms of **mbC** take the value 1 and also MP preserves validity but (k) takes the non-designated value when we assign i or j to A. □

In the work of Carnielli, Coniglio and Marcos, the smallest extension of **mbC** considered was **mCi**, and in the final section of Carnielli et al. (2007), they remarked as follows:

> At any rate, the study of extensions of **mbC** that do not extend **mCi** seems a very attractive enterprise. (Carnielli et al. 2007, p. 72)

Note that the system **mCi** can be obtained by adding the following formulas to the system **mbC**:

$$\neg(\circ A) \supset (A \wedge \neg A) \tag{ci}$$

$$(\circ\neg^n(\circ A)) \quad (n \geq 0) \tag{(cc)$_n$}$$

where $\neg^0 A = A$ and $\neg^{n+1} A = \neg(\neg^n A)$. Now, after the work of Carnielli, Coniglio and Marcos, Avron explicitly introduced the system **Bk** which strictly includes **mbC** and is strictly included in **mCi**. The latter result was also proved by Avron as a corollary of his completeness theorem, but here we again give a direct proof of this result.

Proposition 2. *The system* ***Bk*** *is strictly included in the system* ***mCi****.*

Proof. In order to prove the desired result, it suffices to employ the following four-valued evaluation tables:

$\wedge$	1	i	j	0
1	1	i	j	0
i	i	i	0	0
j	j	0	j	0
0	0	0	0	0

$\vee$	1	i	j	0
1	1	1	1	1
i	1	i	1	i
j	1	1	j	j
0	1	i	j	0

$\supset$	1	i	j	0
1	1	i	j	0
i	1	1	j	j
j	1	i	1	i
0	1	1	1	1

	$\neg$	$\circ$
1	0	1
i	1	j
j	1	i
0	1	1

Note here that 1 is the only designated value. It is straight forward to check that all the axioms of **Bk** take the value 1 and also MP preserves validity but the formula $\neg(\circ A) \supset (A \wedge \neg A)$, which is an axiom of the system **mCi**, takes the non-designated value when we assign i or j to A. □

Finally, we present some theorems of **Bk**. These formulas are not only useful and important in the following proofs, but also interesting in view of da Costa's systems. More concretely, it turns out that many of the characteristic formulas provable in da Costa's systems are also provable in the system **Bk**.

Theorem 1. *The following formulas can be derived in* ***Bk****:*

$$\circ A \vee A \tag{8.1}$$

$$\circ A \vee \neg A \tag{8.2}$$

$$\neg^* A \supset \neg A \tag{8.3}$$

$$\circ A \equiv \neg^*(A \wedge \neg A) \tag{8.4}$$

$$\neg^* \circ A \equiv (A \wedge \neg A) \tag{8.5}$$

$$\circ A \equiv (\neg A \supset \neg^* A) \tag{8.6}$$

$$\circ A \equiv (\neg^* A \equiv \neg A) \tag{8.7}$$

$$\circ A \equiv (\neg^* \neg A \equiv A) \tag{8.8}$$

$$\neg^* A \equiv (\neg A \wedge \circ A) \tag{8.9}$$

$$\neg^* \neg A \equiv (A \wedge \circ A) \tag{8.10}$$

Proof. See Appendix. □

Remark 6. Note that of these statements only (8.3) is provable in the system **mbC**. This can be verified by the four-valued evaluation tables we employed in the proof of Proposition 1. Also, note that we may consider systems which extend **mbC** but do not extend **Bk**. Indeed, if we add only one of (8.1) or (8.2), we obtain such systems. This can be proved again as a corollary of Avron's completeness theorem.

8.3 Main Results on the System Bk

We now present the main results for the system **Bk**. First, we examine some notions for theories which will be essential for the characterization of consistency. A characterization of $\circ A$ follows, which gives a nice intuitive reading of $\circ A$. In the final two subsections we examine the propagation condition.

8.3.1 Relative Maximality, Implication-Saturatedness and Non-triviality of Theories

Here, we examine the relation of three notions for theories. Let us begin with definitions. We refer to the language of **Bk** as $\mathcal{L}_{\mathbf{Bk}}$.

Definition 3. Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. We define the notion of non-triviality as follows:

$$\Gamma \text{ is } \textit{non-trivial} \quad \text{iff} \quad \exists A(\Gamma \nvdash_{\mathbf{Bk}} A)$$

Definition 4 (Batens). Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. We define implication-saturation as follows:

$$\Gamma \text{ is } \textit{implication-saturated} \quad \text{iff} \quad \forall A(A \notin \Gamma \Rightarrow \forall B(A \supset B \in \Gamma))$$

Remark 7. Note that Batens introduced this notion in Batens (1980) in order to provide completeness theorems for some systems of paraconsistent logic.

Definition 5 (Lindenbaum and Łoś). Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. We define maximality relative to F as follows:

$$\Gamma \text{ is } \textit{maximal relative to } F \quad \text{iff } (\Gamma \nvdash_{\mathbf{Bk}} F \ \text{ and } \ \forall G(G \notin \Gamma \Longrightarrow \Gamma \cup \{G\} \vdash_{\mathbf{Bk}} F))$$

Based on this, we define relative maximality as follows:

$$\Gamma \text{ is } \textit{relatively maximal} \quad \text{iff} \quad \Gamma \text{ is maximal relative to } F \text{ for a formula } F$$

Remark 8. In view of the above definition, we can see that relatively maximal theories are a natural generalization of maximal consistent theories of classical logic. Note that Loparić introduced the notion of a theory being maximal relative to F under a different name, F-saturated, in Loparić (1986)[3] for the purpose of proving the completeness theorem with respect to a valuation semantics for da Costa's system C_ω.

We prepare two more lemmas towards the desired result.

Lemma 1. *Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then we have the following equivalence when Γ is relatively maximal:*

$$A \in \Gamma \iff \Gamma \vdash_{\mathbf{Bk}} A$$

Proof. The left to right implication is immediate. For the other direction, suppose that $\Gamma \vdash_{\mathbf{Bk}} A$ and $A \notin \Gamma$. Then by the latter assumption and the definition of Γ being relatively maximal, we obtain $\Gamma \cup \{A\} \vdash_{\mathbf{Bk}} F$ and therefore $\Gamma \vdash_{\mathbf{Bk}} A \supset F$ by the Deduction Theorem. Thus, if we combine this with the first assumption, we get $\Gamma \vdash_{\mathbf{Bk}} F$. But this is absurd in view of Definition 5. □

[3] There is a shorter version of Loparić (1986), which appeared as Loparić (1977), and in this earlier paper the definition of F-saturated theory is slightly different.

Lemma 2. *Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then we have the following equivalence when Γ is non-trivial and implication-saturated:*

$$A \in \Gamma \iff \Gamma \vdash_{\mathbf{Bk}} A$$

Proof. Analogous to the proof of the previous lemma. □

Now we are in a position to prove the relation between having a relatively maximal theory, and an implication-saturated and non-trivial theory.

Proposition 3. *Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then if Γ is non-trivial and implication saturated, then Γ is relatively maximal.*

Proof. By the non-triviality of Γ, we have $\Gamma \nvdash_{\mathbf{Bk}} X_0$ for some formula X_0. Then, by Lemma 2, we have $X_0 \notin \Gamma$. Since Γ is implication saturated, we have $G \notin \Gamma \Longrightarrow G \supset X_0 \in \Gamma$. Again by Lemma 2, we have $G \notin \Gamma \Longrightarrow \Gamma \vdash_{\mathbf{Bk}} G \supset X_0$ which is equivalent to $G \notin \Gamma \Longrightarrow \Gamma \cup \{G\} \vdash_{\mathbf{Bk}} X_0$ by the Deduction Theorem. This together with $\Gamma \nvdash_{\mathbf{Bk}} X_0$ shows that Γ is relatively maximal. □

Proposition 4. *Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then if Γ is relatively maximal, Γ is non-trivial and implication saturated.*

Proof. Suppose that Γ is relatively maximal. Then, by Definition 5, we have $\Gamma \nvdash_{\mathbf{Bk}} F_0$ for some F_0, which implies the non-triviality of Γ. We now prove that Γ is implication saturated by reductio ad absurdum. Suppose that $A \notin \Gamma$ and $A \supset B_0 \notin \Gamma$ for some B_0. Since Γ is relatively maximal, we have $\Gamma \cup \{A\} \vdash_{\mathbf{Bk}} F_0$ and $\Gamma \cup \{A \supset B_0\} \vdash_{\mathbf{Bk}} F_0$ where F_0 is the same formula as above. Therefore, by the Deduction Theorem, we have $\Gamma \vdash_{\mathbf{Bk}} A \supset F_0$ and $\Gamma \vdash_{\mathbf{Bk}} (A \supset B_0) \supset F_0$, which imply $\Gamma \vdash_{\mathbf{Bk}} (A \vee (A \supset B_0)) \supset F_0$. If we combine this with (A9), we obtain $\Gamma \vdash_{\mathbf{Bk}} F_0$, but this is absurd in view of the fact that $\Gamma \nvdash_{\mathbf{Bk}} F_0$ holds for Γ. □

We therefore obtain the following result:

Theorem 2. *Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then the following holds:*

$$\Gamma \textit{ is relatively maximal} \iff \Gamma \textit{ is non-trivial and implication saturated.}$$

8.3.2 Characterization of $\circ A$ in Bk

In this subsection, we provide an intuitively desirable characterization of $\circ A$ in **Bk**. This kind of characterization first appeared in Waragai and Omori (2010) for the system PCL1C, which is the extension of the system PCL1 obtained by adding Peirce's law. These two systems PCL1 and PCL1C also have the notion of behaving consistently being *defined* as in da Costa's systems C_n, but the formula employed for the definition of behaving consistently is different. Indeed, where the formula $\neg(A \wedge \neg A)$ is used in C_1, the formula $A \supset \neg\neg A$ is used in formulating the system PCL1.

Now, as is remarked in Waragai and Omori (2010) the characterization holds not only for the system PCL1C but also for the systems C_n $(1 \leq n < \omega)$. However, the question of whether the characterization holds for LFIs remained open. We shall here answer this question by extending the result of Waragai and Omori (2010) to the system **Bk** which has the notion of behaving consistently as a *primitive* notion.

We begin with the definition of a new notion which is key for the characterization to be given below.

Definition 6. Let Γ be a theory in $\mathcal{L}_{\mathbf{Bk}}$. Then,

- A is normal in Γ with respect to $\neg$ $\iff$ only one of $\Gamma \nvdash_{\mathbf{Bk}} A$ or $\Gamma \nvdash_{\mathbf{Bk}} \neg A$ holds.
- A is non-normal in Γ with respect to $\neg$ $\iff$ both $\Gamma \vdash_{\mathbf{Bk}} A$ and $\Gamma \vdash_{\mathbf{Bk}} \neg A$ hold.

Now we move to the proof for the characterization. For this purpose, we prepare two lemmas.

Lemma 3. *Let Γ be a relatively maximal theory in $\mathcal{L}_{\mathbf{Bk}}$. Then the following holds:*

$$\Gamma \nvdash_{\mathbf{Bk}} A \iff \Gamma \vdash_{\mathbf{Bk}} \neg^* A$$

Proof. For the case from left to right, suppose $\Gamma \nvdash_{\mathbf{Bk}} A$ and $\Gamma \nvdash_{\mathbf{Bk}} \neg^* A$. Then by Lemma 1, we have $A \notin \Gamma$ and $\neg^* A \notin \Gamma$. Therefore by the definition of a theory being relatively maximal, we obtain $\Gamma \cup \{A\} \vdash_{\mathbf{Bk}} F_0$ and $\Gamma \cup \{\neg^* A\} \vdash_{\mathbf{Bk}} F_0$. By applying the Deduction Theorem and (A8), we get $\Gamma \vdash_{\mathbf{Bk}} (A \vee \neg^* A) \supset F_0$, so making use of the law of the excluded middle with respect to $\neg^*$, we have $\Gamma \vdash_{\mathbf{Bk}} F_0$. But this is absurd in view of Definition 5. On the other hand, for the case from right to left, suppose $\Gamma \vdash_{\mathbf{Bk}} A$ and $\Gamma \vdash_{\mathbf{Bk}} \neg^* A$. Then we have $\Gamma \vdash_{\mathbf{Bk}} (A \wedge \neg^* A)$, so by applying the explosion law with respect to $\neg^*$, we get $\Gamma \vdash_{\mathbf{Bk}} F$ for any formula F. But again this is absurd in view of Definition 5. This completes the proof. □

Lemma 4. *Let Γ be a relatively maximal theory in $\mathcal{L}_{\mathbf{Bk}}$. Then it is not the case that both $\Gamma \nvdash_{\mathbf{Bk}} A$ and $\Gamma \nvdash_{\mathbf{Bk}} \neg A$ hold. In other words, either $\Gamma \vdash_{\mathbf{Bk}} A$ or $\Gamma \vdash_{\mathbf{Bk}} \neg A$ holds.*

Proof. Suppose $\Gamma \nvdash_{\mathbf{Bk}} A$ and $\Gamma \nvdash_{\mathbf{Bk}} \neg A$. Then by the former assumption and Lemma 3, we have $\Gamma \vdash_{\mathbf{Bk}} \neg^* A$. On the other hand, we have $\Gamma \vdash_{\mathbf{Bk}} \neg^* A \supset \neg A$ in view of (8.3). Therefore, we obtain $\Gamma \vdash_{\mathbf{Bk}} \neg A$. However, this is absurd in view of the latter assumption. □

Theorem 3. *Let Γ be a relatively maximal theory in $\mathcal{L}_{\mathbf{Bk}}$. Then the following results can be proved:*

- *A is normal in Γ with respect to $\neg$ $\iff$ $\Gamma \vdash_{\mathbf{Bk}} \circ A$*
- *A is non-normal in Γ with respect to $\neg$ $\iff$ $\Gamma \vdash_{\mathbf{Bk}} \neg^*(\circ A)$*

Proof. For the former, the proof runs as follows:

$$
\begin{aligned}
\Gamma \vdash_{\mathbf{Bk}} \circ A &\Longleftrightarrow \Gamma \vdash_{\mathbf{Bk}} \neg^{*}(A \wedge \neg A) \quad \text{(by (8.4))}\\
&\Longleftrightarrow \Gamma \nvdash_{\mathbf{Bk}} (A \wedge \neg A) \quad \text{(by Lemma 3)}\\
&\Longleftrightarrow \Gamma \nvdash_{\mathbf{Bk}} A \ \text{ or } \ \Gamma \nvdash_{\mathbf{Bk}} \neg A\\
&\Longleftrightarrow \Gamma \nvdash_{\mathbf{Bk}} A \ \text{ or } \ \Gamma \nvdash_{\mathbf{Bk}} \neg A) \ \& \ (\Gamma \vdash_{\mathbf{Bk}} A \ \text{ or } \ \Gamma \vdash_{\mathbf{Bk}} \neg A) \ \text{(by Lemma 4)}\\
&\Longleftrightarrow (\Gamma \nvdash_{\mathbf{Bk}} A \ \& \ \Gamma \vdash_{\mathbf{Bk}} \neg A) \ \text{ or } \ (\Gamma \vdash_{\mathbf{Bk}} A \ \& \ \Gamma \nvdash_{\mathbf{Bk}} \neg A)\\
&\Longleftrightarrow A \text{ is } \textbf{normal in } \Gamma \textbf{ with respect to } \neg
\end{aligned}
$$

For the latter, we can prove it as follows:

$$
\begin{aligned}
\Gamma \vdash_{\mathbf{Bk}} \neg^{*}(\circ A) &\Longleftrightarrow \Gamma \vdash_{\mathbf{Bk}} A \wedge \neg A \quad \text{(by (8.5))}\\
&\Longleftrightarrow \Gamma \vdash_{\mathbf{Bk}} A \ \ \& \ \ \Gamma \vdash_{\mathbf{Bk}} \neg A\\
&\Longleftrightarrow A \text{ is } \textbf{non-normal in } \Gamma \textbf{ with respect to } \neg
\end{aligned}
$$

This completes the proof. □

Remark 9. As the above theorem shows, we obtain a nice reading of $\circ A$ in the system **Bk**. This was not possible in **mbC** since (8.4) is not derivable (cf. Proposition 1). However, if we have (8.4) as a theorem in addition to some conditions, then we can read $\circ A$ in a way which seems to reflect the intention of da Costa quite well.

8.3.3 Results Related to Propagation of Consistency I

We shall start to examine the formula expressing the propagation of consistency. It will be proved that in the system **Bk** and its extended systems, formulas expressing the propagation of consistency are actually equivalent to formulas which do not contain the consistency operator at all.

Proposition 5. *The following formulas can be derived in* ***Bk****:*

$$\circ A \supset (\circ(\neg A) \equiv (\neg\neg A \supset A)) \tag{8.11}$$

$$(\circ A \wedge \circ B) \supset (\circ(A \wedge B) \equiv (\neg(A \wedge B) \supset (\neg A \vee \neg B))) \tag{8.12}$$

$$(\circ A \wedge \circ B) \supset (\circ(A \vee B) \equiv (\neg(A \vee B) \supset (\neg A \wedge \neg B))) \tag{8.13}$$

$$(\circ A \wedge \circ B) \supset (\circ(A \supset B) \equiv (\neg(A \supset B) \supset (A \wedge \neg B))) \tag{8.14}$$

Proof. See Appendix. □

We then immediately obtain the following result:

Theorem 4. *The following formulas can be derived in* ***Bk****:*

$$(\circ A \supset (\circ\neg A)) \equiv (\circ A \supset (\neg\neg A \supset A)) \tag{8.15}$$

$$((\circ A \wedge \circ B) \supset \circ (A \wedge B)) \equiv ((\circ A \wedge \circ B) \supset (\neg(A \wedge B) \supset (\neg A \vee \neg B))) \tag{8.16}$$

$$((\circ A \wedge \circ B) \supset \circ (A \vee B)) \equiv ((\circ A \wedge \circ B) \supset (\neg(A \vee B) \supset (\neg A \wedge \neg B))) \tag{8.17}$$

$$((\circ A \wedge \circ B) \supset \circ (A \supset B)) \equiv ((\circ A \wedge \circ B) \supset (\neg(A \supset B) \supset (A \wedge \neg B))) \tag{8.18}$$

Proof. Just apply the theorem $(A \supset (B \equiv C)) \supset ((A \supset B) \equiv (A \supset C))$ of **Bk** to the formulas from (8.11) to (8.14) in order to derive those from (8.15) to (8.18) respectively. □

Remark 10. These equivalences can be seen as a generalization of the results of Waragai and Shidori in their Waragai and Shidori (2007). It was their contribution that propagation axioms can be *proved* by making use of a restricted form of de Morgan's laws. Since their discovery was already examined semantically by Guillaume in Guillaume (2007), we here examined it syntactically.

The next theorem, which was never recognized in the preceding works, shows that the restricted form of de Morgan's laws are actually equivalent to formulas not containing consistency at all.

Theorem 5. *The following formulas can be derived in* ***Bk****:*

$$(\circ A \supset (\neg\neg A \supset A)) \equiv (\neg\neg A \supset A) \tag{8.19}$$

$$((\circ A \wedge \circ B) \supset (\neg(A \wedge B) \supset (\neg A \vee \neg B))) \equiv (\neg(A \wedge B) \supset (\neg A \vee \neg B)) \tag{8.20}$$

$$((\circ A \wedge \circ B) \supset (\neg(A \vee B) \supset (\neg A \wedge \neg B))) \equiv (\neg(A \vee B) \supset ((\neg A \vee (B \wedge \neg B)) \wedge (\neg B \vee (A \wedge \neg A)))) \tag{8.21}$$

$$((\circ A \wedge \circ B) \supset (\neg(A \supset B) \supset (A \wedge \neg B))) \equiv (\neg(A \supset B) \supset ((A \vee (B \wedge \neg B)) \wedge (\neg B \vee (A \wedge \neg A)))) \tag{8.22}$$

Proof. See Appendix. □

Combining Theorems 4 and 5, we obtain the following result:

Theorem 6. *The following formulas can be derived in* ***Bk****:*

$$(\circ A \supset \circ \neg A) \equiv (\neg\neg A \supset A) \tag{8.23}$$

$$((\circ A \wedge \circ B) \supset \circ (A \wedge B)) \equiv (\neg(A \wedge B) \supset (\neg A \vee \neg B)) \tag{8.24}$$

$$((\circ A \wedge \circ B) \supset \circ (A \vee B)) \equiv (\neg(A \vee B) \supset ((\neg A \vee (B \wedge \neg B)) \wedge (\neg B \vee (A \wedge \neg A)))) \tag{8.25}$$

$$((\circ A \wedge \circ B) \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset ((A \vee (B \wedge \neg B)) \wedge (\neg B \vee (A \wedge \neg A)))) \tag{8.26}$$

Remark 11. The above results show that in the system **Bk**, and therefore in all the extensions of **Bk**, the propagation condition for classical connectives can be replaced by certain forms of de Morgan's laws. These kind of equivalences originally appeared in the formulation of C_1 given by Béziau in Béziau 1993, Théorème 3. Although Béziau's work is limited to the system C_1, the fact is that his formulation also works in **Bk** and its extensions which are far more general systems compared to C_1.

8.3.4 Results Related to Propagation of Consistency II

In Sect. 8.3.3, we observed that the propagation conditions in which the antecedent is a *conjunction* can be replaced by certain forms of de Morgan's laws. In this subsection, we consider the case where the antecedent of the conditions is a *disjunction*.

Proposition 6. *The following formulas can be derived in **Bk**:*

$$(\circ A \supset \circ (A \wedge B)) \equiv (\neg(A \wedge B) \supset (B \supset \neg A)) \tag{8.27}$$

$$(\circ A \supset \circ (A \vee B)) \equiv (\neg(A \vee B) \supset (\neg A \wedge (B \supset A))) \tag{8.28}$$

$$(\circ A \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset (A \wedge (B \supset \neg A))) \tag{8.29}$$

$$(\circ B \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset ((A \vee B) \wedge \neg B)) \tag{8.30}$$

Proof. See Appendix. □

Making full use of the above result, we obtain the following result.

Theorem 7. *The following formulas can be derived in **Bk**:*

$$((\circ A \vee \circ B) \supset \circ (A \wedge B)) \equiv (\neg(A \wedge B) \supset ((A \supset \neg B) \wedge (B \supset \neg A))) \tag{8.31}$$

$$((\circ A \vee \circ B) \supset \circ (A \vee B)) \equiv (\neg(A \vee B) \supset ((\neg A \wedge \neg B) \wedge (A \equiv B))) \tag{8.32}$$

$$((\circ A \vee \circ B) \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset ((A \wedge \neg B) \wedge (B \supset \neg A))) \tag{8.33}$$

Proof. See Appendix. □

Remark 12. In view of the above results, we conclude that the systems C_n^+ which were introduced by Béziau in Béziau (1990) can also be formulated by making use of consistency operator only once.

8.4 Extending the Results on Propagation to First Order Predicate Calculus

In Sect. 8.3, we offered some interesting equivalences related to the propagation conditions in the scope of propositional calculus. It is then natural to ask whether those results can be extended to predicate calculus. The answer is in the affirmative as we shall see in the following. We begin with the definition of the system in which we prove the desired results.

Definition 7. Let us consider an extension of **Bk** by adding the following formulas:

$$\frac{C \supset A(x)}{C \supset \forall x A(x)} \tag{R1}$$

$$\forall x A(x) \supset A(t) \tag{A12}$$

$$\frac{A(x) \supset C}{\exists x A(x) \supset C} \tag{R2}$$

$$A(t) \supset \exists x A(x) \tag{A13}$$

Note here that the above postulates are subject to the usual restrictions. We shall refer to this extension of **Bk** as **Bk***.

Remark 13. Note that in Avron and Zamansky (2007) the system **Bk*** enriched with another axiom similar to the axiom (VII) in C_n^* of da Costa (cf. da Costa 1974, p. 501) is proved to be complete with respect to non-deterministic semantics. We could have also followed Avron and Zamansky to add another axiom but since it is not necessary for the result we intend to prove, we shall employ the above system.

Remark 14. Though we provided a characterization of the consistency operator in **Bk**, we will not attempt it in the predicate calculus. We just note that the proof in the propositional case cannot be easily adapted to the case of predicate calculus. Indeed, the deduction theorem does not hold for predicate calculus in general.

8.4.1 Results Related to Propagation of Consistency I

We shall now extend the results we obtained in Sect. 8.3. The first target is a natural extension of Theorem 4 and for this purpose, we prove the following lemma:

Lemma 5. *The following formulas can be derived in* ***Bk****:*

$$\forall x \circ (A(x)) \supset (\circ(\forall x A(x)) \equiv (\neg \forall x A(x) \supset \exists x \neg A(x))) \tag{8.34}$$

$$\forall x \circ (A(x)) \supset (\circ(\exists x A(x)) \equiv (\neg \exists x A(x) \supset \forall x \neg A(x))) \tag{8.35}$$

Proof. See Appendix. □

Theorem 8. *The following formulas can be derived in* ***Bk*****:*

$$(\forall x \circ (A(x)) \supset (\circ \forall x A(x))) \equiv (\forall x \circ A(x) \supset (\neg \forall x A(x) \supset \exists x \neg A(x))) \tag{8.36}$$

$$(\forall x \circ A(x) \supset \circ (\exists x A(x))) \equiv (\forall x \circ A(x) \supset (\neg \exists x A(x) \supset \forall x \neg A(x))) \tag{8.37}$$

Proof. Just apply the theorem $(A \supset (B \equiv C)) \supset ((A \supset B) \equiv (A \supset C))$ of **Bk** to (8.34) and (8.35) in order to obtain (8.36) and (8.37) respectively. □

Remark 15. The above result shows that we can generalize Theorem 4.

The second target is the first order version of Theorem 6. Once again, the proof runs in an analogous manner.

Theorem 9. *The following formulas can be derived in* ***Bk*****:*

$$(\forall x \circ A(x) \supset (\neg \forall x A(x) \supset \exists x \neg A(x))) \equiv (\neg \forall x A(x) \supset \exists x \neg A(x)) \tag{8.38}$$

$$(\forall x \circ A(x) \supset (\neg \exists x A(x) \supset \forall x \neg A(x))) \equiv (\neg \exists x A(x) \supset \forall x (\neg A(x) \vee \exists x (A(x) \wedge \neg A(x)))) \tag{8.39}$$

Proof. See Appendix. □

Combining the above two theorems, we obtain the following result:

Theorem 10. *The following formulas can be derived in* ***Bk*****:*

$$(\forall x \circ A(x) \supset \circ (\forall x A(x))) \equiv (\neg \forall x A(x) \supset \exists x \neg A(x)) \tag{8.40}$$

$$(\forall x \circ A(x) \supset \circ (\exists x A(x))) \equiv (\neg \exists x A(x) \supset \forall x (\neg A(x) \vee \exists x (A(x) \wedge \neg A(x)))) \tag{8.41}$$

8.4.2 Results Related to Propagation of Consistency II

Finally, we prove the first order version of Theorem 7.

Theorem 11. *The following formulas can be derived in* ***Bk*****:*

$$(\exists x \circ A(x) \supset \circ (\forall x A(x))) \equiv (\neg \forall x A(x) \supset (\forall x A(x) \supset \forall x \neg A(x))) \tag{8.42}$$

$$(\exists x \circ A(x) \supset \circ (\exists x A(x))) \equiv (\neg \exists x A(x) \supset ((\exists x A(x) \supset \forall x A(x)) \wedge \forall x \neg A(x))) \tag{8.43}$$

Proof. See Appendix. □

Remark 16. We have thus seen that propagation results hold not only for the propositional calculus, but also for the predicate calculus.

8.5 Concluding Remarks

As we have show, the system **Bk** has the following two interesting properties:

- Two kinds of propagation of consistency can be stated without the presence of $\circ A$.
- In view of the characterization, we obtain a nice reading of the consistency operator which reflects the intuition of da Costa quite well.

Since these results also hold in the extensions of the system **Bk**, which include most of the systems treated in Carnielli et al. (2007) and Carnielli and Marcos (2002), we may say that our results are general enough in this sense. However, there is still scope for further research extending these results. Indeed, in view of the work of Loparić and da Costa (cf. Loparić and da Costa 1984) on a system which is not only paraconsistent but also paracomplete, we may consider a subsystem of **Bk** in which the law of the excluded middle with respect to paraconsistent negation holds only with restriction in the following form:

$$\circ A \supset (A \vee \neg A).$$

With another modification of the axiomatization of **Bk**, we obtain a system referred to as **BS** in Omori and Waragai (2011). In this system **BS**, one can generalize the results obtained in the present paper both the characterization result and the propagation results. We shall leave the detail for another occasion.

Acknowledgements The authors would like to thank the referees for their helpful comments which improved the paper in many ways. We would also like to thank Dr. Andrew Brooke-Taylor who kindly proofread our final draft and made many helpful suggestions to improve our English. The first author is a JSPS (DC1) research fellow and the present work is partially supported by a Grant-in-Aid for JSPS Fellows.

Appendix

In this appendix, we will provide the readers with full proofs of the theorems which we have omitted in the main body. We begin with the preliminaries for the proofs.

Preliminaries

First, note that several theorems, listed below, can be proved in **Bk** since it contains axioms from (A1) to (A9) together with the only rule of inference (MP):

$$(A \supset (B \supset C)) \supset (B \supset (A \supset C)) \tag{8.44}$$

$$((A \wedge B) \supset C) \supset (A \supset (B \supset C)) \tag{8.45}$$

$$(A \supset (B \supset C)) \supset ((A \wedge B) \supset C) \tag{8.46}$$

$$(C \supset A) \supset ((C \supset B) \supset (C \supset (A \wedge B))) \tag{8.47}$$

$$(A \vee (B \wedge C)) \supset (A \vee B) \tag{8.48}$$

$$(A \vee (B \wedge C)) \supset (A \vee C) \tag{8.49}$$

$$(A \supset (B \supset (C \supset D))) \supset ((C \wedge A) \supset (B \supset D) \tag{8.50}$$

$$(A \supset (B \supset (C \supset D))) \supset ((B \wedge A) \supset (C \supset D) \tag{8.51}$$

$$(A \supset (A \wedge B)) \equiv (A \supset B) \tag{8.52}$$

$$(A \supset B) \supset ((B \supset A) \supset (A \equiv B)) \tag{8.53}$$

$$((A \supset B) \supset B) \equiv (A \vee B) \tag{8.54}$$

$$((A \supset B) \supset A) \equiv A \tag{8.55}$$

$$(A \vee (B \supset C)) \equiv (B \supset (C \vee A)) \tag{8.56}$$

$$(A \wedge B \wedge (C \supset A)) \equiv (A \wedge B) \tag{8.57}$$

$$(A \wedge B \wedge (A \vee C) \wedge D) \equiv (A \wedge D \wedge B) \tag{8.58}$$

$$B \supset (A \equiv (A \wedge B)) \tag{8.59}$$

$$((A \vee B) \supset C) \equiv ((A \supset C) \wedge (B \supset C)) \tag{8.60}$$

$$((C \supset A) \wedge (C \supset B)) \equiv (C \supset (A \wedge B)) \tag{8.61}$$

$$(A \equiv B) \supset ((C \supset A) \equiv (C \supset B)) \tag{8.62}$$

$$(A \equiv B) \supset ((C \wedge A) \equiv (C \wedge B)) \tag{8.63}$$

$$(A \equiv B) \supset ((A \vee C) \equiv (B \vee C)) \tag{8.64}$$

$$(A \equiv B) \supset ((C \equiv D) \supset ((A \vee C \vee E) \equiv (B \vee D \vee E)) \tag{8.65}$$

$$(A \equiv B) \supset ((C \equiv D) \supset ((A \supset C) \equiv (B \supset D))) \tag{8.66}$$

$$(A \equiv B) \supset ((C \equiv D) \supset ((A \wedge C) \equiv (B \wedge D))) \tag{8.67}$$

$$(A \supset (B \equiv C)) \supset ((D \supset (E \equiv F)) \supset ((A \wedge D) \supset ((B \vee E) \equiv (C \vee F)))) \tag{8.68}$$

$$((A \wedge B) \supset C) \equiv (B \supset (A \supset C)) \tag{8.69}$$

$$(((A \vee B) \wedge C) \supset D) \equiv (C \supset ((A \supset D) \wedge (B \supset D))) \tag{8.70}$$

$$((A \wedge B) \supset (C \wedge D)) \equiv (B \supset ((A \supset C) \wedge (A \supset D))) \tag{8.71}$$

$$((A \wedge B) \vee (C \wedge D) \vee (E \supset (B \wedge D))) \equiv (E \supset ((B \vee (C \wedge D)) \wedge (D \vee (A \wedge B)))) \tag{8.72}$$

Also note that the following rules of inference can be derived in **Bk**:

$$\frac{A \supset B \quad B \supset C}{A \supset C} \tag{Syll.}$$

$$\frac{A \equiv B \quad B \equiv C}{A \equiv C} \tag{EqSyll.}$$

$$\frac{A \equiv (B \supset C) \quad C \equiv D}{A \equiv (B \supset D)} \tag{8.73}$$

$$\frac{A \supset (B \equiv (C \supset D)) \quad A \supset (D \equiv E)}{A \supset (B \equiv (C \supset E))} \tag{8.74}$$

$$\frac{A \equiv (B \wedge C) \quad B \equiv (D \supset E) \quad C \equiv (D \supset F)}{A \equiv (D \supset (E \wedge F))} \tag{8.75}$$

Second, the following theorems which contain negation will be useful.

$$(A \supset (A \wedge \neg A)) \equiv \neg A \tag{8.76}$$

$$((A \supset B) \supset \neg A) \equiv (A \supset (B \supset \neg A)) \tag{8.77}$$

$$((A \supset B) \supset \neg B) \equiv \neg B \tag{8.78}$$

Finally, as for theorems in the predicate calculus **Bk***, note that the following formulas are provable.

$$\neg^* \forall x A(x) \supset \exists x \neg^* A(x) \tag{8.79}$$

$$\neg^* \forall x A(x) \equiv \exists x \neg^* A(x) \tag{8.80}$$

$$\neg^* \exists x A(x) \supset \forall x \neg^* A(x) \tag{8.81}$$

$$\neg^* \exists x A(x) \equiv \forall x \neg^* A(x) \tag{8.82}$$

$$(\forall x A(x) \vee B) \equiv \forall x (A(x) \vee B) \tag{8.83}$$

Here, of course, x is not free in B of (8.83). Also, note that the following rules of inference can be derived in **Bk***:

$$\frac{A(x) \supset B(x)}{\forall x A(x) \supset \forall x B(x)} \tag{8.84}$$

$$\frac{A(x) \supset B(x)}{\exists x A(x) \supset \exists x B(x)} \tag{8.85}$$

$$\frac{A(x) \equiv B(x)}{\exists x A(x) \equiv \exists x B(x)} \tag{8.86}$$

$$\frac{A(x) \equiv (B(x) \wedge C(x))}{\forall x A(x) \equiv (\forall x B(x) \wedge \forall x C(x))} \tag{8.87}$$

$$\frac{A\supset(B(x) \equiv C(x))}{A\supset(\exists x B(x) \equiv \exists x C(x))} \tag{8.88}$$

Here, of course, x is not free in A of (8.88). We shall make use of the above theorems and rules of inference in the following proofs.

Proof of Theorem 1. For (8.1) and (8.2), just apply (8.48) and (8.49) to (k). For (8.3), we obtain $(A \vee \neg A)\supset(\neg^* A\supset\neg A)$ by (CN) and therefore $\neg^* A\supset\neg A$ follows by making use of (A11) and (MP). For (8.4):

1	$\circ A\supset((A \wedge \neg A)\supset\neg^*(A \wedge \neg A))$	[(A11), (8.46), (Syll.)]
2	$\circ A\supset\neg^*(A \wedge \neg A)$	[1, (CN), (Syll.)]
3	$\circ A \equiv \neg^*(A \wedge \neg A)$	[2, (⋆), (8.53), (MP)]

For (8.5), it immediately follows from (8.4) by (CN) and as for (8.6), just note that we have (8.4) together with $\neg^*(A \wedge \neg A) \equiv (\neg A\supset\neg^* A)$ which can be obtained by (CN). For (8.7), note that we obtain $(\neg A\supset\neg^* A) \equiv (\neg A \equiv \neg^* A)$ by (8.3) and (8.59), and combine it with (8.6). For the proof of (8.8), note that we obtain $(\neg A \equiv \neg^* A) \equiv (\neg^*\neg A \equiv A)$ by (CN). For (8.9):

1	$\neg^* A\supset \circ A$	[(8.1), (CN), (MP)]
2	$\neg^* A\supset(\neg A \wedge \circ A)$	[1, (8.3), (8.47), (MP)]
3	$(\neg A \wedge \circ A)\supset(A\supset\neg^* A)$	[(A11), (8.50), (MP)]
4	$(\neg A \wedge \circ A)\supset\neg^* A$	[3, (CN), (Syll.)]
5	$\neg^* A \equiv (\neg A \wedge \circ A)$	[2, 4, (8.53), (MP)]

For (8.10):

1	$\neg^*\neg A\supset A$	[(8.3), (CN), (MP)]
2	$\neg^*\neg A\supset \circ A$	[(8.2), (CN), (MP)]
3	$\neg^*\neg A\supset(A \wedge \circ A)$	[1, 2, (8.47), (MP)]
4	$(A \wedge \circ A)\supset(\neg A\supset\neg^*\neg A)$	[(A11), (8.51), (MP)]
5	$(A \wedge \circ A)\supset\neg^*\neg A$	[4, (CN), (Syll.)]
6	$\neg^*\neg A \equiv (A \wedge \circ A)$	[3, 5, (8.53), (MP)]

This completes the proof.

Proof of Proposition 5. Proof proceeds analogously for all four formulas. We shall therefore only prove two cases, for (8.11) and (8.12).
For (8.11):

1	$\circ A\supset(\circ\neg A \equiv (\neg\neg A\supset\neg^* \neq \neg^*$	[(8.6), (A1), (MP)]
2	$\circ A\supset(\circ\neg A \equiv (\neg\neg A\supset A))$	[1, (8.8), (8.74)]

For (8.12):

1	$\circ(A \land B) \equiv (\neg(A \land B) \supset \neg^*(A \land B))$	[(8.6)]
2	$\neg^*(A \land B) \equiv (\neg^* A \lor \neg^* B)$	[(CN)]
3	$\circ(A \land B) \equiv (\neg(A \land B) \supset (\neg^* A \lor \neg^* B))$	[1, 2, (8.73)]
4	$(\circ A \land \circ B) \supset (\circ(A \land B) \equiv$ $(\neg(A \land B) \supset (\neg^* A \lor \neg^* B)))$	[3, (A1), (MP)]
5	$\circ A \supset (\neg^* A \equiv \neg A)$	[(8.7), (A4), (MP)]
6	$\circ B \supset (\neg^* B \equiv \neg B)$	[(8.7), (A4), (MP)]
7	$(\circ A \land \circ B^{)} \supset ((\neg^* A \lor \neg^* B) \equiv (\neg A \lor \neg B))$	[5, 6, (8.68), (MP)]
8	$(\circ A \land \circ B) \supset (\circ(A \land B) \equiv (\neg(A \land B) \supset (\neg A \lor \neg B)))$	[4, 7, (8.74)]

This completes the proof.

Proof of Theorem 5. For (8.19):

1	$(A \supset (\neg\neg A \supset A)) \supset ((\circ A \supset (\neg\neg A \supset A)) \supset$ $((\circ A \lor A) \supset (\neg\neg A \supset A)))$	[(A8), (8.44),(MP)]
2	$(\circ A \supset (\neg\neg A \supset A)) \supset ((\circ A \lor A) \supset (\neg\neg A \supset A))$	[1, (A1), (MP)]
3	$(\circ A \lor A) \supset ((\circ A \supset (\neg\neg A \supset A)) \supset (\neg\neg A \supset A))$	[2, (8.44), (MP)]
4	$(\circ A \supset (\neg\neg A \supset A)) \supset (\neg\neg A \supset A)$	[3, (8.1), (MP)]
5	$(\circ A \supset (\neg\neg A \supset A)) \equiv (\neg\neg A \supset A)$	[4, (A1), (8.53), (MP)]

For (8.20): We abbreviate the formula $\neg(A \land B) \supset (\neg A \lor \neg B)$ to X.

1	$(\neg A \supset (\circ B \supset X)) \supset ((\circ A \supset (\circ B \supset X)) \supset$ $((\circ A \lor \neg A) \supset (\circ B \supset X)))$	[(A8), (8.44), (MP)]
2	$(\neg A \supset (\circ B \supset X))$	[(A6), (A1), (8.44), (MP)]
3	$(\circ A \supset (\circ B \supset X)) \supset ((\circ A \lor \neg A) \supset (\circ B \supset X))$	[1, 2, (MP)]
4	$(\circ A \lor \neg A) \supset ((\circ A \supset (\circ B \supset X)) \supset (\circ B \supset X))$	[3, (8.44), (MP)]
5	$(\circ A \supset (\circ B \supset X)) \supset (\circ B \supset X)$	[4, (8.2), (MP)]
6	$\circ B \supset ((\circ A \supset (\circ B \supset X)) \supset X)$	[5, (8.44), (MP)]
7	$\neg B \supset ((\circ A \supset (\circ B \supset X)) \supset X)$	[(A7), (A1), (8.44), (MP)]
8	$(\circ B \lor \neg B) \supset ((\circ A \supset (\circ B \supset X)) \supset X)$	[(A8), 6, 7, (MP)]
9	$(\circ A \supset (\circ B \supset X)) \supset X$	[8, (8.2), (MP)]
10	$((\circ A \land \circ B) \supset X) \supset X$	[(8.45), 9, (Syll.)]
11	$((\circ A \land \circ B) \supset X) \equiv X$	[10, (A1), (8.53), (MP)]

For (8.21):

1 $((\circ A\wedge\circ B)\supset(\neg(A\vee B)\supset(\neg A\wedge\neg B)))\equiv$
$(\neg^*(\circ A\wedge\circ B)\vee(\neg(A\vee B)\supset(\neg A\wedge\neg B)))$ [(CN)]

2 $(\neg^*(\circ A\wedge\circ B)\vee(\neg(A\vee B)\supset(\neg A\wedge\neg B)))\equiv$
$((\neg^*(\circ A)\vee\neg^*(\circ B))\vee(\neg(A\vee B)\supset B)\supset$
$(\neg A\wedge\neg B)))$ [(8.64), (CN), (MP)]

3 $(\neg^*(\circ A)\vee\neg^*(\circ B))\vee(\neg(A\vee B)\supset(\neg A\wedge\neg B)))\equiv$
$(((A\wedge\neg A)\vee(B\wedge\neg B))\vee(\neg(A\vee B)\supset$
$(\neg A\wedge\neg B)))$ [(8.65), (8.5), (MP)]

4 $(((A\wedge\neg A)\vee(B\wedge\neg B))\vee(\neg(A\vee B)\supset(\neg A\wedge$
$\neg B)))\equiv$
$(\neg(A\vee B)\supset((\neg A\vee(B\wedge\neg B))\wedge$ [(8.72)]
$(\neg B\vee(A\wedge\neg A))))$

5 $((\circ A\wedge\circ B)\supset(\neg(A\vee B)\supset(\neg A\wedge\neg B)))\equiv$
$(\neg(A\vee B)\supset((\neg A\vee(B\wedge\neg B))\wedge$
$(\neg B\vee(A\wedge\neg A))))$ [1–4, (EqSyll.)]

For (8.22): Proof is analogous to the proof for (8.21); indeed, we just need to replace $\neg(A\vee B), (\neg A\wedge\neg B)$ with $\neg(A\supset B), (A\wedge\neg B)$ respectively.
This completes the proof.

Proof of Proposition 6. For (8.27):

1 $(\circ A\supset\circ(A\wedge B))\equiv(\neg^*\circ(A\wedge B)\supset\neg^*(\circ A))$ [(CN)]

2 $(\neg^*\circ(A\wedge B)\supset\neg^*(\circ A))\equiv$
$(((A\wedge B)\wedge\neg(A\wedge B))\supset(A\wedge\neg A))$ [(8.66), (8.5), (MP)]

3 $(((A\wedge B)\wedge\neg(A\wedge B))\supset(A\wedge\neg A))\equiv$
$((B\wedge\neg(A\wedge B))\supset(A\supset(A\wedge\neg A)))$ [(8.69)]

4 $((B\wedge\neg(A\wedge B))\supset(A\supset(A\wedge\neg A)))\equiv$
$((B\wedge\neg(A\wedge B))\supset\neg A)$ [(8.62), (8.76), (MP)]

5 $((B\wedge\neg(A\wedge B))\supset\neg A)\equiv$ [(8.69)]
$(\neg(A\wedge B)\supset(B\supset\neg A))$

6 $(\circ A\supset\circ(A\wedge B))\equiv(\neg(A\wedge B)\supset(B\supset\neg A))$ [1–5, (EqSyll.)]

For (8.28):

1 $(\circ A\supset\circ(A\vee B))\equiv(\neg^*\circ(A\vee B)\supset\neg^*(\circ A))$ [(CN)]

2 $(\neg^*\circ(A\vee B)\supset\neg^*(\circ A))\equiv$
$(((A\vee B)\wedge\neg(A\vee B))\supset(A\wedge\neg A))$ [(8.66), (8.5), (MP)]

3 $(((A\vee B)\wedge\neg(A\vee B))\supset(A\wedge\neg A))\equiv$
$((\neg(A\vee B)\supset(A\supset(A\wedge\neg A)))\wedge$
$(\neg(A\vee B)\supset(B\supset(A\wedge\neg A)))$ [(8.70)]

4 $((\neg(A\vee B)\supset(A\supset(A\wedge\neg A)))\equiv((\neg(A\vee$ [(8.62), (8.76), (MP)]
$B)\supset\neg A)$

5 $(\neg(A\vee B)\supset(B\supset(A\wedge\neg A)))\equiv$
$(\neg(A\vee B)\supset((B\supset A)\wedge(B\supset\neg A)))$ [(8.62), (8.61), (MP)]

6	$(((A \vee B) \wedge \neg(A \vee B)) \supset (A \wedge \neg A)) \equiv (\neg(A \vee B) \supset (\neg A \wedge (B \supset A) \wedge (B \supset \neg A)))$	[3, 4, 5, (8.75)]
7	$(\neg(A \vee B) \supset (\neg A \wedge (B \supset A) \wedge (B \supset \neg A))) \equiv (\neg(A \vee B) \supset (\neg A \wedge (B \supset A)))$	[(8.62), (8.57), (MP)]
8	$(\circ A \supset \circ (A \vee B)) \equiv (\neg(A \vee B) \supset (\neg A \wedge (B \supset A)))$	[1, 2, 6, 7, (EqSyll.)]

For (8.29):

1	$(\circ A \supset \circ (A \supset B)) \equiv (\neg^* \circ (A \supset B) \supset \neg^*(\circ A))$	[(CN)]
2	$(\neg^* \circ (A \supset B) \supset \neg^*(\circ A)) \equiv (((A \supset B) \wedge \neg(A \supset B)) \supset (A \wedge \neg A))$	[(8.66), (8.5),(MP)]
3	$(((A \supset B) \wedge \neg(A \supset B)) \supset (A \wedge \neg A)) \equiv (((\neg(A \supset B)) \supset ((A \supset B) \supset A)) \wedge ((\neg(A \supset B)) \supset ((A \supset B) \supset \neg A)))$	[(8.71)]
4	$((\neg(A \supset B)) \supset ((A \supset B) \supset A)) \equiv (\neg(A \supset B) \supset A)$	[(8.62), (8.55), (MP)]
5	$((\neg(A \supset B)) \supset ((A \supset B) \supset \neg A)) \equiv (\neg(A \supset B) \supset (A \supset (B \supset \neg A)))$	[(8.62), (8.77), (MP)]
6	$(((A \supset B) \wedge \neg(A \supset B)) \supset (A \wedge \neg A)) \equiv (\neg(A \supset B) \supset (A \wedge (B \supset \neg A)))$	[3, 4, 5, (8.75)]
7	$(\circ A \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset (A \wedge (B \supset \neg A)))$	[1, 2, 6, (EqSyll.)]

For (8.30):

1	$(\circ B \supset \circ (A \supset B)) \equiv (\neg^* \circ (A \supset B) \supset \neg^*(\circ B))$	[(CN)]
2	$(\neg^* \circ (A \supset B) \supset \neg^*(\circ B)) \equiv (((A \supset B) \wedge \neg(A \supset B)) \supset (B \wedge \neg B))$	[(8.66), (8.5),(MP)]
3	$(((A \supset B) \wedge \neg(A \supset B)) \supset (B \wedge \neg B)) \equiv (((\neg(A \supset B)) \supset ((A \supset B) \supset B)) \wedge ((\neg(A \supset B)) \supset ((A \supset B) \supset \neg B)))$	[(8.71)]
4	$((\neg(A \supset B)) \supset ((A \supset B) \supset B)) \equiv (\neg(A \supset B) \supset (A \vee B))$	[(8.62), (8.54), (MP)]
5	$((\neg(A \supset B)) \supset ((A \supset B) \supset \neg B)) \equiv (\neg(A \supset B) \supset \neg B)$	[(8.62), (8.78), (MP)]
6	$(((A \supset B) \wedge \neg(A \supset B)) \supset (B \wedge \neg B)) \equiv (\neg(A \supset B) \supset ((A \vee B) \wedge \neg B))$	[3, 4, 5, (8.75)]
7	$(\circ B \supset \circ (A \supset B)) \equiv (\neg(A \supset B) \supset ((A \vee B) \wedge \neg B))$	[1, 2, 6, (EqSyll.)]

This completes the proof.

Proof of Theorem 7. For (8.31):

1 $((\circ A \vee \circ B) \supset \circ (A \wedge B)) \equiv$
$((\circ A \supset \circ (A \wedge B)) \wedge (\circ B \supset \circ (A \wedge B)))$ [(8.60)]

2 $((\circ A \supset \circ (A \wedge B)) \wedge (\circ B \supset \circ (A \wedge B))) \equiv$
$((\neg(A \wedge B) \supset (B \supset \neg A)) \wedge$
$(\neg(A \wedge B) \supset (A \supset \neg B)))$ [(8.67), (8.27), (MP)]

3 $((\neg(A \wedge B) \supset (B \supset \neg A)) \wedge$
$(\neg(A \wedge B) \supset (A \supset \neg B))) \equiv$
$(\neg(A \wedge B) \supset ((A \supset \neg B) \wedge (B \supset \neg A)))$ [(8.61)]

4 $((\circ A \vee \circ B) \supset \circ (A \wedge B)) \equiv$
$(\neg(A \wedge B) \supset ((A \supset \neg B) \wedge (B \supset \neg A)))$ [1–3, (EqSyll.)]

For (8.32):

1 $((\circ A \vee \circ B) \supset \circ (A \vee B)) \equiv$
$((\circ A \supset \circ (A \vee B)) \wedge (\circ B \supset \circ (A \vee B)))$ [(8.60)]

2 $((\circ A \supset \circ (A \vee B)) \wedge (\circ B \supset \circ (A \vee B))) \equiv$
$((\neg(A \vee B) \supset (\neg A \wedge (B \supset A))) \wedge$
$(\neg(A \vee B) \supset (\neg B \wedge (A \supset B))))$ [(8.67), (8.28), (MP)]

3 $((\neg(A \vee B) \supset (\neg A \wedge (B \supset A))) \wedge$
$(\neg(A \vee B) \supset (\neg B \wedge (A \supset B)))) \equiv$
$(\neg(A \vee B) \supset ((\neg A \wedge \neg B) \wedge (A \equiv B)))$ [(8.61)]

4 $((\circ A \vee \circ B) \supset \circ (A \vee B)) \equiv$
$(\neg(A \vee B) \supset ((\neg A \wedge \neg B) \wedge (A \equiv B)))$ [1–3, (EqSyll.)]

For (8.33):

1 $((\circ A \vee \circ B) \supset \circ (A \supset B)) \equiv$
$((\circ A \supset \circ (A \supset B)) \wedge (\circ B \supset \circ (A \supset B)))$ [(8.60)]

2 $((\circ A \supset \circ (A \supset B)) \wedge (\circ B \supset \circ (A \supset B))) \equiv$
$((\neg(A \supset B) \supset (A \wedge (B \supset \neg A))) \wedge$
$(\neg(A \supset B) \supset ((A \vee B) \wedge \neg B)))$ [(8.67), (8.29), (8.30), (MP)]

3 $((\neg(A \supset B) \supset (A \wedge (B \supset \neg A))) \wedge$
$(\neg(A \supset B) \supset ((A \vee B) \wedge \neg B))) \equiv$
$(\neg(A \supset B) \supset ((A \wedge (B \supset \neg A)) \wedge ((A \vee B) \wedge \neg B)))$ [(8.61)]

4 $(\neg(A \supset B) \supset ((A \wedge (B \supset \neg A)) \wedge ((A \vee B) \wedge \neg B))) \equiv$ [(8.58), (8.62), (MP)]
$(\neg(A \supset B) \supset (A \wedge \neg B \wedge (B \supset \neg A)))$

5 $((\circ A \vee \circ B) \supset \circ (A \supset B)) \equiv$
$(\neg(A \supset B) \supset ((A \wedge \neg B) \wedge (B \supset \neg A)))$ [1–5, (EqSyll.)]

This completes the proof.

Proof of Lemma 5. Since the proofs for two formulas (8.34) and (8.35) are analogous, we only prove the former whose proof runs as follows:

1 $\circ\forall x A(x) \equiv (\neg\forall x A(x)\supset\neg^*\forall x A(x))$ [(8.6)]

2 $\circ\forall x A(x) \equiv (\neg\forall x A(x)\supset\exists x\neg^* A(x))$ [1, (8.80), (8.73)]

3 $\forall x \circ (A(x))\supset(\circ(\forall x A(x)) \equiv$
$(\neg\forall x A(x)\supset\exists x\neg^* A(x)))$ [2, (A1), (MP)]

4 $\forall x \circ (A(x))\supset(\neg^* A(x) \equiv \neg A(x))$ [(A12), (8.7), (Syll.)]

5 $\forall x \circ (A(x))\supset(\exists x\neg^* A(x) \equiv \exists x\neg A(x))$ [4, (8.88)]

6 $\forall x \circ (A(x))\supset(\circ(\forall x A(x)) \equiv$
$(\neg\forall x A(x)\supset\exists x\neg A(x)))$ [3, 5, (8.74)]

This completes the proof.

Proof of Theorem 9. For (8.38):

1 $\exists x\neg^* \circ (A(x))\supset\exists x\neg A(x)$ [(8.2), (CN), (8.85)]

2 $\neg^*\forall x \circ ((A(x)))\supset\exists x\neg A(x)$ [1, (8.79), (Syll.)]

3 $\neg^*\forall x \circ ((A(x)))\supset(\neg\forall x A(x)\supset\exists x\neg A(x))$ [2, (A1), (MP)]

4 $(\forall x \circ A(x)\supset(\neg\forall x A(x)\supset\exists x\neg A(x)))\supset$
$(\forall x \circ ((A(x)) \vee \neg^*\forall x \circ (A(x)))\supset$
$(\neg\forall x A(x)\supset\exists x\neg A(x)))$ [(A8), (8.44), 3, (MP)]

5 $(\forall x \circ A(x)\supset(\neg\forall x A(x)\supset\exists x\neg A(x)))\supset$
$(\neg\forall x A(x)\supset\exists x\neg A(x))$ [4, (8.44), (CN), (MP)]

6 $(\forall x \circ A(x)\supset(\neg\forall x A(x)\supset\exists x\neg A(x))) \equiv$
$(\neg\forall x A(x)\supset\exists x\neg A(x))$ [5, (A1), (8.53), (MP)]

For (8.39):

1 $(\forall x \circ A(x)\supset(\neg\exists x A(x)\supset\forall x\neg A(x))) \equiv$
$(\neg^*\forall x \circ (A(x)) \vee$ [(CN)]
$(\neg\exists x A(x)\supset\forall x\neg A(x)))$

2 $(\neg^*\forall x \circ (A(x)) \vee (\neg\exists x A(x)\supset\forall x\neg A(x))) \equiv$
$(\exists x\neg^* \circ (A(x)) \vee$ [(8.64), (8.80), (MP)]
$(\neg\exists x A(x)\supset\forall x\neg A(x)))$

3 $\exists x\neg^* \circ (A(x)) \equiv \exists x(A(x) \wedge \neg A(x))$ [(8.5), (8.86)]

4 $(\exists x\neg^* \circ (A(x)) \vee (\neg\exists x A(x)\supset\forall x\neg A(x))) \equiv$
$(\exists x(A(x) \wedge \neg A(x)) \vee$
$(\neg\exists x A(x)\supset\forall x\neg A(x)))$ [(8.64), 3, (MP)]

5 $(\exists x(A(x) \wedge \neg A(x)) \vee$
$(\neg\exists x A(x)\supset\forall x\neg A(x))) \equiv$
$(\neg\exists x A(x)\supset$ [(8.56)]
$(\forall x\neg A(x) \vee \exists x(A(x) \wedge \neg A(x))))$

6 $(\neg\exists x A(x)\supset(\forall x\neg A(x) \vee \exists x(A(x) \wedge \neg A(x)))) \equiv (\neg\exists x A(x)\supset \forall x(\neg A(x) \vee \exists x(A(x) \wedge \neg A(x))))$ [(8.83), (8.62), (MP)]

7 $(\forall x \circ A(x)\supset(\neg\exists x A(x)\supset\forall x\neg A(x))) \equiv (\neg\exists x A(x)\supset \forall x(\neg A(x) \vee \exists x(A(x) \wedge \neg A(x))))$ [1, 2, 4, 5, 6, (EqSyll.)]

This completes the proof.

Proof of Theorem 11. For (8.42):

1 $(\exists x \circ A(x) \supset \circ\forall x A(x)) \equiv (\neg^* \circ \forall x A(x) \supset \neg^*\exists x \circ A(x))$ [(CN)]

2 $(\neg^*\forall x \circ (A(x))\supset\neg^* \circ (\exists x(A(x)))) \equiv ((\forall x A(x)\wedge\neg\forall x A(x))\supset\forall x\neg^*\circ(A(x)))$ [(8.66), (8.5), (8.82), (MP)]

3 $\forall x\neg^* \circ (A(x)) \equiv (\forall x A(x) \wedge \forall x\neg A(x))$ [(8.5), (8.87)]

4 $((\forall x A(x) \wedge \neg\forall x A(x))\supset\forall x\neg^* \circ (A(x))) \equiv ((\forall x A(x) \wedge \neg\forall x A(x))\supset (\forall x A(x) \wedge \forall x\neg A(x)))$ [3, (8.62), (MP)]

5 $((\forall x A(x) \wedge \neg\forall x A(x))\supset(\forall x A(x) \wedge \forall x\neg A(x))) \equiv (\neg\forall x A(x)\supset (\forall x A(x)\supset(\forall x A(x) \wedge \forall x\neg A(x))))$ [(8.45), (8.44), (Syll.)]

6 $(\neg\forall x A(x)\supset(\forall x A(x)\supset(\forall x A(x) \wedge \forall x\neg A(x)))) \equiv (\neg\forall x A(x)\supset(\forall x A(x)\supset\forall x\neg A(x)))$ [(8.62), (8.52), (MP)]

7 $\circ(\exists x(A(x))\supset \circ (\forall x A(x))) \equiv (\neg\forall x A(x)\supset(\forall x A(x)\supset\forall x\neg A(x)))$ [1, 2, 4–6, (EqSyll.)]

For (8.43):

1 $(\exists x \circ A(x) \supset \circ\exists x A(x)) \equiv (\neg^* \circ \exists x A(x) \supset \neg^*\exists x \circ A(x))$ [(CN)]

2 $(\neg^* \circ \exists x A(x) \supset \neg^*\exists x \circ A(x)) \equiv ((\exists x A(x) \wedge \neg\exists x A(x)) \supset \forall x\neg^* \circ A(x))$ [(8.66), (8.5), (8.82), (MP)]

3 $\forall x\neg^* \circ (A(x)) \equiv (\forall x A(x) \wedge \forall x\neg A(x))$ [(8.5), (8.87)]

4 $((\exists x A(x) \wedge \neg\exists x A(x))\supset\forall x\neg^* \circ (A(x))) \equiv ((\exists x A(x) \wedge \neg\exists x A(x))\supset (\forall x A(x) \wedge \forall x\neg A(x)))$ [3, (8.62), (MP)]

5 $((\exists x A(x)\wedge\neg\exists x A(x))\supset(\forall x A(x)\wedge \forall x\neg A(x)))\equiv (\neg\exists x A(x)\supset((\exists x A(x)\supset\forall x A(x))\wedge (\exists x A(x)\supset\forall x\neg A(x))))$ [(8.71)]

6 $(\neg^*\exists x A(x)\supset\forall x\neg A(x))\supset((\exists x A(x)\supset\forall x\neg A(x))\supset$
$((\exists x A(x)\vee\neg^*\exists x A(x))\supset\forall x\neg A(x)))$ [(A8), (8.44), (MP)]

7 $\forall x\neg^* A(x)\supset\forall x\neg A(x)$ [(8.3), (8.84)]

8 $\neg^*\exists x A(x)\supset\forall x\neg A(x)$ [(8.79), 7, (Syll.)]

9 $(\exists x A(x)\supset\forall x\neg A(x))\supset$
$((\exists x A(x)\vee\neg^*\exists x A(x))\supset\forall x\neg A(x))$ [6, 8, (MP)]

10 $(\exists x A(x)\supset\forall x\neg A(x))\supset\forall x\neg A(x)$ [9, (8.44), (CN), (MP)]

11 $(\exists x A(x)\supset\forall x\neg A(x))\equiv\forall x\neg A(x)$ [10, (A1), (8.53), (MP)]

12 $((\exists x A(x)\supset\forall x A(x))\wedge$
$(\exists x A(x)\supset\forall x\neg A(x)))\equiv$
$((\exists x A(x)\supset\forall x A(x))\wedge\forall x\neg A(x))$ [11, (8.63), (MP)]

13 $(\neg\exists x A(x)\supset((\exists x A(x)\supset\forall x A(x))\wedge$
$(\exists x A(x)\supset\forall x\neg A(x))))\equiv(\neg\exists x A(x)\supset$
$((\exists x A(x)\supset\forall x A(x))\wedge\forall x\neg A(x)))$ [12, (8.62), (MP)]

14 $(\exists x\circ A(x)\supset\circ(\exists x A(x)))\equiv$
$(\neg\exists x A(x)\supset((\exists x A(x)\supset\forall x A(x))\wedge$
$\forall x\neg A(x)))$ [1, 2, 4, 5, 13, (EqSyll.)]

This completes the proof.

References

Avron, A. (2009). Modular semantics for some basic logics of formal inconsistency. In W. A. Carnielli, M. E. Coniglio, & I. M. L. D' Ottaviano (Eds.), *The many sides of logic* (Studies in logic, Vol. 21, pp. 15–26). London: College Publications.

Avron, A., & Zamansky, A. (2007). Many-valued non-deterministic semantics for first-order logics of formal (in)consistency. In S. Aguzzoli, A. Ciabattoni, B. Gerla, C. Manara, & V. Marra (Eds.), *Algebraic and proof-theoretic aspects of non-classical logics*. (LNAI, Vol. 4460, pp. 1–24). Berlin/New York: Springer.

Batens, D. (1980). A completeness-proof method for extensions of the implicational fragment of the propositional calculus. *Notre Dame Journal of Formal Logic, 21*(3), 509–517.

Béziau, J. Y. (1990). Loiques construites suivant les méthodes de da Costa. *Logique et Analyse, 131–132*, 259–272.

Béziau, J. Y. (1993). Nouveaux resultats et nouveau regard sur la logique paraconsistente C1. *Logique et Analyse, 141–142*, 45–58.

Carnielli, W. A., & Marcos, J. (2002). A taxonomy of C-systems. In W. A. Carnielli, M. E. Coniglio, & I. M. L. d'Ottaviano (Eds.), *Paraconsistency: The logical way to the inconsistent* (pp. 1–94). New York: Marcel Dekker.

Carnielli, W. A., Coniglio, M. E., & Marcos, J. (2007). Logics of formal inconsistency. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 14, pp. 1–93). Dordrecht/London: Springer.

da Costa, N. C. A. (1974). On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic, 15*(4), 497–510.

Guillaume, M. (2007). da Costa 1964 logical seminar: Revisited memories. In J. Y. Béziau, W. A. Carnielli, & D. Gabbay (Eds.), *Handbook of paraconsistency* (pp. 3–62). London: College Publications.

Loparić, A. (1977). Une étude sémantique de quelques calculs propositionnels. *Comptes Rendus de l'Academie de Sciences de Paris, 284*, 835–838.

Loparić, A. (1986). A semantical study of some propositional calculi. *Journal of Non-classical Logic, 3*, 73–95.

Loparić, A., & da Costa, N. C. A. (1984). Paraconsistency, paracompleteness, and valuations. *Logique et Analyse, 106*, 119–131.

Omori, H., & Waragai, T. (2011). Some observations on the systems **LFI1** and **LFI1***. In *DEXA Workshops 2011*, Toulouse (pp. 320–324).

Waragai, T., & Omori, H. (2010). Some new results on PCL1 and its related systems. *Logic and Logical Philosophy, 19*, 129–158.

Waragai, T., & Shidori, T. (2007). A system of paraconsistent logic that has the notion of "behaving classically" in terms of the law of double negation and its relation to s5. In J. Y. Béziau, W. A. Carnielli, & D. Gabbay (Eds.), *Handbook of paraconsistency* (pp. 177–187). London: College Publications.

Chapter 9
Degrees of Validity and the Logical Paradoxes

Francesco Orilia

9.1 Introduction

We traditionally accept a sharp distinction between deductive and inductive arguments. The former are taken to be undefeasible and thus we accept a principle about deductions, *PRD*, which, roughly, goes as follows: given a deductively valid argument with premises you believe, you should also believe the conclusion of the argument. In contrast, inductive arguments are defeasible and consequently we acknowledge that their validity admits of degrees: strong ("highly valid") arguments can be "blocked" by equally strong or stronger ("equally valid" or "more valid") arguments leading to the opposite conclusion. Accordingly, we accept a principle about inductive reasoning, *PRI*, along these lines: if it is important to make up your mind as to whether C or $\sim C$, given an inductively valid argument A leading to C from premises you believe, then you should believe C, unless you know of another argument A' leading to $\sim C$ from premises you believe, such that A' is at least as valid as A.

Now, logical paradoxes such as the Liar, Russell's or Curry's cast doubts on PRD. For, at least prima facie, they are deductively valid arguments, but they can lead to any conclusion we please, either directly (as in Curry's case) or via *Ex Falso Quodlibet*. Yet, in spite of them, we do not believe, e.g., that the moon is made of blue cheese. Traditional reactions to this problem question either grammar (e.g., by invoking type-theoretical distinctions) or logic (by regarding as not really deductive some inference rules that are traditionally taken to be deductive) so as to claim that the paradoxes fail to be deductively valid after all.

F. Orilia (✉)
Department of Humanities, Section of Philosophy and Human Sciences,
University of Macerata, Macerata, Italy
e-mail: orilia@unimc.it

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic,
Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_9,
© Springer Science+Business Media Dordrecht 2014

Here I shall explore a different approach, based on the intuition that deductive arguments can be treated on analogy with inductive arguments, for in their case as well we can have two valid arguments that lead to opposite conclusions. For instance, Curry's paradox can be used to show that the moon is made of blue cheese, but also to show that it is not so made. My proposal then will be to view deductive arguments as defeasible and accordingly I shall suggest an extension to them of the notion of degrees of validity. In the light of this, the principle PRD will dropped and PRI will be generalized so as to cover both deductive and inductive arguments in a single principle of reasoning, to be called *PR*.

9.2 Deductive Arguments

We can distinguish two fundamental rational reasons for adding, on the basis of arguments, a new belief N to the set S of one's beliefs. One consists in having found a deductively valid argument (a deduction) that leads to N from members of S. The other consists in having found an inductively valid argument (an induction) that leads to N from members of S.

I shall focus on deductions for the time being. Deductions are taken to be *undefeasible*, because they are taken to be both *truth-preserving* and *monotonic*. That is, if the premises are true, it is necessary that the conclusion is also true, or more generally, the conclusion must have the same degree of credibility of the premises. Moreover, if a deductive argument is valid, it remains valid, even if we add new premises. Now, obviously, deductions are constructed in a stepwise fashion, on the basis of inferential deductive rules such as (to take peculiarly uncontroversial examples) conjunction elimination or existential generalization. It is then natural to say that these rules themselves are undefeasible and in particular truth-preserving. That is, in allowing us to move from one step to another in constructing a deduction, they guarantee that we do not add a false proposition to preceding true propositions. This is, we may say, the standard view about deductions and deductive rules or, more generally, laws.[1]

For reasons that will be clear below, it is important to record a crucial point about deductions. As the use of many examples in introductory logic books testifies, there are *paradigmatic* cases of deductions, i.e., arguments for which it seems particularly self-evident that they are undefeasible. Here is one:

Example 1. (a) If Peter has still a fever and he is coughing a lot, then he might have pneumonia; (b) if Peter might have pneumonia, then he should take a chest X-ray; (c) Peter has still a fever; (d) Peter is coughing a lot. Therefore, (e) Peter should take a chest X-ray.

[1] I use "law" to also cover (alleged) logical truths such as the law of excluded middle (LEM) and the principle of non-contradiction (PNC); it is normally assumed that for each of them there is a corresponding deductive rule that says that the law in question can be freely introduced as a step in a deductive argument.

Clearly, if the premises are true, so must the conclusion be, and we must accept this conclusion, even if we add a further premise, e.g., that Peter may simply have a bronchitis.

Patently, there are deductive laws that we can *abstract* from paradigmatic deductions (see Arnold and Shapiro 2007). For example, we can abstract from (an obvious reconstruction of) Example 1 deductive inferential rules such as conjunction introduction and modus ponens (MP). The rules that are so identified can be called not only deductive, but, more specifically, *paradigmatically* deductive, as they can be abstracted from paradigmatic deductions. The paradigmatically deductive rules do not exhaust all the rules that are taken to be deductive. For example, at least classical logicians regard the rule that allows one to infer any proposition from a contradiction (Ex Falso Quodlibet; EFQ, in brief) as deductive. However, an argument that directly appeals to such a rule can hardly be regarded as a paradigm of deduction. If we present Example 1 to beginners in logic we immediately gain their assent. But we hardly have the same reaction if we present something like:

Example 2. (a) Jerry is hungry; (b) Jerry is not hungry. Therefore, (c) snow is black.

However, textbooks distinguish between basic and derived deductive rules. Once the basic rules are assumed, we can derive other rules from them in the sense that an argument that appeals to a derived rule can be seen as an abbreviation of an argument that only uses basic rules. For example, as is well known, an argument that uses EFQ, such as Example 2, can be viewed as an abbreviation of an argument that does not use it and that appeals only to basic deductive rules such as disjunction introduction and disjunctive syllogism, or negation elimination. Now, the basic rules typically found in textbooks are arguably paradigmatic deductive rules and since they are appealed to in paradigmatic deductions it is natural to take them as deductive. Moreover, since the derived rules are simply abbreviations of basic rules it is also natural to take them as deductive.

9.3 The Global Deductive System

The deductive laws presented in introductory logic textbooks simply have to do with the standard connectives and quantifiers and thus do not exhaust all the laws that we actually use in our argumentative practices and that can be considered deductive. For example, we certainly also appeal to deductive rules governing modal and temporal notions such as these:

$$\frac{\text{necessarily, } A}{A.}$$

$$\frac{\text{It was the case that } A}{\text{it will be the case that it was the case that } A.}$$

If we consider all the deductive rules that we actually appeal to in our argumentative practices, we get what we could call "Our Global Deductive System", which, let us conveniently assume, can be presented as a system of natural deduction with variables standing for (appropriately formalized) sentences, which in turn are taken to express propositions.

But what are the deductive laws granted by our global deductive system? At least prima facie, our global deductive system incorporates the laws of classical logic (**CL**), for arguably **CL** can be identified via a set of basic rules that are paradigmatically deductive. And in fact these rules have been more or less explicitly taken for granted for centuries (in spite of occasional qualms such as those in Aristotle's famous discussion of future contingents) and only in a relatively recent past, with the birth of non-classical logics, the deductive status of some of them has been seriously questioned. Moreover, as we have noted, our global deductive system must involve other rules in addition to those for standard connectives and quantifiers. Thus we can at least prima facie characterize our global deductive system as the system $\mathbf{CL}^{+}$, where "**CL**" indicates that it incorporates the laws of **CL** and "+" that it involves other laws as well.

Among these additional laws there are some that are of specific concerns to us here. First of all, "disquotational T-rules" governing the use of the notion of truth as applied to sentences:

$$\text{TE.} \qquad \frac{T(\text{`}A\text{'})}{A.}$$

$$\text{TI.} \qquad \frac{A}{T(\text{`}A\text{'}).}$$

Moreover, "lambda rules" that can be viewed, depending on one's tastes (we need not be picky for present purposes), as either rules governing the attribution of complex properties to objects (viewing "$\in$" as standing for predication) or as rules telling us when an object belongs to a certain set (viewing "$\in$" as standing for set membership):

$$\lambda\text{I.} \qquad \frac{A}{a \in [\lambda x\ A(a/x)].}$$

$$\lambda\text{E.} \qquad \frac{a \in [\lambda x\ A(a/x)]}{A.}$$

These rules should be part of $\mathbf{CL}^{+}$, since they seem to be appealed to in ordinary argumentative practices in arguments that look like paradigms of deduction. Consider for example these arguments:

Example 3. (a) All that Anthony says as physician is true; (b) Anthony said as physician "if Peter has still a fever he may have pneumonia" and "if Peter may have pneumonia he should take a chest X-ray"; (c) Peter has still a fever. Hence, (d) Peter should take a chest X-ray.

Example 4. (a) John belongs to the class of unmarried male adults (is unmarried, male and adult). Therefore, (b) John is not married.

Clearly, arguments such as these are paradigmatically valid arguments that appeal to TE and to λE, respectively, and paradigmatically valid arguments that appeal to TI and λI can also be easily presented.

9.4 The Principles of Reasoning

As explained by Harman in Harman (1986), there are *principles of reasoning* that found our being rational. They guide the addition of new beliefs to given ones on the basis of arguments and should not be confused with the inferential laws used for the constructions of such arguments. Given the standard view about deductive laws and deductions, we typically accept the following principle of reasoning concerning deductive arguments:

PRD. If one finds a deductive argument with conclusion C such that one believes its premises, then one should believe C as well, with the proviso that it is important, in view of a given goal, to have an opinion as to whether C is true or not.

The proviso is important. Without it we should retain as belief any conclusion of a deduction that we take to be sound, even if we could not care less about it, thereby going against the principles of reasoning that Harman calls "Clutter avoidance" and "Interest condition" (Harman 1986, p. 55). According to the former, we should not clutter our mind with trivialities and according to the latter, one should add a new proposition N to one's beliefs only if one is interested (given one's goals) in knowing whether N is true or not. Clearly, these principles are very useful, since our reasoning resources are limited and we have bounded storage capacities in our minds.

To see PRD at work, imagine the following story.

The feverish Peter. Peter is ill and Tom has an interest in taking care of him. It so happens that Tom believes the premises (a), (b) and (c) of Example 3 and (more or less explicitly) constructs an argument that, via TE, universal instantiation and MP (inter alia), leads to the conclusion (d) Peter should take a chest X-ray.

Clearly, Tom should believe the conclusion (d) of Example 3 and this is precisely what PRD tells us. On the basis of PRD, we can easily imagine a continuation of the story in which Tom comes to believe (d) and, accordingly, takes Peter to the lab, to have his X-ray.

9.5 The Logical Paradoxes and Explosion

It is well known that, in the light of logical (self-referential) paradoxes such as the Liar, Russell's or Curry's, **CL** plus the above disquotational T-rules and/or the lambda rules generates explosion, i.e., that every proposition can be deductively inferred (from premises that we are forced to believe). We can arrive to explosion in two ways, either by first generating a contradiction, e.g., with versions of the Liar or Russell's paradox, and then applying EFQ; or, without appealing to EFQ, by using a version of Curry's paradox. In either cases, we can either rely on contingent premises that we can hardly reject,[2] or we appeal to no contingent premises at all, as when we show, using Curry's paradox, that both $[\lambda x\ x \in x \longrightarrow A] \in [\lambda x\ x \in x \longrightarrow A]$ and its negation lead to A (see Meyer et al. 1979).

In sum, $\mathbf{CL}^{+}$ is explosive and to the extent that our global deductive system is identified with $\mathbf{CL}^{+}$ we get the conclusion that our global deductive system is explosive, a conclusion which may be called *Tarski's thesis*. For presumably this is what Tarski meant when he claimed, in the light of the Liar, that natural language is inconsistent (cf. Azzouni 2006).

9.6 A Dilemma About Rationality

Now, Tarski's thesis leads us to a dilemma that we can vividly convey by relying on a story such as this.

The starving Tom. Tom is starving and runs into some mushrooms. It is vital for him to know whether they are edible or poisonous. He had once met a logician who had shown to him various ways in which one could construct deductive arguments that could reach any desired conclusion from a contingently true premise or even from no premises at all. He thus considers that he can take advantage of this to get out of his predicament. He then writes on a piece of paper, call it p_1, just one sentence, "if the longest sentence on this piece of paper is true then the mushrooms are edible", and thus comes to believe that the following is true:

(CP) The longest sentence written on p_1 is "if the longest sentence on this piece of paper is true then the mushrooms are edible".

He then reasons as in Curry's paradox to infer that the mushrooms are edible.

Or perhaps he exploits a version of the Liar and then applies EFQ, after having written on another piece of paper, p_2, just one sentence, namely "the sentence written on this piece of paper is false", thereby coming to believe:

[2] For instance, we write just one sentence on the blackboard, "the sentence on the blackboard is false", and then we unfold the Liar reasoning, starting from the premise, which we must believe, that the only sentence on the blackboard is "the sentence on the blackboard is false."

(LP) The only sentence written on p_2 is "the sentence written on this piece of paper is false".

Alternatively, he considers the proposition [$\lambda x\ x \in x \longrightarrow$ the mushrooms are edible] $\in$ [$\lambda x\ x \in x \longrightarrow$ the mushrooms are edible] and exploits a version of Curry's paradox with no contingent premises.

You choose which of these options is preferred by Tom. Whatever the choice, he deductively reaches the conclusion that the mushrooms are poisonous, from contingent premises that he believes (either (CP) or (LP)) or from no premises at all.

Arguments of this kind may be called *paradoxically explosive* arguments. Once Tom has reached his favorite conclusion on the basis of a paradoxically explosive argument, we can claim that, given PRD, (a) it is rational for him to believe that the mushrooms are edible. On the other hand, (b) it is irrational for Tom to believe that the mushrooms are edible. For clearly the mushrooms might very well be poisonous in spite of the deductive skills exhibited by Tom in (re-)constructing one of these arguments. And if Tom comes to believe on these grounds that the mushrooms are edible, he seriously risks his life. Of course, Tom will not eat the mushrooms on these grounds, but the problem remains that we must face a perplexing dilemma about rationality, because there are serious reasons in favor of both (a) and (b).

How to get out of this dilemma? Let us record that an interesting way out must somehow preserve the idea, embodied in PRD, that, to act rationally, one *must* add to one's beliefs the conclusions of "normal" deductive arguments such as those of Examples 1 and 3, if one is aware of them and believes their premises (and the arguments are relevant for one's goals). For example, we should regard as unsatisfactory a way out that consists in simply saying that PRD is not a principle of reasoning, without replacing it with something that tells us that, in the context of the feverish Peter story, Tom should believe the conclusion (d) of Example 3.

9.7 The Traditional Response

There is a "traditional response" to our dilemma, implicit in all the more or less well-known approaches to the paradoxes in current logico-philosophical literature. According to this response, there is no reason to deny that we are rational when we fail to believe the conclusion of a paradoxically explosive argument, because any such argument presupposes that $\mathbf{CL}^+$ is our global deductive system, whereas in fact this is not the case. That is, we should not consider the paradoxically explosive arguments as deductive after all. The idea is that we can be wrong about what our global deductive system really is and, according to the traditional response, this is precisely what happens when we take $\mathbf{CL}^+$ to be our global deductive system. We might well have some prima facie evidence that $\mathbf{CL}^+$ has this status, but, since deductions are undefeasible, in particular truth-preserving, and $\mathbf{CL}^+$ allows for arguments that lead us from true premises to a false conclusion, this evidence must be set aside. $\mathbf{CL}^+$ cannot be a global *deductive* system and thus a fortiori cannot be *our* global deductive system.

This point of view has been explicitly expressed, e.g., by Louise Antony in this quotation:

> I may, perhaps, feel that I have direct intuitions as to the content of the true logical laws, but if I am right on a distinction between possessing a logic, and possessing a belief about a logic, there should be room for divergencies. Certainly, it must be possible . . . for us to have an intuition . . . that some proposition is [logically] true, when it's actually false. Frege, let us not forget, thought that the Axiom of Comprehension was self-evident. (Antony 2004, p. 12)

Now, since we make deductions, there must exist, according to the traditional response, another system that differs from $\mathbf{CL}^+$ in that it does not contain the former's flaws and that can be identified with our deductive system. There are two variants of the traditional response (a "grammatical" and an "inferential" one, we may say): (a) the flaws are in the grammar of $\mathbf{CL}^+$, as it allows for self-reference, and (b) the flaws are in the inferential laws, i.e., some of them must not really be deductive after all.

9.7.1 The Grammatical Version of the Traditional Response

There are two main approaches that follow the variant (a) of the traditional approach: Russell's ramified type theory and Tarski's construction of a hierarchy of languages. We can be very brief here since these options are well known. They both avoid self-reference so as to rule out the very formulation of the paradoxically explosive arguments. Yet, they do not prevent the formulation of "normal" arguments such as those in *The feverish Peter*, to the extent that they do not involve self-reference.

But these approaches are also in clear conflict with our normal way of speaking and formulating arguments in natural language and thus can hardly be utilized to characterize our everyday inferential practises, which exploit natural language in ways that permit self-reference. Hence, these approaches can hardly provide an adequate response to our dilemma.

In fact, Tarski himself thought that his theory was not so much a description of how we speak and reason in everyday life, but rather a prescription for the construction of a rigorous language for mathematics and science. Russell's attitude toward his type theory was, I think, similar, and, in any case, he was not quite happy with it and would have preferred a different approach, one granting the type-freedom he had accepted in his early years (see Landini 1998).

9.7.2 The Inferential Version of the Traditional Response

According to the variant (b) of the traditional response, grammar is fine and the problem lies with the inferential laws of $\mathbf{CL}^+$: some of them may perhaps look deductive, but they are not, since they license arguments that are not truth-preserving. We are thus wrong in thinking that our global deductive system coincides with $\mathbf{CL}^{+}$. For our global deductive system must rule out those laws of $\mathbf{CL}^{+}$ that are not truth-preserving. It their stead it will presumably encompass some weaker, but more reliable, truth-preserving, laws.

For example, in ZF set theory Frege's comprehension axiom (which more or less corresponds to the above lambda rules) is replaced by a weaker version of it and by other axioms for the postulation of some of the sets granted by Frege. In Gupta's and Belnap's theory of truth (see Gupta and Belnap 1993), the disquotational T-rules are replaced by a set of "circular definitions" on the basis of which, roughly speaking, these T-rules hold in a vast number of cases, but not in general. Similarly, in the property theory of Orilia (see Orilia 2000), Gupta's and Belnap's method of circular definitions is used to circumscribe the generality of the lambda rules. Alternatively, there are theories that replace some laws of **CL**. For instance, the truth theory of Kripke (see Kripke 1975) or the property theory of Field (see Field 2004), where LEM is not generally valid. Or the paraconsistent logic of Priest (see Priest 1987), where MP and LNC are rejected.

All the versions of variant (b) of the traditional response have this problem: they deny that certain inferential laws that can be abstracted from arguments that appear to be paradigmatically deductive are really deductive and thus all versions of variant (b) classify as not really deductive arguments of this sort. For instance, according to Priest's approach, in which MP is not deductively valid, the argument of Example 3, which relies on MP, is not deductive. Now, it does not seem compatible with rationality that Tom (in *The feverish Peter*) does not come to believe that Peter needs an X-ray on the basis of the argument of Example 3. But if this argument is not viewed as deductive, PRD does not grant that the argument provides a reason for Tom to believe the conclusion that Peter needs an X-ray. For PRD has to do with deductive arguments and the argument in question, according to Priest, is not deductive.

Obviously, this approach must be complemented with some classificatory method that attributes some sort of weaker validity to those arguments that we classify as non-deductive although they seem deductive. This would allow us to insist that, when one discovers an argument of this kind, whose premises one believes, then one must, to be rational, believe its conclusion, at least in typical circumstances. Priest (1987) has appreciated this point and thus he considers arguments that rely on those rules of **CL** that he rejects (e.g., MP and PNC) as "quasi-valid". These arguments are not valid in an absolute sense, but as long as they don't have to do with an inconsistent domain, we can rely on them in that they will not lead from false premises to true conclusions. This is analogous to what we find in Batens' adaptive logics.[3] In these logics one distinguishes a lower-limit and an upper-limit logic. The former provides a set of rules that hold no matter what (lower-limit logic), and a

[3] See, e.g., Batens (1999) and the "Adaptive Logics homepage" at http://logica.ugent.be/adlog/al.html.

set of rules that hold only in “normal situations” and thus may lead to conclusions that are taken back in the light of “abnormalities”. In typical examples of adaptive systems the lower-limit logic is paraconsistent and the higher-limit logic is **CL**, though one can also have **CL** as lower-limit logic and a non-monotonic logic as upper-level logic. Batens has a pragmatic instrumentalist approach to logic: the choice of a logic depends on the intended domain of application. From this point of view, Priest’s approach to quasi-validity consists in one among many possible choices of a lower and an upper-level logic. On the contrary, Priest takes his paraconsistent system **LP** as the “true logic”, which tells us which arguments are really valid. Once we move from it to the realm of **CL**, we have only quasi-validity and thus conclusions that cannot be taken for granted in spite of the truth of the premises.[4]

Harman is, like Priest, sensitive to the need of resorting, once we follow the traditional response, to some sort of quasi-validity for those rules that seem to be valid, but really are not (see Harman 1986). But for Harman the rules of **CL** are deductive and thus he attributes this weaker validity to the T-rules, which are, in his terminology “default rules”.

Who is right, Priest or Harman? It seems very difficult to find a principled reason to pick some rules of $\mathbf{CL}^+$, as opposed to some others, as the culprits that are not truly deductive and at most exhibit this weaker sort of validity, for all the rules of $\mathbf{CL}^+$ are abstracted from paradigmatic deductions (or at least are derived from rules abstracted from paradigmatic deductions). And perhaps this is why logicians cannot come to an agreement regarding which rules of $\mathbf{CL}^+$ are to count as really deductive and which ones are to be rejected and count as deductive at most in a weaker sense. As a matter of fact, in everyday life as well as in mathematical and scientific practices, we keep using all the rules of $\mathbf{CL}^+$, despite the paradoxes, as equally valid. I thus think that we must look for an alternative to the traditional response.

[4] The fact that I am discussing Batens’ approach in this section is not meant to indicate that it should be seen as a version of what I have called the “traditional response”. Given his pragmatic approach, Batens commits himself to a specific set of deductive rules only in the context of a certain domain and not in absolute terms. In contrast, the traditional response is attributed to those who take a certain set of rules as constituting the true deductive logic, independently of any context. A typical example is Priest when he proposes his paraconsistent system **LP** as the true logic. I am discussing Batens’ approach in this section merely because Priest’s quasi-validity can be understood in the framework of adaptive logic (see, e.g., the Sect. “What are ampliative adaptive logics?” at the “Adaptive Logics homepage” (http://logica.ugent.be/adlog/al.html)) in a way that helps us to better understand what Priest’s quasi-validity amounts to. Rather than assimilating Batens’ approach to logic to the traditional response, I would rather like to suggest (see Sect. 9.10 below) that such an approach may be resorted to in developing in a certain direction the alternative to the traditional response that I shall propose in the following. Moreover, an anonymous referee has even suggested that adaptive logic may provide the appropriate tools to formalize my proposal. Indeed, this is a line that I wish to explore in the future.

9.8 An Alternative Response

To find an alternative path, it is useful, I want to suggest, to have a look at how we exploit inductive arguments. Since their conclusions must be handled with care, it may not be rational to believe them, even if we believe their premises. Let us then take a little detour on induction.

9.8.1 The Defeasibility of Inductive Arguments

Inductive arguments may use deductive laws, but what distinguishes them is their appeal to inductive laws such as enumerative induction and abduction (infer that all F's are G's from the attribution of the property G to a good number of F's; infer P from P implies Q and Q). Just because these arguments avail themselves of such rules they are not truth-preserving and monotone: they are defeasible.

Thus, the conclusion of an inductive argument may well be false, despite the truth of the premises and we are all typically well aware of this. Nonetheless, we often choose to believe an inductively gained conclusion, because, to orient our actions toward the satisfaction of our goals, we typically need beliefs that cannot be reached deductively. After all, the conclusion of a good inductive argument may well have a good chance of being true, given the truth of the premises.

The defeasibility of inductive arguments however entails that the validity of inductive arguments cannot be regarded as absolute: it must have degrees. We can thus have two conflicting inductive arguments, arguments that are both inductively valid, with premises that we believe, and yet with incompatible conclusions. For example, consider these three arguments:

Example 5. (a) Measles causes red spots; (b) Peter has red spots. (c) Therefore Peter has measles.

Example 6. (a) Scarlet fever causes red spots; (b) Peter has red spots. (c) Therefore, Peter has scarlet fever.

Example 7. (a) Scarlet fever causes red spots; (b) Peter has red spots; (c) there is a scarlet fever epidemics. (d) Therefore, Peter has scarlet fever.

We can, perhaps debatably, regard Examples 5 and 6 as equally (and poorly) valid. But we certainly consider Example 7 as more valid (stronger) than both. If one is presented with both Examples 5 and 6 and believes their premises, one should hardly accept either conclusion (on the reasonable assumption that our ill friend does not have two diseases). But if one is then presented with Example 7, because new evidence has been found, one may well accept the conclusion that Peter has scarlet fever.

As this discussion illustrates, the defeasibility of inductive arguments suggests a prudent analog of PRD:

PRI. *Principle of Inductive Reasoning*. If (i) it is important, in view of a given goal, to have an opinion as to whether C is true or not, (ii) one finds an inductive argument A with conclusion C, (iii) one believes its premises; then one must believe C as well, provided one is unable to construct another argument B with conclusion C' such that (i) B is at least as valid as A, (ii) one believes its premises, (iii) C' is incompatible with C.

9.8.2 The Principle PR

The acceptance of a less prudent principle for deductive reasoning, i.e. PRD, is motivated, as we have seen, by the idea that deductions are undefeasible. We noted however that there are reasons to be skeptical about the traditional response to the dilemma of Sect. 9.6. Now, both the traditional response and the dilemma presuppose PRD. Perhaps, then, we can find a way out of the dilemma by rejecting PRD. But there is a problem here. As we have seen, PRD is motivated by the thesis of the undefeasibility of deductions and thus before rejecting it we should first get rid of this thesis. It might seem however that this can hardly be done, for the standard view about deductions may lead us to take them as essentially undefeasible or even undefeasible by definition. Given this way of characterizing deductions, we are plainly incoherent if we claim that deductions are not defeasible after all. In order to be able to claim this, we need a way of characterizing deductions that is not based on undefeasibility.

There is, I think, a way of doing this by relying on what we said above about how we abstract certain inferential laws from paradigmatically deductive arguments. We can thus first define *deductive* laws not so much as those laws which are undefeasible, but simply as those laws that are paradigmatically deductive, i.e., can be abstracted from paradigmatically deductive arguments,[5] or can be derived from paradigmatically deductive laws in the sense that we have explained above in discussing EFQ. It is worth noting that, in defining deductive laws in this way, without directly appealing to undefeasibility, we come close to the so-called proof-theoretic conception of validity (see, e.g., Priest 2006, Sect. 11.3), typically associated to Dummett. According to it, there are certain "logical" notions, such as conjunction, negation, etc., that are the notions they are precisely because they are governed by certain *crucial* inference rules governing their behavior in arguments; in other words, they are individuated or identified, at least in part, precisely by these rules. When any such crucial rule contributes to the identification of a certain notion,

[5] We shall not dwell here on the issue of whether, in a similar way, we can identify inductive laws as those that can be abstracted from paradigmatically inductive arguments or that are somehow derived therefrom.

this rule is a deductive rule.[6] This approach however requires an identification of the logical notions and of the crucial inference rules governing them. My appeal to paradigmatically deductive arguments can be seen as a way of doing precisely this: once we focus on paradigmatic deductions, we can enucleate from them the notions that, by being governed by certain inferential rules, appear to give such deductions their self-evident undefeasibility.[7] Such notions are thus seen as logical notions and the inference rules in question as deductive rules.

Deductions can then be seen as those arguments that are constructed solely on the basis of deductive laws. This view of deductive laws and arguments at most guarantees the undefeasibility of paradigmatically deductive arguments and is thus compatible with the existence of defeasible deductive arguments. The logical paradoxes can thus be taken to show precisely this, that there are deductive arguments that are defeasible. Indeed, the difficulties of the traditional response support this line.

Once we follow it, the difference between deductive and inductive arguments that motivated the distinction between PRD and PRI vanishes and the thought arises that, on analogy with inductive arguments, one should speak not so much of deductive validity simpliciter, but of degrees of validity. For on the basis of this notion of deductive law one cannot deny that paradoxically explosive arguments are valid. Yet, one could say that such arguments are hardly as valid as, e.g., the ones in

[6] This is usually put in semantic terms by saying that the meaning of a "logical constant" is provided by appropriate instructions (deductive inference rules) that tell us how the constant can be used in constructing deductive arguments. Thus, e.g., conjunction is nothing over and above that very concept $*$ governed by inference (deductive) rules such as these:

$$*E(a) \qquad \frac{A * B}{\mathrm{A}}$$

$$*E(b) \qquad \frac{A * B}{\mathrm{B}}$$

$$*I. \qquad \frac{\begin{matrix} A \\ B \end{matrix}}{A * B}$$

In the light of these rules, one can say that the meaning of "and" is fully conveyed by saying that in constructing a deduction, one is allowed to infer "*A* and *B*" from "*A*" and "*B*" and both "*A*" and "*B*" from "*A* and *B*."

[7] The reliance on paradigmatic deductions can perhaps also be used to defy the main worry usually associated with the proof-theoretic conception of validity, namely that we can introduce a connective, tonk, governed by inference rules that lead to explosion (see Prior 1960). The point is that the connective tonk does not occur in paradigmatically valid arguments, since it is an artifact that is not found in natural language. Be this as it may, it should be noted that the way in which I propose to deal with explosion in Sect. 9.8 below can also be applied to explosions generated by means of the connective tonk.

The feverish Peter. We can thus conceive of a general principle of reasoning, modeled on PRI, that takes care simultaneously of both deductive and inductive arguments:

PR. If (i) it is important, in relation to a given goal, to ascertain whether C is true or not, (ii) one finds a valid argument A with conclusion C, (iii) one believes the premises of A; then one must also believe C, provided one is unable to construct another valid argument B with conclusion C' such that (i) B is as valid as A, (ii) one believes the premises of B (or at least can in principle come to believe such premises),[8] (iii) C' is incompatible with C.

9.8.3 Degrees of Validity

Ideally, PR should be joined to a general criterion that tells us when a deduction A is more valid, or at least as valid as, another deduction A'. This may be difficult to obtain. But we know enough to face what I think is the most serious problem posed the logical paradoxes, namely the dilemma about rationality considered above. In a nutshell the idea is this. When, say via the Liar + EFQ or via Curry arguments, we reach an explosive conclusion C (say, that the mushrooms are edible) we can by the same token construct a parallel argument that leads to the opposite conclusion and that is certainly as valid as the original one. Moreover, this parallel argument, if it has any premises at all, can have premises that we can believe just like the premises of the original one. Let us see this in more detail.

Say that two deductions D and D' are isomorphic when D' is obtained from D by replacing propositions $P_1, \ldots, P_n$ in D with propositions $Q_1, \ldots, Q_n$ (without changing the justifications for the passage from one step to the next step). To illustrate, consider the following deduction:

1. $F \longrightarrow P$ (premise)
2. $P \longrightarrow R$ (premise)
3. F (hypothesis)
4. P (from 3 and 1 by MP)
5. R (from 2 and 4 by MP)
6. $F \longrightarrow R$ (from 3–5 by hypothetical reasoning).

If one replaces proposition F with proposition $\sim F$ in this deduction, one obtains this isomorphic deduction:

1. $\sim F \longrightarrow P$ (premise)
2. $P \longrightarrow R$ (premise)
3. $\sim F$ (hypothesis)

[8] The reason why this parenthetical remark is appropriate will become apparent below.

4. P (from 3 and 1 by MP)
5. R (from 2 and 4 by MP)
6. $\sim F \longrightarrow R$ (from 3–5 by hypothetical reasoning).

Now, we must clearly accept this thesis:

ID. If two deductions are isomorphic, then they are equally valid.

Consider now the version of Curry's paradox based on $[\lambda x \ \ x \in x \longrightarrow$ the mushrooms are edible$] \in [\lambda x \ \ x \in x \longrightarrow$ the mushrooms are edible$]$, with no premise, envisaged in Sect. 9.6, which allows Tom to conclude that the mushrooms are edible. Clearly, by replacing in this argument the proposition that the mushrooms are edible with its negation, we get an argument, isomorphic to the original one, with the conclusion that the mushrooms are not edible. By the above principle the two arguments must be equally valid.

Consider now the version of Curry's paradox based on the contingent premise (CP). One can in principle move from it to a parallel argument by (i) writing on a piece of paper, p_1', just the sentence "if the longest sentence on this piece of paper is true then the mushrooms are not edible"; (ii) replacing the proposition that the mushrooms are edible with its negation and (iii) replacing (CP) with

(CP′) The longest sentence written on p_1' is "if the longest sentence on this piece of paper is true then the mushrooms are not edible".

Clearly, this new argument is isomorphic to the original one and thus, given the above principle IP, as valid as the former. Moreover, it has a true contingent premise, just like the original one.

The paradoxically explosive arguments are then, we may say, those deductive arguments without false premises for which it is possible to construct, in the way we have just seen, corresponding isomorphic arguments, with the opposite conclusion. The following principle is of course very plausible:

PA. Any valid argument (whether inductive or deductive) that is not paradoxically explosive is more valid than a paradoxically explosive argument.

9.9 Another Look at the Dilemma

Let us now see how my proposal deals with the dilemma of Sect. 9.6. According to it, we need not deny that we are rational when we fail to believe the conclusion of a paradoxically explosive argument, because, in so doing, we do not fail to comply with a principle of reasoning. It is true that we do not comply with PRD, but this is not really a principle of reasoning. The principle that, instead of PRD, grounds our rational behavior when it comes to take advantage of our argumentative practices, is PR. And clearly, in the light of PR, Tom need not believe that the mushrooms are edible, because he knows (more or less consciously) that, just as he has constructed a deductive argument D (with premises he believes) that leads to the conclusion that the mushrooms are edible, he can similarly construct an equally valid (isomorphic) argument D' (with premises he believes) that leads to the conclusion that the mushrooms are not edible.

It is important to note that this approach is in line with the fact that in *The feverish Peter*, Tom is rational only if he believes the conclusion of the argument, D, in the story question. For in this case, we can assume, Tom is not aware of another argument D' such that (i) D' has premises that he believes, (ii) D' is as valid as, or more valid than, D; (iii) the conclusion of D is the opposite of the conclusion of D. In sum, PR is sufficient to grant the fact that, in these "normal" cases, one must believe, to be rational, the conclusion of a deductive argument that one knows and that has premises that one believes. The role of PA should be emphasized here, for without it, PR would be too weak for this job. In fact, without PA, PR could not prescribe that a subject should believe a conclusion C obtained from a "normal" valid argument with believed premises, if the subject can construct a paradoxically explosive argument with conclusion $\sim C$.

9.10 Hierarchies of Rules and Logical Truth

We have seen how to deal with paradoxically explosive arguments on the basis of plausible criteria that allow us to say that, at least in certain cases, some arguments have a higher degree of validity than other arguments or that two arguments must be equally valid. What to say about all the other cases?

One option is agnostic. We don't really know how things are in general. To illustrate this option, consider two arguments, A and B. Both of them have to do with a liar sentence L. Argument A just uses the laws of **CL** and concludes that the following must be true: $\sim(L\& \sim L)$. Argument B also appeals to the T-rules and concludes thus: $L\& \sim L$. Which argument is more valid? The agnostic option tells us that we have no way to tell and thus we had better consider both arguments as equally valid. Hence, given PR, neither conclusion can be believed.

Another, anti-agnostic, option is to think that there is some criterion to order all arguments by assigning them degrees of validity. We can do this by placing all basic laws in a hierarchy with the understanding that a higher place in the hierarchy means greater validity. Thus, for example, the rules at the top of the hierarchy will have a validity of degree 1 and any rule lower in the hierarchy will have a lower degree of validity $1 - r$, where r is some positive real less than 1; The number r is, we may say, the *distance* that a "lower rule" has from the topmost rules. Once rules are so ordered, we can assign a degree of validity to an argument by subtracting from 1 the sum of all values d such that d is a distance that a lower rule used in the argument has from the topmost rules. To illustrate, imagine that all rules of **CL** are higher in the hierarchy than the T-rules.[9] In this case the argument A envisaged

[9] If we wish to avoid that an argument is assigned a negative number as degree of validity we must choose appropriately small reals as degrees of validity of rules that are not topmost in the hierarchy.

in the previous paragraph wins and we prefer the conclusion $\sim(L\& \sim L)$ to the conclusion $L\& \sim L$. Contrariwise, imagine that some of the rules of **CL**, say those not supported by Priest's system **LP**, have a degree of validity lower than that of the T-rules. Then it can turn out that argument A has a lower degree of validity than the rival argument B and we should accept $L\& \sim L$ rather than $\sim(L\& \sim L)$ (thereby accepting dialetheism, the thesis, embraced by Priest, that there are true contradictions).

The anti-agnostic option raises the issue of finding criteria that allow us to choose a certain specific hierarchy. This may be difficult for one might run into problems analogous to those of the inferential version of the traditional response (see Sect. 9.7.2). But perhaps one can pursue different versions of the anti-agnostic option either for exploratory reasons or from the instrumentalist perspective defended by Batens in his approach to logic (see Batens 2014).

In any case, $\mathbf{CL}^+$ is, strictly speaking, explosive, but it enjoys a "partial non-explosivity", because not all its sequences of well-formed formulas are deductions and, among deductions, some are less valid than others. It thus seems appropriate to reformulate the notion of logical truth as follows: A is *logically true* iff there is a deduction D from no premises (or from premises that can be considered as "logical axioms") and there is no deduction D' with conclusion $\sim A$ from no premises (or from premises that can be considered as "logical axioms") such that D' is at least as valid as D.

Clearly, in this final section we are just scratching the surface of a number of complex matters and we must leave for another occasion a deeper investigation of them.

Acknowledgements I wish to thank an anonymous referee for his/her very useful comments.

References

Antony, L. (2004). A naturalized approach to the a priori. *Philosophical Issues, 14*, 1–17.

Arnold, J., & Shapiro, S. (2007). Where in the (world wide) web of belief is the law of non-contradiction? *Noûs, 41*, 276–297.

Azzouni, J. (2006). *Tracking reason*. Oxford: Oxford University Press.

Batens, D. (1999). Inconsistency-adaptive logics. In E. Orłowska (Ed.), *Logic at work: Essays dedicated to the memory of Helena Rasiowa* (pp. 445–472). Heidelberg/New York: Physica Verlag (Springer).

Batens, D. (2014). Adaptive logics as a necessary tool for relative rationality: Including a section on logical pluralism. In E. Weber, D. Wouters, & J. Meheus (Eds.), *Logic, reasoning, and rationality* (Vol. 5). Dordrecht: Springer.

Field, H. (2004). The consistency of the naive theory of properties. *Philosophical Quarterly, 54*, 78–104.

Gupta, A., & Belnap, N. (1993). *The revision theory of truth*. Cambridge: MIT.

Harman, G. (1986). *Change in view*. Cambridge: MIT.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*, 690–716.

Landini, G. (1998). *Russell's hidden substitutional theory*. Oxford: Oxford University Press.

Meyer, R. K., Routley, R., & Dunn, J. M. (1979). Curry's paradox. *Analysis, 39*, 124–128.

Orilia, F. (2000). Property theory and the revision theory of definitions. *Journal of Symbolic Logic, 65*, 212–246.

Priest, G. (1987). *In contradiction*. Dordrecht: Nijhoff.

Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Clarendon Press.

Prior, A. N. (1960). The runabout inference-ticket. *Analysis, 21*, 38–39.

Chapter 10
Contradictory Concepts

Graham Priest

10.1 Introduction

That we have concepts which are contradictory is not news. That there may be things which satisfy them, dialetheism, is, by contrast, a contentious view. My aim here is not to defend it, however[1]; and in what follows, I shall simply assume its possibility. Those who disagree are invited to assume the same for the sake of argument. The point of this essay is to think through a raft of issues that the view raises. In particular, we will be concerned with three inter-related questions:

1. Are the contradictions involved simply in our conceptual/linguistic representations, or are they in reality? And what exactly does this distinction amount to anyway?
2. Assuming that it is only in the former, can we get rid of them simply by changing these?
3. If we can, should we do so?

I will take up these issues, in that order, in the three parts of the paper. The journey will take us through a number of important issues in metaphysics, semantics, and epistemology.[2]

[1] This is done in Priest (1987, 1995, 2006). The topic is discussed by numerous people in the essays in Priest et al. (2004) and the references cited therein.

[2] Much of the paper has been provoked by many years of enjoyable discussion with Diderik Batens—including his generous comments on some earlier drafts of this paper. I thank him for all of this. The paper was originally written for a conference in honour of his 60th birthday.

G. Priest (✉)
Graduate Center, City University of New York, New York, NY, USA

Department of Philosophy, University of Melbourne, Parkville, VIC, Australia
e-mail: g.priest@unimelb.edu.au

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_10,
© Springer Science+Business Media Dordrecht 2014

10.2 Dialetheism, Concepts, and the World

10.2.1 Contradiction by Fiat

A dialetheia is a pair of statements of the form A and $\neg A$ which are both true (Priest 1987, p. 4). We may think of statements as (interpreted) sentences expressed in some language—a public language, a language of thought, or whatever. In this way they contrast, crucially, with whatever it is that the statements are *about*. Let us call this, for want of a better name, *the world*.

One thing that partly determines the truth value of a statement is its constituents: the meanings of the words in the sentence, or the concepts the words express. (Conceivably, one might draw a distinction here, but not one that seems relevant for present purposes.) Let us call these things, again for want of a better word, *semantic*. In certain limit cases, such as 'Red is a colour', semantic factors may completely determine the truth value of a statement. In general, however, the world is also involved in determining the truth value. Thus, the statement that Melbourne is in Australia is made true, in part, by a certain city, a certain country—literally part of this world.[3]

Given that dialetheias are linguistic, one natural way for them to arise is simply in virtue of linguistic/conceptual fiat. Thus, suppose we coin a new word/concept, 'Adult', and stipulate that it is to be used thus (see Priest 2001):

- If a person is 16 years or over, they are an Adult
- If a person is 18 years or under, they are not an Adult

Now suppose there is a person, Pat, who is 17. Then we have:

(*) Pat is both an Adult and not an Adult.

Of course, one can contest the claim that the stipulation succeeds in giving the new predicate a sense. Deep issues lurk here, but I will not go into them, since my concern is with other matters. I comment only that the stipulation would seem to be just as successful as the dual kind, endorsed by a number of people (e.g., Soames 1999), which under-determine truth values—such as the following for 'Child':

- If a person is 16 years or under, they are a Child
- If a person is 18 years or over, they are not a Child

The conference did not eventuate; but I'm delighted to dedicate the paper to him anyway. Diderik and I come at dialetheism from very different general perspectives. In particular, he gives much more importance to the role of context in semantics and epistemology than I do. See, e.g., Batens (1985, 1992) and Meheus and Batens (1996). Some of the matters I discuss here are difficult disengaged from these differences. I do my best.

[3] Quineans would, of course, reject the distinction being made here between semantic and worldly factors. This is not the place to defend the notion of analyticity. I do so in Priest (1979).

Assuming the stipulation of the kind involved in ‘Adult’ to work, we have a certain sort of dialetheia here. We might call it, following Mares (2004), a semantic dialetheia. Note that, in terms of the distinction just drawn between semantic and worldly factors, the epithet is not entirely appropriate. The truth of (*) is determined only in part by semantics; some worldly factors are also required, such as Pat and Pat’s age. Still, let us adopt this nomenclature.

10.2.2 Semantic Dialetheism

The dialetheism engendered by the definition of ‘Adult’ is transparent. There are other examples which are, plausibly, of the same kind, though they are less transparent. One of these concerns dialetheias apparently generated by bodies of laws, rules, or constitutions, which can also be made to hold by fiat. Thus, suppose that an appropriately legitimated constitution or statute rules that[4]:

- Every property-holder shall have the right to vote
- No woman shall have the right to vote

As long as no woman holds property, all is consistent. But suppose that, for whatever reason, a woman, Pat, comes to own property, then:

- Pat both has and has not got the right to vote.

Examples that are arguably of the same kind are given by multi-criterial terms, see Priest (1987, 4.8) and Priest and Routley (1989, Section 2.2.1). Thus, suppose that a criterion for being a male is having male genitalia; and that another criterion is the possession of a certain chromosomic structure. These criteria may come apart, perhaps as the result of surgery of some kind. Thus, suppose that Pat has female genitalia, but a male chromosomic structure. Then:

- Pat is a male and not a male.

In this case, there is no fiat about the matter. One cannot, therefore, argue that the contradiction can be avoided by supposing that the act of fiat misfires. What one has to do, instead, is to argue that the conditions in question are not criterial. Again, I shall not pursue the matter here.

A final example that is, arguably, in the same camp, is generated by the Abstraction Principle of naive set theory (see Priest 1987, Chap. 0):

Abs Something is a member of the collection $\{x : A(x)\}$ iff it satisfies the condition $A(x)$.

[4] The example comes from Priest (1987, 13.2).

This leads to contradiction in the form of Russell's paradox.[5] Again, there is no fiat here.[6] If one wishes to avoid the contradiction, what one must contest is the claim that satisfying condition $A(x)$ is criterial for being a member of the set $\{x : A(x)\}$—or, what arguably amounts to the same thing in this case, that **Abs** is true solely in the virtue of the meanings of the words involved, such as 'is a member of'.

Again, let us not go into this here. The point of the preceding discussion is not to establish that the contradictions involved are true, but to show that they may arise for reasons that are, generally speaking, linguistic/conceptual.

10.2.3 Contradictions in the World

Some have felt that there may be a more profound sort of contradiction, a contradiction in the world itself, independent of any linguistic/conceptual considerations. True, these are not strictly dialetheias as I have defined them, but let us call such things, following Mares (2004) again, *metaphysical dialetheias*.[7]

A major problem here is to see exactly what a metaphysical dialetheia might be. Even someone who supposes that dialetheias are solely semantic will accede to the thought that there are contradictions in the world, in one sense. None of the contradictions we considered in the previous sections, with perhaps the exception of Russell's paradox, is generated purely by semantic considerations. In each case, the world has to cooperate by producing an object of the appropriate kind, such as the much over-worked Pat. The world, then, is such that it renders certain contradictions true. In that sense, the world is contradictory. But this is not the sense of contradiction that is of interest to metaphysical dialetheism. The contradictions in question are still semantically dependent in some way. Metaphysical dialetheias are not dependent on language at all; only the world.

But how to make sense of the idea? If the world comprises objects, events, processes, or similar things, then to say that the world is contradictory is simply a category mistake, as, then, is metaphysical dialetheism.[8] For the notion to get a grip, the world must be constituted by things of which one can say that they are true or false—or at least something ontologically similar.

Are there accounts of the nature of the world of this kind? There are. The most obvious is a Tractarian view of the world, according to which it is composed of facts. One cannot say that these are true or false, but one can say that they obtain

[5] Take $A(x)$ to be $x \notin x$, and r to be $\{x : x \notin x\}$. Then we have $y \in r$ iff $y \notin y$. Hence, $r \in r$ iff $r \notin r$, and so $r \in r$ and $r \notin r$.

[6] An example of a similar kind, which does have an explicit element of fiat, is that of the Secretaries' Liberation League, given by Chihara in Chihara (1979).

[7] A number of people have taken me (mistakenly) to be committed to this kind of dialetheism. See Priest (1987, 20.6).

[8] The point is made in Priest (1987, 11.1).

or do not, which is the ontological equivalent. Given an ontology of facts to make sense, metaphysical dialetheism may be interpreted as the claim that there are facts of the form A and $\neg A$, say the facts that Socrates is sitting and that Socrates is not sitting. But as this makes clear, there must be facts of the form $\neg A$; and since we are supposing that this is language-independent, the negation involved must be intrinsic to the fact. That is, there must be facts that are in some sense negational, negative facts.[9] Now, negative facts have had a somewhat rocky road in metaphysics, but there are at least certain well-known ways of making sense of the notion, so I will not discuss the matter here.[10]

If one accepts an ontology of facts, fact-like structures, or something of this kind, then metaphysical dialetheism makes sense. Note, moreover, that if one accepts such an ontology, metaphysical dialetheism is a simple corollary of dialetheism. Since there are true statements of the form A and $\neg A$ then there are facts, or fact-like structures, corresponding to both of these.[11] All the hard work here is being done by the metaphysics; dialetheism itself is playing only an auxiliary role.

10.3 Conceptual Revision

10.3.1 Desiderata for Revision

Still, a metaphysics of facts (including negative facts) is too rich for many stomachs. Suppose that we set this view aside. If we do, all dialetheias are essentially language/concept dependent. In this way, they are, of course, no different from any other truths. But some have felt that, if this be so, contradictions are relatively superficial. They can be avoided simply by changing our concepts/language. Compare the corresponding view concerning vagueness, held, for example, by Russell (1923). All vagueness is in language. Reality itself is perfectly precise. Vague language and its problems may, therefore, be avoided by changing to a language which mirrors this precision.

[9] This isn't quite right. Facts may not themselves be intrinsically negative: the *relation* between the facts that A and that $\neg A$ must be intrinsic. But this does not change matters much: there must still be some kind of negativity in reality. There are other ways of making sense of the idea that the world itself is contradictory. For example, it may be held that reality is composed of properties, and that objects are bundles of properties. Then a contadictory world would be one in which there are property-bundles which contain the properties P and $\neg P$, for some P. Again, there must be some kind of negativity in reality. This time, negative properties.

[10] In situation semantics, states of affairs come with an internal "polarity bit", 1 or 0. Facts with a 0 bit are negative. Alternatively, a positive fact may be a whole comprising objects and a positive property/relation; whilst a negative fact may be a whole comprising objects and a negative property/relation. For a fuller discussion of a dialetheic theory of facts, see Priest (2006, Chap. 2).

[11] This assumes that all truths correspond to facts. In principle, anyway, one could endorse a view to the effect that some kinds of sentence are true in virtue of the existence of corresponding facts, whilst others may have different kinds of truth-makers.

Contradictions may certainly be resolved sometimes. Thus, consider the legal example concerning Pat and her rights. If and when a situation of this kind arises, the law would, presumably, be changed to straighten out the conflicting conditions for being able to vote. Note, however, that this is not to deny dialetheism. The situation before the change was dialetheic. The point of the change is to render it not so. Note, also, there is no a priori guarantee that making changes that resolve this particular contradiction will guarantee freedom from contradiction *in toto*. There may well be others. Indeed, making changes to resolve this contradiction may well introduce others. Laws comprise a complex of conceptual inter-connections, and the concepts apply to an unpredictable world. There is certainly no decision procedure for consistency in this sort of case; nor, therefore, any guarantee of success in avoiding dialetheism in practice.[12]

But maybe we could always succeed in principle. Consider the following conjecture:

- Whenever we have a language or set of concepts that are dialetheic, we can change to another set, at least as good, that is consistent.

The suggestion is, of course, vague, since it depends on the phrase ‘at least as good’. Language has many purposes: conveying information, getting people to do things, expressing emotions. Given the motley of language use, I see no reason to suppose that an inconsistent language/set of concepts can be replaced by a consistent set which is just as good for all the things that language does. I don’t even know how one could go about arguing for this.

Maybe we stand more chance if we are a little more modest. It might be suggested that language has a primary function, namely, making statements (truth-apt sentences); and, at least for this function, given an inconsistent language/set of concepts, one can always replace it with a consistent one that is just as good. The claim that this is the primary function of language may, of course, be contested; but let us grant it here. We still have to face the question of what ‘just as good’ means now, but a natural understanding suggests itself: the replacement is just as good if it can describe every situation that the old language describes. Let us then consider the following conjecture[13]:

[12] Actually, I think that the change here is not so much a change of concepts as a change of the world. Arguably, the change of the law does not affect the meanings of ‘vote’, ‘right’, etc. The statement ‘Pat has the right to vote’ may simply change its truth value, in virtue of a change in the legal “facts”.

[13] Batens (1999, p. 267) suggests that a denial of this conjecture is the best way to understand a claim to the effect that the world is inconsistent. ‘[I]f one claims that the world is consistent, one can only intend to claim that, whatever the world looks like, there is a language L and a [correspondence] relation R such that the true description of the world as determined by L and R is consistent.’ He maintains an agnostic view on the matter. See also Batens (2002, p. 131).

- Any language (set of concepts), L, that describes things in a dialetheic way, can be replaced by a consistent language (set of concepts), L', that can describe every situation that L represents, but in a consistent way.

The conjecture is still ambiguous, depending how one understands the possibility of replacement here. Are we to suppose this to be a practical possibility, or a merely theoretical one? If the distinction is not clear, just consider the case of vagueness again. If there is no such thing as vagueness *in re*, we could, in principle, replace our language with vague predicates by one whose only predicates are crisp. But the result would not be humanly usable. We can perceive that something is red. We cannot perceive that it has a wavelength of between exactly x and y Ångstroms, where x and y are real numbers. A language with precise colour predicates would not, therefore, be humanly usable. Any language that can be used only by someone with superhuman powers of computation, perception, etc., would be useless.

To return to the case of inconsistency, we have, then, two questions:

- Can the language be replaced in theory?
- Would the replacement be possible in practice?

A few things I say will bear on the practical question,[14] but by and large I shall restrict my remarks to the theoretical one. This is because to address the practical question properly one has to understand what the theoretical replacement is like. In other words, not only must the answer to the theoretical question must be ‘yes’, the answer must provide a sufficiently clear picture of the nature of the replacement. Nothing I go on to say will succeed in doing this. I have stressed the distinction mainly to point out that even if the answer to the theoretical question is ‘yes’, the replaceability conjecture has another hurdle to jump if the victory for those who urge replacement is to be more than Phyrric.

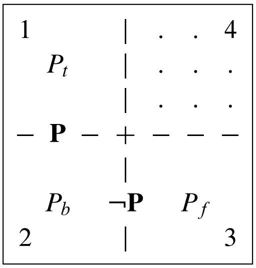
So let us address the theoretical question. Is it true? Yes, but for entirely trivial reasons. L' can be the language with just one sentence, ¥. ¥ is true of any situation. Thus, every situation is describable, and consistently so. (The language does not even contain negation.) But this is not an interesting answer to the question, and the reason is obvious. We have purchased consistency at the cost of the loss of expressive power. To make the question interesting, we should require L' to have the same expressive power as L—or more. That is, everything that L is able to express, L' is able to express. The idea is vague. What, exactly is it for different languages to be able to express the same thing? But it is at least precise enough for us to be able to engage with the question in a meaningful way.

[14] I note that Batens (2002, p. 131, fn. 7) suggests that a consistent replacement for an inconsistent language might well be required to have a non-denumerable number of constants, which would make it humanly unusable.

10.3.2 The Possibility of Revision

Return to the case of multiple criteria. A natural thought here is that we may effect an appropriate revision by replacing the predicate/concept *male* with two others, $male_1$, corresponding to the first criterion, and $male_2$, corresponding to the second. Pat is a male_2, but not a male_1, so the contradiction is resolved, and what used to be expressed by 'x is male', can now be expressed by 'x is $\text{male}_1 \vee x$ is male_2'. So far so good; but note that there is no guarantee that in this complex and unpredictable world the result will be consistent. The predicates 'male_1' and 'male_2' may themselves turn out to behave in the same inconsistent way, due to the fact that we have different criteria for 'genitalia' or 'chromosome'. More importantly, the resolution of this dialetheia depends on the fact that the old predicate falls neatly apart into two, individuated by different criteria. This will not be the case in general.—Just consider the case of 'Adult', for example, which is not multi-criterial in the same way.

We might attempt a more general way of resolving dialetheias as follows. Suppose we have some predicate, P, (like 'Adult') whose extension (the set of things of which it is true) and co-extension (the set of things of which it is false) overlap. Given that we are taking it that our predicates do not have to answer to anything in the world, we may simply replace P with the three new predicates, P_t, P_f, and P_b, such that the things in the extension of P_t are the things that are in the extension of P but not its co-extension; the things in the extension of P_f are the things that are in the co-extension of P but not its extension; the things in the extension of P_b are the things that are in both the extension and co-extension of P. The co-extension, in each case, is simply the complement. The situation may be depicted by the following diagram. For future reference, I call this the *Quadrant Diagram*. The numbers refer to the quadrants.



The left-hand side is the extension of P. The bottom half is the co-extension of P. Quadrant 4 comprises those things of which P is neither true nor false, and for present purposes we may take this to be empty.[15] The three new predicates have as extensions the other three quadrants. Each of the new predicates behaves

[15] Note that, if it is not, the same procedure can be used to get rid of truth value gaps.

consistently. Any dialetheia of the form $Pa \wedge \neg Pa$ is expressed by the quite consistent $P_b a$, and the predicate Px is now expressed, again, as a disjunction, $P_t x \vee P_b x$.[16]

So far so good. But recall that the new language must be able to express everything that the old language expressed. A necessary condition for this is that any situation described by the old language can be described by the new. To keep matters simple for the moment, let us suppose that the old language contains only the predicate P and the propositional operators of conjunction, disjunction, and negation. We have seen how any atomic sentence, A, of the old language can be expressed equivalently by one, A^+, in the new. If this translation can be extended to all sentences, then any situation describable in the old language is describable in the new. The natural translation is a recursive one. For the positive connectives:

- $(A \vee B)^+$ is $A^+ \vee B^+$
- $(A \wedge B)^+$ is $A^+ \wedge B^+$

But what of $\neg A$? We certainly cannot take $(\neg A)^+$ to be $\neg(A^+)$. $\neg Px$ is true in the bottom half of the Quadrant Diagram, whilst $\neg(P_t x \vee P_b x)$ is not true in quadrant 2. In this case there is an easy fix. $\neg Px$ is equivalent to $P_b x \vee P_f x$. So we can deal with the atomic case. What of the others? There is a simple recipe that works:

- $(\neg(A \vee B))^+$ is $\neg(A^+) \wedge \neg(B^+)$
- $(\neg(A \wedge B))^+$ is $\neg(A^+) \vee \neg(B^+)$
- $(\neg\neg A)^+$ is A^+

In other words, we can drive the negations inwards using De Morgan laws and double negation until they arrive at the atoms, where they are absorbed into the predicate. In this way, every sentence of the old language is equivalent to a consistent one in the new language.

The end can therefore be achieved for this simple language. But, for the strategy to work, it must be implementable with much more complex and realistic languages. In particular, it must work for conditionals, quantifiers of all kinds, modal and other intentional operators; and is not at all clear that it can be made to do so. At the very least, then, the onus is on the proponent of the strategy to show that it can.

Moreover, there are general reasons for supposing that it cannot. Extending the translation to intentional operators would seem to provide insuperable difficulties. Take an operator such as ‘John believes that’, $\mathfrak{B}$. How are we to handle $\mathfrak{B}A$? The only obvious suggestion that $(\mathfrak{B}A)^+$ is $\mathfrak{B}(A^+)$, and this will clearly not work. Even logical equivalence does not guarantee equivalence of belief: one can believe $\neg\neg A$ without believing A, for example. Hence, even if A and A^+ express the same situation in some sense, one could have $\mathfrak{B}A$ without having $\mathfrak{B}A^+$. The trouble is that belief and similar mental states are intentional, directed towards

[16] Batens notes this idea in Batens (1999, p. 271, 2002, p. 132). He also notes that in such a transition the theory expressed in the new language may lose its coherence and conceptual clarity, making it worse.

propositions/sentences. These seem to be integral to the intentional state in question, and so cannot be eliminated if we are to describe the intensional state. (Indeed, the same is true of *all* conceptual revisions. If people's thoughts are individuated in terms of old concepts, one cannot describe those thoughts if the concepts are junked.)

One possible suggestion at this point is simply to take $(\mathfrak{B}A)^+$ to be $\mathfrak{B}A$ itself. Of course, if we leave it at that, we have not rid ourselves of the dialetheic concepts, since these are still occurring in the language. But we might just treat $\mathfrak{B}A$ as a new atomic sentence—a single conceptual unit. The problem with this is clear. There would be an infinite number of independent atomic sentences, and the language would not be humanly learnable. The construction would fail the practicality test. And even then, given that the language contains other standard machinery, there would still be expressive loss. For example, we would no longer have a way of expressing things such as $\exists x(Px \wedge \mathfrak{B}Px)$ or $\forall p(\mathfrak{B}p \rightarrow p)$.

Nor is this just a problem about mental states. It applies to intensional notions generally. Thus, consider the statement 'That A confirms that B'. This is not invariant under extensional equivalence. Let us make the familiar assumption that all creatures with hearts are creatures with kidneys.[17] Consider the information that $a_1, \ldots, a_n$ are creatures of kind k with a heart. This confirms the claim that all creatures of kind k have a blood circulation system. The information is extensionally equivalent to the information that $a_1, \ldots, a_n$ are creatures of kind k with kidneys. This does not confirm the claim that all creatures of kind k have a blood circulation system.[18]

10.3.3 Expressive Loss

But worse is yet to come for the conjecture that we can, in theory, always replace an inconsistent language with a consistent one. Suppose that the project of showing that every situation describable in the old language can be described in the new can be carried out, in the way just illustrated or some similar way. This is not sufficient to guarantee that there is no expressive loss.

Consider the naive notion of set again. This is characterised by the schema:

Abs $x \in \{y : A(y)\} \leftrightarrow A(x)$

which gives rise to inconsistency, as we have noted. Let us suppose that it were replaced with different notions in the way that we have just considered. Thus, we

[17] As a matter of fact, Diderik (an amateur beekeeper) tells me, this is false. Bees have a heart, but no kidneys.

[18] More generally, relations relevant to confirmation are well known not to be invariant under linguistic transformations. See, e.g., Miller (1974).

have three predicates $\in_t$, $\in_b$, and $\in_f$, where $x \in y$ is expressed by $x \in_t y \vee x \in_b y$. Let us write this as $x \in' y$. Given the above schema, we have:

Abs′ $x \in' \{y : A(y)\} \leftrightarrow A(x)$

and in particular:

$$x \in' \{y : \neg y \in' y\} \leftrightarrow \neg x \in' x$$

Substituting $\{y : y \notin' y\}$ for x gives us Russell's paradox, as usual. We have not, therefore, avoided dialetheism.[19] Why is this not in conflict with the discussion of the last section? The reason is essentially that the procedure of driving negations inwards, and finally absorbing them in the predicate, produces a language in which there is no negation. The instance of Abs′ that delivers Russell's paradox cannot, therefore, even be formed in this language, since it contains negation. The procedure guarantees, at best, only those instances of Abs′ where $A(x)$ is positive (negation-free).

We face a choice, then. Either dialetheism is still with us, or we lose the general schema that we had before. But the Schema effectively characterizes the naive concept of set membership. So if we go the latter way, notwithstanding anything heretofore, there is still an expressive loss. We have lost a concept which we had before, with no equivalent replacement. We have lost the ability to express arbitrary set formation.

This provides us with an argument as to why we may not always be able to replace an inconsistent language/conceptual scheme with one that is consistent. There are cases where this can be done only with conceptual impoverishment. That one may achieve consistency by throwing away a concept is not surprising. The notion of truth gives rise to contradictions. No problem: just throw it away! But such a conceptual impoverishment will leave us the poorer.

If we were throwing away useless things, then, one might argue, this is no loss. But contradictory concepts may be useful; indeed, *highly* useful—contradictions notwithstanding. Thus, for example, the ability to think of the totality of all objects of a certain kind—closely related to our ability to quantify over all such objects, and to form them into a set—would seem to be inherent in our conceptual repertoires. It plays an essential role in certain kinds of mathematics (such as category theory), and in our ruminations about the way that language and other conceptual processes work. But abilities of this kind drive us into contradictions of the sort involved in discussions of the limits of thought.[20] If we threw away the ability to totalise in this way,[21] maybe this would restore consistency; but the cost would be to cripple the kind of mathematical and philosophical investigations that depend on it. To do so simply in the name of consistency would be like doing so in the name of an arbitrary and repressive government *diktat*.

[19] This is observed by Batens in Batens (2002, p. 132). See also Batens (1999, p. 272).

[20] A detailed discussion of all this can be found in Priest (1995).

[21] And can we? If one has such an ability, how *can* one lose it, short of some brain trauma?

The situation is not to be confused with that in which the concept of phlogiston was “replaced” by that of oxygen. We did not, in fact, dispense with the concept of phlogiston. We can still talk about it now. What was rejected there was the claim that something satisfies this notion. We now think that nothing does; in consequence, the concept is of no scientific use.

10.4 The Norm of Revision

10.4.1 Methodological Consistency

As we see, one cannot always replace an inconsistent language/set of concepts with a consistent one in a satisfactory way. But if we can, should we? Inconsistency should certainly be replaced sometimes. One of the functions of law is to guide action. Contradictory laws may frustrate this purpose—should we or should we not allow Pat to vote? But as far as the purely descriptive function of language goes, there would appear to be little point. The language/concepts provide a perfectly adequate representation of reality. If it ain’t broke, don’t try to fix it (see Priest 1987, 13.6).

There is no obvious reason why we should do so, but Batens (1999, 2002) has argued that it is sound methodology to replace an inconsistent set of concepts with a consistent one if we can do so, *ceteris paribus*. He cites Earman according to whom, though we have no reason to suppose the world to be deterministic, there is methodological virtue in trying to find deterministic theories. The same, according to Batens, is true of consistency. The virtue in the case of consistency is, of course, somewhat different. Batens calls it ‘precision’ and illustrates as follows[22]:

> Let P be a unary predicate of the language of an inconsistent theory, and let some paraconsistent logic **PL** be the underlying logic of the theory. ... P divides the objects into three subsets: those that are P only, those that are $\neg P$ only, and those that are both P and $\neg P$. The sentence $Pa \wedge \neg Pa$ unequivocally locates a amongst the objects that are inconsistent with respect to P. There is no way, however, to locate a in the union of the first and [second][23] set, not in the [third] only. Compare this situation to the one in which P belongs to a consistent theory (of which the underlying logic validates EFQ). Here P introduces two sets only; Pa unequivocally locates a in the first set, $\neg Pa$ unequivocally locates it in the second one. If there is a need for three sets, then one introduces a family of predicates (Carnap’s term), say P_1, P_2, and P_3. The predicates of the family are exhaustive and mutually exclusive. So they divide the objects into three sets, P_1a unequivocally locates a is the first, P_2a in the second, and P_3a in the third. Whether you need two or three sets (this depends on ‘the world’) the consistent theory is more precise. (Batens 1999, p. 271)

What to say about this argument?

[22] I change his notation to bring it into line with the rest of this essay.

[23] The text actually interchanges ‘second’ and ‘third’, but I take this to be a slip. The union of the first and third sets in Batens’ enumeration is characterised by P.

10.4.2 Precision

To evaluate it, let us start by getting clear about the notion of precision in play. Note that this has nothing to do with the truth conditions of negation: these are just as precise in the paraconsistent as in the classical case. Rather, the sense of precision at issue[24] is as follows. Refer again to the Quadrant Diagram. If we want to express the claim that an object, a, is in the union of quadrants 1 and 3, and we have the consistent language at our disposal, we can say $P_t a \vee P_f a$. But if we have only the inconsistent concepts at our disposal, the best we can do is:

(1) $Pa \vee \neg Pa$

We cannot rule out a's being in quadrant 2. In particular, the following won't do:

(2) $(Pa \vee \neg Pa) \wedge \neg(Pa \wedge \neg Pa)$

Given the standard semantics of negation and conjunction, $\neg(Pa \wedge \neg Pa)$ is true in quadrants 1, 2, and 3. The precision that Batens has in mind then, is the ability to characterize a's status in a more fine-grained way. We may now ask two crucial questions. First, is precision in this sense, a virtue? Second, does an inconsistent theory lack it? Take them in that order.

Precision is not necessarily a virtue. Recall the case of vagueness again. In our ordinary colour vocabulary, we can say that something is red, or some hue thereof, but we have no way of saying that it has some *precise* redness. Neither is this a problem. Our colour language is quite adequate for normal purposes. True, we can resort to the language of frequencies, but such discourse has imprecision of its own. We cannot specify a range of between x and y Ångstroms if x and y are real numbers not referred to by names or descriptions in our language. (There will always be such numbers, since the totality of real numbers is uncountable.) Nor, generally speaking, does this matter. Indeed, precision may not just fail to be a virtue; it may be a vice. As already observed, our colour language works only because its vagueness matches the limitations of our perceptual apparatus: a precise colour vocabulary would be unworkable. Another example: you do not know how to play cricket, and ask a friend. Reading out the rule book would provide a very precise answer, but it would not be very helpful. One needs the main points; details obfuscate. Sufficient to the occasion is the precision therefore.

Let us turn now to the second point. Are paraconsistent theories imprecise in the way suggested? Note, at the start, that there is nothing about paraconsistency, or even dialetheism as such, that prevents the language containing an operator that behaves as does classical negation. It is just that the operator isn't negation.

[24] Clarified by Batens in correspondence.

Of course, this possibility is ruled out if one wishes to run a dialetheic or paraconsistent line on the paradoxes of self-reference, since such an operator gives rise to triviality-producing contradictions.

However, assuming that there is no operator with the powers of classical negation in the language, is it the case that using a paraconsistent logic we cannot express the consistent parts of the Quadrant Diagram? As noted, $\neg(Pa \wedge \neg Pa)$ will not do. But it can be expressed if there is a truth predicate, T, and some naming device for sentences, $\langle . \rangle$, in the language. The four quadrants can be expressed by the following four conditions:

1. $T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle$
2. $T\langle Pa\rangle \wedge T\langle \neg Pa\rangle$
3. $\neg T\langle Pa\rangle \wedge T\langle \neg Pa\rangle$
4. $\neg T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle$

In particular, the union quadrants 1 and 3 can be specified by:

(*) $(T\langle Pa\rangle \vee T\langle \neg Pa\rangle) \wedge \neg(T\langle Pa\rangle \wedge T\langle \neg Pa\rangle)$.

Note that $\neg T\langle Pa\rangle$ is not equivalent to $T\langle \neg Pa\rangle$ (see Priest 1987, 4.9); if it were, and given the T-schema, $\neg(T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle)$ would be equivalent to $\neg(Pa \wedge \neg Pa)$.

Batens would no doubt object at this point. If the negation used in (*) is paraconsistent (which I take it to be), the sentence could be true even though $T\langle Pa\rangle \wedge T\langle \neg Pa\rangle$ (second quadrant) holds as well. The diagram itself might be inconsistent. If one objects in this way, the point is no longer that the facts of the diagram cannot be represented, but that they cannot be represented in a way that guarantees consistency. This is true: there is nothing a paraconsistent logician can say that enforces consistency. But this is no objection, since exactly the same holds of one who subscribes to classical logic! Such a person can, of course, assert $\neg(Pa \wedge \neg Pa)$ where $\neg$ is, or is taken to be, classical negation; this does prevent them endorsing $Pa \wedge \neg Pa$ as well. If they do, then they will be committed to everything. This, I take it, is the relevance of the reference to EFQ (*ex falso quodlibet*, $\{A, \neg A\} \vdash B$) in Batens' words. As he says elsewhere:

> To adopt the ex falso quodlibet has dramatic consequences. Someone who asserts $\neg A$ is truly committed to the rejection of A: asserting A would commit one to triviality. The dramatic character lies in the fact that triviality constitutes the end of all thinking... (Batens 1990, p. 222)

However, enforcing collapse into triviality can be secured by perfectly legitimate paraconsistent means as well. A paraconsistent logician may endorse a claim of the form:

(**) $(Pa \wedge \neg Pa) \to \bot$

where $\to$ is a detachable conditional, and $\bot$ is a logical constant that implies everything. (It may be defined as $\forall x Tx$.) A subsequent endorsement of $Pa \wedge \neg Pa$ will then commit them to everything (and so, presumably, force then to give up something to which they are committed).[25] The classical logician is, in the end, then, no better off than the paraconsistent logician.

Batens addresses essentially this matter explicitly in a later article (Batens 2002, p. 142–4).[26] He claims that, at least without Boolean negation, there is no way to express the thought that two claims, A and B, are incompatible, or not jointly possible. You can't simply say $\neg\Diamond(A \wedge B)$. For that could be the case, even though $A \wedge B$ is true. Much the same considerations apply. Ruling out in the pertinent sense is a function of EFQ,[27] and this can be done by a paraconsistent logician using $\perp$. Batens points out that even the trio of claims A, B, and $(A \wedge B) \rightarrow \perp$ can be endorsed by someone who is prepared to accept that everything is true—trivialism. But classical logic is no defence against trivialism: trivialists *are* classical logicians! They endorse, and reason in accord with, all the principles of classical logic, including EFQ.

Batens goes on to argue for a further claim: the fact that it is *logically* possible to accept everything in a paraconsistent logic is a shortcoming in the context of theory-revision: to handle such revision in the case of inconsistent data requires an adaptive logic. Now, for a start, dialetheists can use adaptive logics. One is endorsed in Priest (1987, Chap. 16). But the point that theory-revision goes beyond logic is correct. Theory-revision uses norms of rationality that go beyond those of mere logic—adaptive or otherwise. It is rational to replace an old theory with a new one if that theory performs better on the aggregate of positive criteria for theory-choice, such as simplicity, unifying power, etc.[28] And because the mechanism is broader than that of logic, it can account for change in the received logical theory too.[29]

10.4.3 Boolean Negation: Again

There is another, and harder, point here.[30] Batens is, in fact, advocating not just replacing inconsistent concepts with consistent ones, but replacing concepts employing a paraconsistent logic with concepts employing classical logic.

[25] The conditional (**) is not a logical truth; one can think of it as part of the theory of P. (Classical logic, in effect, promotes this contingent truth into a necessary (logical) one.) This is beside the point, though; what is at issue is whether triviality can be triggered in a paraconsistent context.

[26] The objection is given under the rubric 'objections to dialetheism'. This is misleading, for he himself is a dialetheist. He holds that there are inconsistent concepts, and so dialetheias.

[27] There are other senses to do with denial. On these, see Priest (2006, Chap. 6). The claim about the expressive limitations of paraconsistent logic has been pressed most strongly by Shapiro in Shapiro (2004). See the discussion in Priest (1987, 20.4).

[28] One of these criteria may well be consistency. See Priest (2006, Part 3) for a full account of the details.

[29] See Priest (2006, p. 151) and for more detail, Priest (to appear).

[30] Many of the arguments in this section are given in more detail in Priest (1987, 2006). Hence, the treatment here can be reasonably terse.

The possibility of this presupposes that the notions of classical logic make sense, and, in particular, that Boolean negation does so. It seems to me that it does not.

The idea may seem absurd. Can't we simply recognise the meaning of classical negation? Unfortunately, no. Things do not wear their meaning—or lack thereof—on their face. Whether something is meaningful can be determined only by the articulation and application of a theory of meaning. A classical theory of meaning may deliver the result that Boolean negation is meaningful. But the adequacy of a classical theory of meaning is, in part, what the debate at hand is all about. And as far as Boolean negation goes, a dialetheic theory of meaning can side with an intuitionistic theory of meaning in holding that it does not. Nor need a classical logician feel smug about the matter. *No one*, on pain of triviality, can endorse both a classical notion of negation and an unrestricted truth predicate (see Priest 2006, Chap. 5). Hence a classical logician must deny the meaningfulness of the latter notion, which seems just as bad, if not worse.[31]

Why should we suppose that classical negation does not make sense? In a nutshell, the argument goes as follows (see Priest 2006, Chap. 5). A connective that satisfies the rules of Boolean negation appears to be in the same camp as Prior's connective *tonk* (a connective, †, satisfying the rules $A \vdash A \dagger B$, $A \dagger B \vdash B$). If such a connective is in the language, then any sentence entails any sentence. Similarly, if a connective obeying the rules of Boolean negation is in the language, then any sentence entails any sentence (in the context of self-reference and the T-schema). Since *tonk* is meaningless; so is Boolean negation.[32]

But may we not show that Boolean negation is legitimate by giving it truth conditions in the standard way? Say:

- $\neg A$ is true in (a world of) an interpretation if A is not true (there).

The truth conditions may determine a perfectly legitimate notion, but to establish that they deliver a notion underwriting EFQ we need to do more than state truth conditions; we need to reason about what follows from them. And—to cut a long story short—we have no reason to suppose that the conditions do so unless we reason classically—in particular, using Boolean negation—in the metalanguage, and so presuppose the meaningfulness of the very notion whose meaningfulness we are supposed to be establishing.

This argument has been contested by Batens, who raises a number of objections (Batens 2002, p. 141 ff.).[33] One is the following. Negation may not actually be

[31] Of course, in consistent contexts, a paraconsistent negation may behave indistinguisably from classical negation. That does not mean that it is classical negation that is being used.

[32] A number of people have suggested to me that *tonk* is perfectly meaningful. Its meaning is just defective. I have no objection to this if one can give a satisfactory account of defective meaning (which I don't know how to do). The point is that, whatever one says about *tonk*, the same applies to Boolean negation.

[33] Batens' own views about meaning depend heavily on his contextualism, and he would dispute my whole approach to meaning, but that is far too big an issue to take on here. In what follows, his objections may be taken to be ad hominem.

needed to give the truth conditions of negation. Thus, assume for the sake of illustration that we have a three-valued semantics with the truth values $\{t\}$, $\{t, f\}$, and $\{f\}$, the first two being designated. We may give the truth conditions of negation without using negation, and thereby presupposing its properties, as follows:

- $\neg A$ is $\{t\}$ if the value of A is $\{f\}$
- $\neg A$ is $\{f\}$ if the value of A it is $\{t\}$ or $\{t, f\}$.

or as:

- $\neg A$ has the value $\{t\}$ if the value of A is designated, and $\{f\}$ otherwise.

Indeed we may; but in either case we have to reason using negation to show that the $\neg$ so defined grounds EFQ. Thus, we may establish that A and $\neg A$ never take a designated value together. But to establish that $A, \neg A \vDash B$, we need to reason that, since the premises are never both designated, whenever the premises are designated, so is the conclusion. This is an argument of the form $\neg A \vdash A \rightarrow B$, which a paraconsistent logician is not going to accept if $\rightarrow$ is a detachable conditional. Alternatively, if validity is defined in terms of truth preservation using the material conditional, the inference to the validity of EFQ is of the form $\neg A \vdash A \supset B$. The inference is perfectly correct; but now we cannot get to B from A and $\neg A$ because the material conditional does not detach. We still do not have the force of explosion.

Batens' next point concerns classical logic and metatheoretic reasoning. It might appear that paraconsistent logicians are committed to the meaningfulness of Boolean negation, since they use it themselves when doing the metatheory of paraconsistent logic. Certainly, it is often assumed that the metatheory of a logic, classical or non-classical, must be undertaken in classical logic; but this is false. There is no reason why, in principle, the metatheory of a logic cannot be undertaken in a non-classical logic. In practice, metatheory is done in informal set theory; the question is how to regiment this formally. Standardly, classical logic and ZF set theory suffice; this does not show that they are necessary. Batens points out that paraconsistent logicians have not given a great deal of thought to the paraconsistent regimentation of metatheory. This is a fair point. How best to turn the trick is still moot. One way of doing it is explained in Priest (1987, Chap. 18). I will not reproduce the details here, but the idea is that a certain understanding of paraconsistent set theory allows a paraconsistent logician simply to appropriate classical metatheoretic arguments.

Batens' third point is swift. He says:

> Once a ... dialetheist theory for handling functions will be around, things will get serious. The question will not any more be whether some metalinguistic negation is paraconsistent rather than classical, but whether $t \in \{f\}$ and hence $t = f$. (Batens 2002, p. 141)

The point, I take it, is that if triviality looms anyway, the blame cannot be laid at the door of Boolean negation. Why, exactly, triviality looms, Batens does not spell out here, but in Batens (1990, Sect. 5), when he considers the matter, he notes that if ν is an evaluation function, and we can show that $\{t\} = \nu(A) \neq \nu(B) = \{t, f\}$, this does not '*rule out*' (his italics) $\nu(A) = \nu(B)$, and so $\{t\} = \{t, f\}$. Indeed, in a sense, it does not. The question, though, is whether there is any reason to believe what is not ruled out to be true. Without this, the point has no bite.[34]

Batens' final major objections is to the effect that Boolean negation must be meaningful because people can reason in accord with the rules of classical logic—or any other logic. However, this may be explained without resort to an appeal to the meaningfulness of Boolean negation. Using a non-monotonic adaptive logic—of the kind pioneered by Batens—a paraconsistent logician may reason in exactly the same way as a classical logician in consistent situations, though the negation employed has exactly the same meaning as it does in the underlying monotonic paraconsistent logic (see Priest 1987, 8.6 and especially Priest 1987, Chap. 16). One cannot say the same thing about reasoning in accord with other logics, or (fortunately!) about reasoning using classical logic in inconsistent situations; but there is a much more general point here. Given *any* set of putatively logical rules—at least as long as they are not too complex—a person may follow them and know that they are doing so. Nothing follows about meaning at all, however. One can just as well follow the rules for reasoning with *tonk*. And just as with *tonk*, following the rules may lead to disaster.[35]

10.5 Conclusion

This has been an essay about contradictory concepts, concepts which generate dialetheias. Assuming there to be such things, three further claims are tempting. (1) Dialetheias are merely in our concepts; there are no such things as contradictions in re. (2) Dialetheias may always be removed by revising our concepts. (3) Even if this is not the case, if they can be, they should be, ceteris paribus. We have seen that there are ways in which one may resist all of these suggestions. I think that Hegel would have been delighted; but that is another matter.[36]

[34] There is, in fact, a small industry of people (not including Batens) who have attempted to produce arguments to this effect, based on "strengthened paradoxes". The arguments are discussed and rejected in Priest (1987, 20.3).

[35] Batens has one more objection, not about Boolean negation, but about the classical material conditional. The claim is that this is needed to deal with restricted quantification. This objection is answered in Priest (1987, 18.3). A more general discussion can be found in Beall et al. (2006).

[36] Versions of this paper, or parts of it, have been given under various titles at a number of philosophy departments and conferences over the last few years: the University of Melbourne, the University of Queensland, the *Australasian Association of Philosophy* (Australian National University), the University of Chapel Hill (North Carolina), the University of Connecticut, the Massachusetts Institute for Technology, *Logic and Reality* (Universities of Namur and Louvain la Neuve), the University of Gent, the City University of New York (Graduate Center), the *Fourth Cambridge Graduate Conference on the Philosophy of Logic and Mathematics*. I thank the participants for many lively discussions and helpful comments. Thanks for comments go also to two referees for the volume in which this essay appears.

References

Batens, D. (1985). Meaning, acceptance, and dialectics. In J. C. Pitt (Ed.), *Change and progress in modern science* (pp. 333–360). Dordrecht: Reidel.

Batens, D. (1990). Against global paraconsistency. *Studies in Soviet Thought, 39*, 209–229.

Batens, D. (1992). Do we need a hierarchical model of science? In J. Earman (Ed.), *Inference, explanation, and other frustrations: Essays in the philosophy of science* (pp. 199–215). Berkeley: University of California Press.

Batens, D. (1999). Paraconsistency and its relation to worldviews. *Foundations of Science, 3*, 259–283.

Batens, D. (2002). In defence of a programme for handling inconsistencies. In J. Meheus (Ed.), *Inconsistency in science* (pp. 129–150). Dordrecht: Kluwer.

Beall, J., Brady, R., Hazen, A., Priest, G., & Restall, G. (2006). Restricted quantification in relevant logics. *Journal of Philosophical Logic, 35*, 587–598.

Chihara, C. (1979). The semantic paradoxes: A diagnostic investigation. *Philosophical Review, 13*, 117–124.

Mares, E. (2004). Semantic dialetheism. In G. Priest, J. Beall, & B. Armour-Garb (Eds.), *The law of non-contradiction: New philosophical essays* (chap. 16). Oxford: Oxford University Press.

Meheus, J., & Batens, D. (1996). Steering problem solving between cliff incoherence and cliff solitude. *Philosophica, 58*, 153–187.

Miller, D. (1974). Popper's qualitative theory of verisimilitude. *British Journal for the Philosophy of Science, 25*, 160–177.

Priest, G. (1979). Two dogmas of quineanism. *Philosophical Quarterly, 29*, 289–30l.

Priest, G. (1987). *In contradiction: A study of the transconsistent*. Dordrecht: Martinus Nijhoff. Second edition, 2006, Oxford: Oxford University Press; references are to the second edition.

Priest, G. (1995). *Beyond the limits of thought*. Cambridge: Cambridge University Press. Second edition, 2002, Oxford University Press, Oxford.

Priest, G. (2001). Review of Soames (1999). *British Journal for the Philosophy of Science, 52*, 211–215.

Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.

Priest, G. (to appear). Logical disputes and the a priori.

Priest, G., & Routley, R. (1989). The philosophical significance of paraconsistent logic. In G. Priest, R. Routley, & J. Norman (Eds.), *Paraconsistent logic: Essays on the inconsistent* (chap. 18). Munich: Philosophia Verlag.

Priest, G., Beall, J., & Armour-Garb, B. (Eds.). (2004). The law of non-contradiction: New philosophical essays. Oxford: Oxford University Press.

Russell, B. (1923). Vagueness. *Australasian Journal Philosophy, 1*, 84–92. (Reprinted as Ch. 3 of R. Keefe & P. Smith (Eds.) (1999). *Vagueness: A reader*. Cambridge: MIT).

Shapiro, S. (2004). Simple truth, contradiction, and consistency. In *The law of non-contradiction: New philosophical essays* (chap. 20). Oxford: Oxford University Press.

Soames, S. (1999). *Understanding truth*. Oxford: Oxford University Press.

Chapter 11
Bloody Analogical Reasoning

Dagmar Provijn

11.1 Introduction

There is more to William Harvey than his discovery of the blood circulation, first presented in the *Exercitatio anatomica de motu cordis et sanguinis in animalibus* (1628) (henceforth: DMC). This major breakthrough in the history of medicine and biology was preceded by Harvey's thesis on the forceful systole and its corollary on the true nature of the pulse, both occurring in his Anatomical Lecture Notes—*Prelectiones Anatomiae Universalis*—(henceforth: ALN), originating from 1616 and gradually adapted to new findings over the course of several years. The thesis on the forceful systole and its corollary on the pulse are another fine accomplishment of Harvey, i.e. one that solved a problem that already lasted for several centuries. Though Galen, whose theory on the physiology of man had been predominant for hundreds of years, had his own account on the action of the heart and the nature of the pulse, the interrelation of both phenomena remained a topic of discussion in anatomical studies.[1] But there is even more. Anyone reading Harvey's DMC will notice that it contains plentiful of analogies. In fact, Harvey's writings reveal more than a mere use of illustrative analogies to embellish his arguments with commonplace analogues. Hence, the DMC is a fine subject for studying different manifestations of analogies, i.e. the different ways in which analogies and analogical

[1] Pietro d'Abano, for example, tackled the question whether the heart and arteries can dilate at the same time and answered it in the negative as they cannot attract (during dilatation) nor expel (during constriction) blood at the same time—(Pagel 1976, p. 67).

D. Provijn (✉)
Centre for Logic and Philosophy of Science, Ghent University (UGent), Ghent, Belgium
e-mail: dagmar.provijn@gmail.com

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_11,
© Springer Science+Business Media Dordrecht 2014

reasoning can be applied both in solving a problem and in the presentation of its results. However, even more important is the question in what respect these analogies may have contributed to the process in which Harvey finally came up with a new perspective on the heart.

In this paper I will show that, apart from their argumentation value, analogies and analogical reasoning also played a considerable role in Harvey's investigations from which his views on the real action of the heart, the true nature of the pulse and the movement of the blood originated. In Sect. 11.2 I will sketch the background that is needed to understand Harvey's project, presented as a problem solving process in terms of Batens' contextual model (see, for instance Batens 1992a,b). This background contains: (i) a presentation of Batens' contextual model, that is a variant of Nickles' constraint-inclusion model (see, for example Nickles 1981), dealing with the different components of a problem solving process; (ii) a general view on analogies and analogical reasoning; (iii) the physiological theory of Galen that was overturned by Harvey's novel view on the functioning of the cardio-vascular system but that also played an important role in the theoretical basis of Harvey's investigations; and (iv) a sketch of Harvey's problem solving processes that ultimately led to the discovery of the blood circulation. Section 11.3 is dedicated to the main topic of the paper, i.e, Harvey's use of analogical reasoning, situated in the sketch of his problem solving processes.

11.2 Background

11.2.1 Contextual Approach

Harvey's investigations will be presented as a sequence of problem solving processes in order to situate the instances of analogical reasoning that will be presented in Sect. 11.3 of this paper. In fact, this presentation is a 'rough' reconstruction of Harvey's reasoning processes leading to his discoveries and as such should not be interpreted as a 'rational reconstruction' of Harvey's discovery processes. I refer to a 'rough' reconstruction as it is implausible that a fully detailed reconstruction of Harvey's problem solving processes can be obtained. However, pinpointing the problems (a goal to be obtained, a question to be answered,...) of Harvey's investigations and an important part of the set of constraints constituting the relevant information and techniques at hand to find a solution is possible. At least, it is enough to determine where exactly, in the process that led to the discovery of the blood circulation, the instances of analogical reasoning may have played a considerable role. In a problem solving perspective on discovery processes, a problem always is connected to a set of constraints imposing restrictions on possible solutions and possible search paths that may lead to these solutions. In addition, Batens' contextual model distinguishes different kinds of constraints for the problem, hence a context (i.e. problem solving situation) consists of: (i) a problem, (ii) certainties, (iii) relevant items, (iv) methodological instructions and (v) participants in the problem solving process.[2] Certainties are considered necessarily true and they primarily function to determine the meaning of words, the meaning and structure of concepts and other components that are contained in the other context-elements and to determine the inferential operations that are justified within the context. Hence, the certainties fix the possible solutions of the problem. The relevant items, on the other hand, are considered contingently true and function to impose further restrictions on the possible solutions in order to derive the correct solution or to eliminate some of the possible solutions. Methodological instructions refer to all kinds of operations—ordered sets of instructions, well-defined problem solving methods, vague heuristic rules, rules of thumb—that may bring us closer to a solution.

Of course, it is possible that a context does not lead to a (unique) solution in which case we will speak of an ill-defined problem. Hence, if the problem solver still wants to obtain a (unique) solution, he or she will need to tackle a derived problem (i.e., what is going wrong in the original problem solving situation?) by moving to another context.

11.2.2 Analogical Reasoning

My approach on analogical reasoning is logico-philosophical and based on the perspectives on analogical reasoning as proposed in Meheus (2000) and Gentner (2003). Consequently, 'analogical reasoning will refer to processes in which inferences are made based on certain similarities between two domains which can be two objects, two classes of objects, . . .'—(Meheus 2000, pp. 24–25), or in other terms it is 'a kind of reasoning that applies between specific exemplars or cases, in which what is known about one exemplar is used to infer new information about another exemplar'—(Gentner 2003, p. 106). While considering analogies as such, one should realize that their use and understanding always requires some kind of analogical reasoning, whether to construct a novel analogy or to check whether a commonplace analogy applies to a specific situation. The application of analogies can serve different purposes. First of all, they can be useful in communication, both for pedagogical reasons in view of clarifying a concept by means of a well-accepted analogue or for rhetorical reasons aiming at persuasion, in most situations combined with an informative component. Secondly, they may allow for predications, for example, in the specific case of extrapolation. Finally, in the domain of problem solving and discovery processes they may enable the change of existing concepts and methods and they may even facilitate the creation of completely new concepts. For each of these functions analogies may have an inter-domain or intra-domain origin. In the field of problem solving and creativity one can also distinguish

[2] As Harvey's case does not fall in the category of non-individual problem solving, we may skip the complications involved in these situations.

between weak and strong analogies, based on the criterion whether the analogy as such is a sufficient reason to accept the conclusions that are derived from them. In the case of a weak analogy, the results only serve a heuristic function to obtain tentative solutions that need further support.

11.2.3 Galen

As Galen's medical doctrine was an important part of Harvey's medical background, I will sketch some of its main elements. In view of the present project, four characteristics of Galen's doctrine are primordial to fully grasp the change that Harvey's DMC engendered: (i) the influence of Hippocratic dietetic and humoral medicine (ii) the distinction between the venal and arterial system, (iii) the attractive force or faculty of the organs, and (iv) the centrifugal flow of the venal blood.

Humoral and dietetic medicine. Long before and even after Harvey, physicians conceived of the organism as a rather unstable entity for which a balance of the humors was primordial. In fact, the role of the physician was to understand the normal state of the bodily functions, how these were susceptible for variation, causing imbalance, and how to restore normality. As the human diet is a substantial cause of instability, dietetic medicine represented a major part of physiological knowledge. Though ancient natural philosophers mainly focused on the healthy organism and applied dissections in their teleological program to understand 'what it is to be—a specific organ', the majority of physicians was utterly cautious to learn from both dissection and vivisection. In view of the normality interpretation of medicine this suspicion is easily understood: dissection concerns death bodies which are not representative for the normal state of the living body and vivisection cause a violent disruption of the normal state of the body. It is important to notice that even Vesalius, for example, explicitly states that the cutting in venesection causes a flux of blood and spirits in the direction of the slash and therefore causes the bleeding. Hence, blood is attracted because of the disruption of the normal state and as such the bleeding is not conceived of as a simple release of blood from its normal flux.[3]

Two systems. In Galen's physiology, the venal and arterial system are clearly separated entities. The venal system serves a nutritional function and contains the liver, the veins and the right ventricle of the heart. The liver is the source of all blood; it receives *chyle*, a product of digested food, through the mesenteric veins, from which it produces blood that is distributed to all parts of the body. The right ventricle of the heart, just like any other organ in the body, attracts venal blood for different reasons: first of all for its own nutrition, furthermore to feed the lungs via the *vena arteriosa* (the pulmonary artery, in Galen's system a vein) after the blood has been further concocted and refined and finally to provide a source for the arterial

[3] Bylebyl (1979, pp. 28–62).

blood that is passed from the right ventricle to the left ventricle through pores in the interventricular septum. The arterial system contains the lungs, the left ventricle of the heart and the arteries. While dilating, the heart and the arteries attract air; the heart receives air from the lungs via the *arteria venosa* (the pulmonary vein, in Galen's system an artery) and the arteries through pores in the skin.[4] As such the venal blood is mixed with vital spirits from the air and via the arteries all organs can attract arterial blood that provides them with heat and spirits. The *arteria venosa* also has another function, i.e., the provision of arterial blood to the lungs and the removal of vaporous wastes created by the heating of the arterial blood.

Attraction. Galen's medical doctrine provides a nutritive physiology by pinpointing the anatomical components linking the dietetic and humoral parts of Hippocratic medicine. This doctrine aimed to establish the presence of four teleological powers in each part of the body to *attract*, *retain* and *assimilate* what is needed and to *expel* what is useless. The heart and arteries were attributed a special faculty causing the *active pulsatile movement* these incessantly undergo. In fact, for Galen, diastole is the most important and most active moment of the heart's movement, allowing the attraction of venal blood into the right ventricle and the attraction of air in the left ventricle to control the innate heat and the assimilation of vital spirits from the air. In diastole, the heart sets over a wave of dilatation to the arteries for the distribution of arterial blood throughout the body. Hence, the *arteries pulsate actively* and are filled *like a pair of bellows*.[5] It is worth mentioning that Galen did not consider the heart to be a muscle, as it's action could not be modified by the will. As he found that voluntary motion depends on the combination of muscles and nerves and as centrally controlled nerves appeared to be missing in the heart, Galen concluded that the heart is not a muscle. Hence, the concept of a forcibly dilating heart could not contradict the forcible contraction of a muscle if the heart would have been conceived of as such.[6]

Centrifugal flow. Venal and arterial blood flow respectively from the liver and the heart to the outer parts of the body. It is the efficacy of the tricuspid valve, allowing only for an influx of blood in the right ventricle, that convinced Galen to diverge from Aristotle's opinion that the heart is the sole central organ of the venal and arterial system.

[4] Actually, Galen is not unambiguously clear on this point. In his opposition against the Erasistratean view of the arteries only containing air, Galen never unequivocally stated whether air really entered the heart and arteries. The cooling function of the air could as well happen by means of contact rather than mixture—(Pagel 1967, pp. 129–132). However, in Harvey's age, the mixture viewpoint was predominant—(Bylebyl 1981, p. 153). Moreover, it was in attacking the idea that the *arteria venosa* predominantly contains air that Columbus posed the lesser circulation as a corollary—(French 1994, pp. 82–83).

[5] Bylebyl (1979, pp. 38–51), Pagel (1967, pp. 127–136) and Harvey (1628/1976, p. 15).

[6] French (1994, pp. 72–73).

11.2.4 Aristotelian Project

As mentioned before, the reconstruction of Harvey's problem solving processes will be incomplete by all means. Still, it will allow for situating Harvey's applications of analogies and analogical reasoning and for an evaluation of their importance. The reconstruction is based on Harvey's own ALN and DMC and is supported by different works of Harvey scholars as there are Walter Pagel, Jerome Bylebyl, Roger French and Gwenneth Whitteridge.[7] The final product will be a sketch of Harvey's overall project, pinpointing the central related problem solving processes that were involved in it and part of the contextual set-up of the processes.

Harvey's first problem was establishing the 'movement of the heart', an altogether Aristotelian project in the light of determining the true nature of the heart or 'what it is to be a heart'. Actually, this problem was part of a more extensive problem that envisaged to answer the relation between form and function of the heart. The description of the form of a specific organ was contained in the 'historia' that was obtained through dissection. The determination of the function, on the other hand, should be understood as containing different tasks, i.e. distinguishing the particular or proper *movement* of the organ from the different movements it may show; establishing the single overall *action* of the organ resulting from the addition of its different movements; defining the *use* or *purpose* of the organ to specify what kind of thing the organ is and what *final cause* it serves; and finally answering the teleological question of what the *utilitas* of the organ is or how it is well-suited for its *final cause*.[8]

So, Harvey's intellectual background certainly was influenced by Aristotle's natural philosophy, an influence that has its main origin in Harvey's stay at the University of Padua where he most probably came in contact both with the revival of Galenism and of Aristotle's teleological project, and more specifically with the works of Realdo Colombo and Fabricius of Aquapendente. Colombo first portrayed the pulmonary transit of blood from the right ventricle of the heart to the left as a corollary of his main discovery stating that the *arteria venosa* solely contained blood and not air or sooty wastes as maintained by Galen's adherents. Moreover he also seemed to have insight in the proper movement of the heart, though his writings on this matter are not unambiguous. As French remarks, Colombo's main contribution to the works of Harvey may have been his demonstration that Galen's work was not devoid of mistakes and even more important that vivisection was a method to trace them.[9] Fabricius' influence on Harvey is double. First of all he incorporated Aristotle's biological works into

[7] Thought not all of them agree on what exactly triggered Harvey's discovery of the circulation as such, their works are most valuable and sufficiently in line to make a 'rough' reconstruction of Harvey's problem solving processes.

[8] Bylebyl (1977, pp. 143–144) and French (1994, p. 67).

[9] Bylebyl (1981, pp. 154–156), French (1994, pp. 82–83), Harvey (1616–1619/1961, p. 185/folio 77r), Pagel (1967, pp. 215–216), and Whitteridge (1971, pp. 41–77).

the natural philosophy program and as such formulated an anatomical Aristotelian project by ‘founding a theatrum of the whole animal fabric’. Hence, what had to be obtained was knowledge (in terms of universals) of what lies behind the particular appearances of observation. Accordingly, different structures point at different functions. However, in spite of this, variation can also be seen in the same organ when different animals are observed in view of the ‘form follows function’ dictum; in other words, in their different appearances, these organs can be well-suited to serve the same *final cause*. Secondly, Fabricius discovered the valves in the veins.[10] On the other hand, Fabricius and Colombo and their adherents, in a sense, also stayed loyal to Galen. Both the discovery of the pulmonary transit and the discovery of the valves in the veins were interpreted as compatible corrections to Galen’s physiology; the pulmonary transit solved the still ongoing problem of the pores in the interventricular septum allowing for the transit of blood from the right to the left ventricle and the valves in the veins were interpreted as mediators for a slow centrifugal flow of venal blood. So, in Harvey’s time it was possible to comment on Galen in spite of Galen’s doctrine still being predominant for interpreting human physiology.[11] This possibility for commenting on the work of Galen partly originated from a growing influence of physiological principles derived from Aristotle’s biological works that steadily caused changes in the interpretation of anatomy and physiology.

Summarizing some of Harvey’s theoretical background from which his initial problem arose, we obtain the following: (i) Harvey certainly was influenced by Galen’s physiolgy; (ii) sixteenth century commentaries and corrections on Galen’s doctrine became gradually accepted and allowed for further research; and (iii) Harvey took part in a revival of Aristotle’s biological program within natural philosophy as received from the Middle Ages. As a matter of fact, Harvey, at an early stage in his intellectual growth, adhered to the Aristotelian idea that the heart is the central organ in physiology and that it should be fully considered as a coherent sanguineous organ. Opposed to Aristotle’s view, he considered the blood as primary to the heart.[12] In fact, Harvey believed in the ‘primacy of the blood’, i.e. blood being the single homogeneous substance from which life originates as already stated in the ALN[13]:

> There is no other organ of contained blood [so] filled to capacity, wherefore Aristotle, contrary to the physicians, [states that] the origin of the blood is in the heart, not in the liver, because there is no extravenate blood in the liver. **WH** Blood is rather the origin of both, as I have seen. (Harvey 1616–1619/1961, p. 180/folio 75r)

[10] Bylebyl (1979, p. 71), French (1994, pp. 67–68), Pagel (1967, 1976, pp. 214–216|p. 13), and Whitteridge (1971, pp. 21–23).

[11] See, for instance, Harvey’s use of Galen’s claims on the attraction of nutrition on folios 23v, 24r, 43v and 44v in the ALN (Harvey 1616–1619/1961).

[12] Bylebyl (1979, p. 65), French (1994, p. 74), Pagel (1967, p. 43), and Whitteridge (1971, pp. 143–144).

[13] Despite this divergence from Aristotle, it still is in line with the latter’s monistic conception of living substance—(Bylebyl 1981, p. 152).

An investigation in the true movement of the heart and its corollary on the true nature of the pulse was not unusual. Especially because both the movement of the heart and the pulse could be felt in the intact body as much as in a vivisected one. Hence, it certainly were natural phenomena related to the normal state of the body that could be investigated. Moreover, despite Galen's view on the matter, the proper movement of the heart and arteries (pulse) and especially their interrelation were still a topic of discussion at Harvey's time. The method at hand was observation by vivisection on warm-blooded animals, a method that had already been described (used) by Galen in detail and successfully applied by Colombo, among others. Actually, for natural philosophers the methods of dissection and vivisection were a bit less problematic than they were for physicians who were extremely reluctant to draw conclusions from situations in which the normal state of the body has been violated. Still, it should be remembered that also Harvey was a trained physician who knew very well what influence an instigated flow of humors could have, but on the other hand, the movement of the heart was also observable in the normal functioning of the organism. The method failed however. Observing the fast beating heart of warm-blooded animals after vivisection did not generate the perspicuous observations needed to draw conclusions on what is the real active movement of the heart. As such, the original problem solving situation did not render a solution to the problem. Two separate phases could be discerned, but it was impossible to pinpoint one of these as 'the proper movement of the heart'. Harvey solved this (derived) observation problem through the observation of cold-blooded animals and dying hearts. From a medical stance, both were problematic; the dying heart (plus it being observed during vivisection) could hardly count for the normal situation and the hearts of cold-blooded animals diverged too much from the ones observed in warm-blooded animals. At this point, Harvey's adherence to a natural philosophy position, and not a medical one, allowed him to observe what was needed to draw conclusion on the proper movement of the heart.[14] For Galen information from animals decreased in interest the more they differed from man (French 1994, p. 84). Moreover, many considered the body of man as an image of God and consequently by rule different from the body of animals—see for example the positions of Du Laurens and Parigiano on this matter, (French 1994, pp. 40–42, pp. 228–234). Harvey, however, was enquiring the true nature of the heart, wherever it appeared—see Sect. 11.3.1 on the application of analogical reasoning at this stage of the problem solving process.

By means of the observations and considerations on the anatomy of the heart and vessels, Harvey, already in the ALN, came to the following conclusion—see also Sect. 11.3.2:

> Action: *thus relaxed receyves blood. Contracted propell it over*. In the whole body of the artery compares *as my breath in a glove*. (Harvey 1616–1619/1961, p. 190/folio 79v)

[14] Bylebyl (1973, pp. 434–439, 1979, p. 68), French (1994, pp. 74–75, 84), and Pagel (1967, pp. 28–47, 214–218).

As a matter of fact, this quote enlightens both the active movement of the heart and its proper action. Forceful systole is the active and powerful true action of the heart, diastole is the moment of rest after forceful systole during which the heart receives blood. The action of the heart is the propulsion of blood, first received in the right ventricle from the *vena cava inferior*, then sent through the lungs to the left ventricle and finally propelled in the arteries through the *aorta*. Though an approximately true perspective on the active movement of the heart and the pulmonary transit can also be found in the works of Colombo, it is Harvey who presents a coherent and systematized account on both phenomena. What is more, he drastically changes the action of the heart from a slow attraction of blood from the *vena cava inferior* during diastole to a incessant outward propulsion of blood into the arteries during forceful systole. Over and above that, he rejects the active dilatation of the arteries as causing of the pulse and reinterprets it as a passive mechanical consequence of the intrusion of blood into the arteries.

Harvey's new findings on the movement and action of the heart showed to be incompatible with Galen's distinction of the venal and arterial systems that respectively had the liver and the heart as central organs. How exactly this new problem was triggered is not important for the present discussion. Whether Harvey knew of Emilio Parigiano's quantification argument for the reflux of arterial blood to the heart (a result published in 1923) (Bylebyl 1979, pp. 76–77; 1981, p. 155), or whether it was by defending his new conception of the heart (French 1994, pp. 89–90), what is important is it being very likely that what is known as the famous 'quantitative argument' in Chap. 8 of DMC (Harvey 1628/1976, pp. 74–75; Bates 1992, p. 364) was in first instance a problem of inconsistency between Galen's conception of the cardio-vascular system and the new concept of the heart propagated by Harvey. Even a rough estimation of the quantity of blood that is propelled with every stroke of the heart and the accumulating effect of this propulsion of blood in one day could never be compensated by the part of the venous blood, produced in the liver, that was reserved for the heart. So, to safeguard his own theses on the forceful systole and the propulsive action of the heart, Harvey had to find a solution to the quantitative problem. His solution is very well known as the 'blood circulation'. There are different scenario's on how Harvey came to this solution. Though this is not the place to discuss this subject in detail, a short sketch of the position that is followed in this text is needed in view of Sect. 11.3.3. Gweneth Whitteridge, in the main, looks upon Harvey as a modern and considers Harvey's reflection and experiments on the valves in the veins as the main elements in the discovery process (Harvey 1628/1976, pp. xxvii–xl). Walter Pagel and Jerome Bylebyl emphasize Harvey's roots in natural philosophy and focus on the quantitative problem as the main trigger and the circular symbolism as an important clue in the construction of the solution (Pagel 1967, pp. 71–124, 1976, pp. 1–6, 14–23; Bylebyl 1979, p. 73–90). The influence of Paduan natural philosophy on Harvey cannot be denied and as Pagel suggests it is very plausible that Harvey's meditation on the valves in the veins is as much an investigation in the true nature of things as his meditations on the true nature of the heart (Pagel 1976, p. 4). In line with French (1994, pp. 85–93), I agree on the significance attributed to the quantitative problem by Pagel and Bylebyl and consider many of the triggers that may have contributed to the initiation of the search process—that after many experiments led to the solution—as very plausible; Sect. 11.3.3, in fact, focusses on one of these triggers. Furthermore, also in line with French, I believe that the valves in the veins have played a prominent role in the final and definite closing of the circle.

11.3 Harvey's Analogies

Before considering some of Harvey's applications of analogical reasoning, it is interesting to focus on one more detail of Harvey's mode of procedure, i.e. the 'rule of Socrates' *per similitudinem* (Harvey 1616–1619/1961, p. 27 folio 4r, p. 62 folio 20r), as highlighted by French in French (1994, pp. 83–85). According to French the 'rule of Socrates'—in fact also propagated by Galen—definitely refers to a method of enquiry proposed by Socrates in the *Republic* connoting "looking for the same thing in different contexts to see it more clearly" (French 1994, p. 85). The *per similitudinem* on the other hand refers to the Aristotelian project of searching for the similarity of function or the 'what it is to be a [in this case] heart'.

11.3.1 Extrapolation

As mentioned in Sect. 11.2, Harvey, just like anyone else, was unable to perspicuously observe the 'characteristic' movement of the heart during vivisecting warm-blooded animals. However, it ought to be noted that Colombo had been able to draw conclusions from vivisecting mammals, i.e. on the pulmonary transit as a corollary of his results on the *arteria venosa* and on the true motion of the heart (though the last not unambiguously). Harvey, on the other hand, claims both in the ALN (Harvey 1616–1619/1961, p. 185/folio 77r) and the DMC (Harvey 1628/1976, p. 29), finding it very hard to draw conclusions (or to confirm Colombo's theses) from these same observations. His Paduan background and the idea behind the 'rule of Socrates'—which he most probably learned from the work of Galen, instigated him to broaden his field of research and to overcome the problem of observation by vivisecting cold-blooded animals.[15] "All this is more evident in the heart of colder creatures, as toads, snakes, frogs, snails, lobsters, crustaceans, molluscs, shrimps and all manner of little fish" (Harvey 1628/1976, p. 32). In the ALN, he even focusses on the fish as in them the observation is most obvious, "In fish it is clearly compressed by extension and blood is given forth." (Harvey 1616–1619/1961, p. 186/folio 77v). This extension however presupposes that the hearts of animals show sufficient

[15] Bylebyl (1979, p. 34) and French (1994, pp. 84–85).

similarity with that of man. In the case of mammals there is a similar structure of the heart, in reptiles, amphibians and especially fish, this similarity is less evident and convincing. Harvey's observations of dying hearts were no better solution as a dying heart was not as such representative for the living heart.

Thus, Harvey was able to discern clearly the movement and the action of the heart in fish. Yet, this could not automatically lead to the construction of universals as the structure of the heart of fish is considerably different from that of man. However, it is highly probable that Harvey, presupposing that all hearts have the same function and as such should, notwithstanding any morphological differences, display analogous processes, extrapolated his findings from the vivisection of fish to other animals. Furthermore, to justify these extrapolations, he relied on analogies and even more, he explained why the differences in morphology of the hearts did not prevent that "all hearts served to eject blood in forceful systole and in so doing generate the pulse" (French 1994, p. 85). These claims are certainly supported by the following fragment from Chap. 6 from the DMC:

> First of all then, in fish which have but one ventricle of the heart, they having no lungs, the matter is clear enough. For it is certain that the bladder of blood set at the base of the heart and analogous to an auricle sends the blood into the heart, and that the heart then clearly sends on the blood again through the pipe or artery or vessel analogous to an artery; and this can be confirmed before our eyes either by looking or by the cutting of that artery from which the blood then leaps forth at every pulsation of the heart. Next, it is not difficult to see the same thing in all animals that have but one ventricle, or as it were but one, as toads, frogs, snakes and lizards, which although they are said in some manner to have lungs because they have a voice [. . .], yet it is plainly to be seen from actual inspection that in them the blood is transferred in the same way from the veins into the arteries by the pulsation of the heart [. . .]. For in these animals the case is as it might be in man were the septum of his heart perforated or taken away or one ventricle made out of the two; that done, I believe no man would then doubt by which way the blood could pass out of the veins into the arteries. [. . .] I have, moreover, considered in my own mind that the same thing is most clearly to be seen in the embryos of those animals that have lungs. (Harvey 1628/1976, pp. 56–57)

Though this fragment of course is a piece of argumentation in defense of the pulmonary transit, it seems to contain the gradual process Harvey himself ran through in order to sort out his observations. By claiming the bladder of blood and the pipe or artery or vessel in fish to be analogous to respectively an auricle and an artery, its clearly observable movement and action can be expected to occur in analogously functioning (by presupposition) but not identical structures. And this seems to be the case in other cold-blooded animals that have a more complex heart, and as not specifically mentioned in this fragment, also in the dying hearts of warm-blooded animals. Even more, Harvey not only takes the morphological differences according to different needs into account, he also focusses on the similarity of structures whenever the needs are similar, i.e., the *foramen ovale* and *ductus arteriosus* in the foetus of animals that have lungs but at that stage of development don't use them.

11.3.2 Glove

In the conclusion on the action of the heart as rendered in the ALN (see Sect. 11.2.4), Harvey applies the analogy of the glove with respect to the pulse. His conclusion on the true movement of the heart, i.e. the forceful systole, changed the attracting movement of the heart into a propulsive one.[16] In fact, Harvey had to conclude that blood was forced into the arteries and also in the *vena arteriosa*. But, in Galenic physiology, the *vena arteriosa*, though having the structure of an artery, was considered to be a vein as it contained venal blood to nourish the lungs. Harvey's thesis on the forceful systole and observations by vivisection on the timing of the pulse in the heart and in the arteries make him conclude that the pulse in the arteries is caused by the violent intrusion of blood during the diastole of the arteries. So, arteries do not dilate actively and only show a pulse for the mechanical reason of forcefully impelled blood. Consequently, on folio 78r (Harvey 1616–1619/1961, p. 186) Harvey inserted an enquiry whether the *vena arteriosa* pulsates and on folio 78v (Harvey 1616–1619/1961, p. 187) he inserted that it effectively does. In view of this result, Harvey concludes that the *vena arteriosa* and the *arteria venosa* respectively are an artery and a vein and not vice versa.

However, there still was a problem related to the pule, i.e. "There was more general, though not unanimous, agreement that the arteries pulsate actively, largely because it was thought that all of the arteries would not dilate simultaneously if the cause of the pulse were purely mechanical (Bylebyl 1981, p. 153)". Whether Harvey came across the analogy of the glove in the writings of Gabriele Falloppio (a teacher of Fabricius) as suggested by Bylebyl in Bylebyl (1981, p. 154) and got convinced of its efficacy or whether he conceived of it himself, it is a remarkable analogy that closes his argumentation on the true nature of the pulse. The analogy makes the simultaneous mechanical dilatation of the arteries fully conceivable and as such constitutes an important argument in Harvey's doctrine on the passive arteries that pulsate because of the forceful intrusion of blood.

11.3.3 Pulmonary Transit—Lesser Circulation

As mentioned before, the discovery of the blood circulation was most probably instigated by the quantitative problem that arose after the determination of the

[16] That Harvey envisaged the propulsive power of the heart as quite ferocious can also be inferred from his reinterpretation of the function of the valves in the veins (added in later version of the ALN) (Whitteridge 1971, p. xxix). "Hence neither the vena cava nor the pulmonary vein [is] of such structure, because they do not pulsate but rather [the blood] is drawn [from them]; and this because the opposed valves break the pulse in the heart and in the rest of the veins. **WH** Wherefore there are many valves in the veins opposed to the heart; the arteries have none except at the exit from the heart" (Harvey 1616–1619/1961, p. 191/folio 80r).

true movement and action of the heart and the pulse. In line with Pagel (1967, p. 54), I want to focus on an analogy that immediately follows the famous passage containing the quantitative argument and Harvey's own account on the discovery of the blood circulation in Chap. 8 of DMC (Harvey 1628/1976, p. 74–75).

> I began privately to think that it might rather have a certain movement, as it were, in a circle, which I afterwards found to be true, and that the blood is thrust out from the heart through the arteries, and driven forward into the habit of the body and to all parts, by the beat of the left ventricle of the heart, just as [the blood is thrust out] by the [the beat of the] right [ventricle] into the lungs through the arterial vein; and returns back again through the veins into the vena cava and up to the right auricle, just as [the blood returns] from the lungs through the so called venous artery to the left ventricle, as was previously said. (Bates 1992, p. 364)

In the light of Chaps. 6 and 7, focussing on the pulmonary transit of the blood, it certainly is, for reasons of argumentation, significant that Harvey refers to the pulmonary transit in order to defend the idea that an analogous passage of blood may start from the left ventricle of the heart throughout the body to the right ventricle of the heart. On the other hand, after the pulmonary transit seemed the most obvious path of the blood returning to the heart after its intrusion in the *vena arteriosa*, Harvey searched for further arguments to make the passage of the blood conceivable, and again arrives at the formulation of some analogies: "[. . .] when we consider how water passing through the substance of the earth brings forth rivers and springs, or observe how sweat passes through the skin, or how urine flows through the parenchyma of the kidneys. [. . .] The parenchyma of the liver and that of the kidneys likewise is denser by far than that of the lungs, which is of a much finer texture and spongy in comparison with that of kidneys and liver" (Harvey 1628/1976, pp. 66–67). It is very likely that the analogy of the situation with the pulmonary transit has contributed to Harvey's formulation of the blood circulation. Especially when considering the fact that the capillary transit of blood was still unknown to Harvey. Just as there was no immediate path from the *vena arteriosa* to the *arteria venosa* that had to account for the transit of blood through the lungs, there was none to be found that had to account for the transit of blood throughout the body.[17] For as long as there was no circulation, one can hardly speak of the lesser circulation of course, but as the pulmonary transit may certainly have contributed as a source of an analogous transit, it may be considered as the lesser circulation preceding and contributing to the conception of the full circulation.

[17] Regarding this passage of blood and the connection between the venal and arterial system, Harvey did not rely on the 'anastomoses' that formed a connection between the arterial and venal system in Erasistratean and Galenic physiology (among other reasons to account for the arterial haemorrhage Bylebyl 1979, pp. 46–51). In the DMC Harvey refers to the 'anastomoses' as a subject for enquiry in Chap. 9 (Harvey 1628/1976, p. 83) and in Chap. 11 (Harvey 1628/1976, p. 93) as a hypothesis derived from experimental observation.

11.3.4 The Heart as a Muscle and Some Concluding Remarks on the Use of Analogical Reasoning by Harvey

Considering Harvey's problem solving process leading to the determination of the movement and action of the heart, we may locate a mixture of elements from Galen's doctrine and especially Aristotle's natural philosophy combined with the 'Socratic rule'. Harvey's natural philosophy background allowed to overcome the problem that the method of vivisection of warm-blooded animals did not provide the relevant items to draw conclusions. Focussing on cold-blooded animals, Harvey was able to make more perspicuous observations that were relevant for the problem at hand. But as argued in Sect. 11.3.1, he had to rely on a kind of analogical reasoning, i.e. extrapolation, to draw similar conclusions for warm-blooded hearts that are morphologically considerably different. Somehow, Harvey's conviction in the existence of a true movement and action of all hearts, allowed him to extrapolate his findings in some type of hearts to draw conclusions on other types. As such, the hearts of fish and other cold-blooded animals served as a source domain to get a grasp on what was happening in the warm-blooded hearts that were morphologically different. From this, Harvey was able to draw conclusion on how differences in needs caused the differences in morphology while the movement and action of the hearts was the same. Therefore, Harvey had to apply defeasible reasoning allowing to transfer information observable in cold-blooded animals (the source domain) to warm-blooded animals (the target) in order to extend his relevant information from which the solution had to be derived. So, his certainties must in some way have contained ampliative rules that allow for these inference steps.

A last case of Harvey's use of analogies, in line with Sect. 11.3.1, should further support this claim. In Chap. 2 of DMC (and less systematic also in the ALN) Harvey describes three important characteristics of the moving heart: (i) the heart rises up, lifting upward into a point, (ii) it is contracted on all sides (especially to be seen in an eel, or little fish or other cold-blooded animals that have conical and rather elongated hearts and (iii) it feels harder. Moreover, in fish and cold-blooded animals like snakes and frogs, the heart becomes paler in color during its movement and richly-dyed blood-red when quiet.[18] From these characteristics Harvey concludes the following:

> [. . .] its movement was like that of the muscles when a contraction is being made [. . .] for when muscles are moving and in action, they gain strength and become tense, from soft they become hard, they are lifted up and thickened, and so likewise the heart. (Harvey 1628/1976, p. 33)

Harvey's observations, his knowledge of the contraction of muscles and the analogy drawn between the movement of the heart and the contraction of a muscle, allow him to suppose that the cavities of the heart become smaller during this action and consequently that blood is trusted out. This is further supported by the

[18] Harvey (1628/1976, pp. 32–33).

observation of the paler heart during movement and the richly-dyed blood-red color when the heart is quiet. So, the analogy with the muscle allows Harvey to conclude that the cavities of the heart must become smaller, which is further supported by an observation that can only be made in fish and other cold-blooded animals. Hence, analogical reasoning is applied to draw conclusions on the movement of the heart that certainly occurs in cold-blooded animals and extrapolation is used to draw the same conclusion for warm-blooded animals—see also Sect. 11.3.1.

Quite astonishing is the observation that the second part of DMC, in which the circulation is defended, contains way less applications of analogical reasoning. Moreover, these chapters show an abundance of experiments. In view of the fact that both parts deal with another problem and that the one from the second part is a derived problem from the new doctrine on the 'action of the heart' (and as such also came later in time), it would be interesting to investigate whether Harvey grew as a 'modern' during his discovery of the blood circulation. Moreover, this would be a basis to study the question whether there was a difference in reception between the doctrines that were defended in the two parts of DMC.

Acknowledgements The paper was written while Dagmar Provijn was a Postdoctoral Fellow of the Fund for Scientific Research Flanders. The author is indebted to the two referees for their valuable comments.

References

Batens, D. (1992a). Do we need a hierarchical model of science? In J. Earman (Ed.), *Inference, explanation, and other frustrations. Essays in the philosophy of science* (pp. 199–215). Berkeley/Los Angeles, Oxford: University of California Press.

Batens, D. (1992b). *Menselijke kennis. Pleidooi voor een bruikbare rationaliteit* (2nd ed., 2004; 3rd ed., 2008). Antwerpen/Apeldoorn: Garant.

Bates, D. G. (1992). Harvey's account of his "discovery". *Medical History, 36*, 361–378.

Bylebyl, J. J. (1973). The growth of Harvey's *De Motu Cordis*. *Bulletin of the History of Medicine, 47*(5), 427–470.

Bylebyl, J. J. (1977). "De Motu Cordis": Written in two stages? (response). *Bulletin of the History of Medicine, 51*(1), 140–150.

Bylebyl, J. J. (1979). The medical side of Harvey's discovery. In J. J. Bylebyl (Ed.), *William Harvey and his age. The professional and social context of the discovery of the circulation* (pp. 28–102). Baltimore/London: The Johns Hopkins University Press.

Bylebyl, J. J. (1981). Harvey, William. In C. Gillispie, J. Hachette, J. Hyrtl (Eds.), *Dictionary of scientific biography* (Vol. 6, pp. 150–162). New York: Charles Scribner's Sons.

French, R. (1994). *William Harvey's natural philosophy*. Cambridge: Cambridge University Press.

Gentner, D. (2003). Analogical reasoning, psychology of. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 106–112). London: Nature Publishing Group.

Harvey, W. (1616–1619/1961). *Lectures On the whole of anatomy*. Berkeley/Los Angeles: University of California Press. (Trans. with introduction and notes by C. D. O'Malley, F. N. L. Poynter, & K. F. Russell).

Harvey, W. (1628/1976). *An anatomical disputation concerning the movement of the heart and the blood in living creatures*. Oxford: Blackwell. (Trans. with introduction and notes by Gweneth Whitteridge).

Meheus, J. (2000). *Analogical reasoning in creative problem solving processes: Logico-philosophical perspectives*. In F. Hallyn (Ed.), *Metaphor and analogy in the sciences* (pp. 17–34). Dordrecht: Kluwer.

Nickles, T. (1981). What is a problem that we may solve it? *Synthese, 47*, 85–118.

Pagel, W. (1967). *William Harvey's biological ideas*. Basel/New York: S. Karger.

Pagel, W. (1976). *New light on william harvey*. Basel/New York: S. Karger.

Whitteridge, G. (1971). *William Harvey and the circulation of the blood*. Amsterdam: Elsevier.

Chapter 12
Another Look at Mathematical Style, as Inspired by Le Lionnais and the OuLiPo

Jean Paul Van Bendegem and Bart Van Kerkhove

12.1 Introduction

As has recently been observed by Mancosu (2010), the topic of 'mathematical style' is an underdeveloped one. In our contribution, next to briefly reviewing what (kinds of) approaches are already at hand, we add another perspective on the matter, inspired by a continental development, more particularly in France. As early as in 1948, in a paper in his edited volume *Les grands courants de la pensée mathématique* (only much later translated as *Great currents in mathematical thought*, see Le Lionnais 1971b), François Le Lionnais discussed the aesthetics of mathematics. Yet instead of focusing on expected issues such as symmetry, golden ratios and the like, Le Lionnais in it distinguished romantic versus classicist ideas of style in mathematics, corresponding roughly with an invention versus systematic exploration classification of mathematical domains.

The development referred to did not only involve Le Lionnais and, as it is the same period, Bourbaki and its structuralist program, but also mathematician-writers such as Raymond Queneau and Jacques Roubaud, founders, together with Georges Perec, of the OuLiPo (*Ouvroir de Littérature Potentielle*, or *Workshop for Potential Literature*). And although the Bourbaki approach clearly became the dominant one, this paper can also be read as a reassessment of the value and importance of that other side, especially since it tended to focus on the full complexity of mathematical and literary practice.

In Sect. 12.2 of this paper, we first briefly put the particular discussion into its wider context, sketching the contours of a new, or rather rejuvenated, direction in recent philosophy of mathematics. In Sect. 12.3 the notion of mathematical style is introduced as a good means of serving the purposes of that resurfacing perspective,

J.P. Van Bendegem (✉) • B. Van Kerkhove
Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Belgium
e-mail: jpvbende@vub.ac.be; bvkerkho@vub.ac.be

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_12,
© Springer Science+Business Media Dordrecht 2014

and an overview of different interpretations is given. In Sect. 12.4, we present our own suggestion for an alternative or complimentary look at mathematical style, as inspired by Le Lionnais, and in Sect. 12.5, we discuss the possibility for a comparison between mathematical and literary styles by looking at the above mentioned OuLiPo as a specific example. In the concluding Sect. 12.6 we recapitulate the main theses of the paper.[1]

12.2 Context of the Discussion

A growing part of contemporary philosophy of mathematics has been an attempt to restore an old tradition, one that was brutally interrupted in the twentieth century by the adoption of the ahistorical and systematic approach of then reigning analyticism. As Ferreirós and Gray have put it:

> Until the late nineteenth century it was customary to accept that studying the soil into which [the] foundations [of mathematics] penetrate was not itself a mathematical question, but rather a philosophical one, and the diagnosis about where the pillars end and the soil begins was very different from today. [. . .] By contrast, a very important ingredient in twentieth-century images of mathematics (though not omnipresent) has been the idea of a self-contained discipline, one that would account for its own foundations. (Ferreirós and Gray 2006, p. 3)

Without any exaggeration, this latter approach, commonly referred to as 'foundationalism', has virtually monopolized the philosophy of mathematics for most of the last century. Successful as these logico-mathematical approaches by the likes of Russell or Hilbert might have been, they have tended "to alienate both historians of mathematics and practicing mathematicians" (Ferreirós and Gray 2006, p. 10). Recent decades have however seen a gradual restoration of the tradition in question, thanks to the analytic approach having ceased to be monolithic (increasing its scope), and the history of mathematics having been professionalized (increasing its quality). As a result, some of the philosophical sensitivities hitherto referred to as 'maverick' have actually been absorbed in running discussions, and the fields of philosophy and history of mathematics are currently converging again. A forceful way of grasping what has happened in the course, is the giving away of the 'foundational filter', an implicit screening device so called by David Corfield, that used to reduce the themes of philosophical interest and as such had outlived itself. That is, way past the 'crisis' period of 1880 through 1930, where the priority of foundational studies and thus the preoccupation with the most elementary levels of mathematics was indisputably justified. Corfield aptly observes:

> Straight away, from simple inductive consideration, it should strike us as implausible that mathematicians dealing with number, function and space have produced nothing of philosophical significance in the past seventy years in view of their record over the previous three centuries. (Corfield 2003, pp. 7–8)

[1] We wish to thank the two anonymous referees for their penetrating comments, which allowed us to improve both clarity and coherence of the paper.

In view of this, an extension of the epistemological scope beyond foundational studies was clearly in order, exploring "the broader space in which questions about mathematical knowledge, its underpinnings, and its development are asked—including questions about mathematical practices" (Ferreirós and Gray 2006, p. 6). Indeed topics such as fruitfulness, evidence, visualization, diagrammatic reasoning, understanding, explanation, etc., topics bearing upon the conceptual and methodological dynamics of the discipline, and as such exhibiting an obvious historical dimension, could not be properly addressed by systematic philosophy, and were thus largely left unattended to Ferreirós and Gray (2006, pp. 10–11).

12.3 Conceptions of Mathematical Style

If one wants to properly address any of the topics listed at the end of Sect. 12.2, then it seems clear that intense interaction and even cooperation are required between the disciplines of philosophy and history of mathematics. One is inevitably reminded here of the beautiful words by Imre Lakatos, taken down in his renowned study *Proofs and refutations*:

> Formalism disconnects the history of mathematics from the philosophy of mathematics, since, according to the formalist conception of mathematics, there is no history of mathematics proper. [...] Under the present dominance of formalism, one is tempted to paraphrase Kant: the history of mathematics, lacking the guidance of philosophy, has become *blind*, while the philosophy of mathematics, turning its back on the most intriguing phenomena in the history of mathematics, has become *empty*. (Lakatos 1976, pp. 1–2, original emphasis)

From this, it seems clear that in order to arrive at a relevant general theory of mathematics (as it is done), philosophy and history urgently need to engage in a much more intense two way traffic and collaboration. Particularly, the exchange of conceptual tools (philosophy to history) and empirical studies (history to philosophy) should lead to both increasing their relevance. One specific proposed tool facilitating this kind of philosophically relevant historical inquiry is precisely that of 'style'.

Given our remarks on the development of twentieth century philosophy of mathematics, one might expect that philosophers of science were quicker to realize the importance of the notion of style as a methodological and unifying concept. A starting point no doubt is Ian Hacking's influential 1992 paper *'Style' for historians and philosophers* (Hacking 1992). There, Hacking indeed presents the notion of 'style' as an analytic tool for philosophers and historians of science to be shared, where he takes differing styles to denote particular methods of scientific reasoning (thinking, arguing, showing, explaining). In response to Alistair Crombie's sixfold division of styles to be identified in the history of scientific endeavor, viz. (1) postulation, (2) experimentation, (3) modeling, (4) taxonomy, (5) statistical analysis, and (6) genealogy, Hacking has made quite some inspiring general observations, e.g. as to style either or not being peculiar to specific man or epoch, or having generalizing versus mere personal meaning, and has tried to distillate from a variety of historical examples any necessary conditions for being a 'style' as specified. We shall here not be going into the general discussion but retain that, in terms of dimensional axes, any definition of (reasoning) 'style', whatever it may be, must involve the local vs. global, the individual vs. community and the diverse vs. universal distinctions. It is worth mentioning that, as for mathematics, although he does not get into much detail, Hacking points out that it too belongs among, not above or beyond, the sciences, so style is or should as much be an issue there. That is, to be more explicit: several kinds of style (e.g., as distinguished by Crombie) can play a potential role there, and not the strictest of mathematical reasonings on the basis of postulation only.

Browsing for literature on our particular topic, however, it appears that references are few, and come with big intervals as well as hardly any connection between them (Chevalley 1935; de Lorenzo 1971; Granger 1968). Only very recently, Mancosu (2010) has undertaken a preliminary attempt at a systematic treatment of the issue and the existing literature, to find, with a touch of irony, that indeed it "is not one of the canonical areas of investigation in philosophy of mathematics". Summarizing, in what can be found, there seem to be two main conceptions at work: an historiographical one, dealing with what one might call various forms of 'local' styles, and a methodological one, rather being occupied with 'universal' styles. Local styles are attributed by historical reconstructions to either persons (e.g., Bishop-style constructivism, Weierstrassian-style calculus), groups or schools (e.g., the Bourbaki, or unification programs such as the Erlangen or Langlands ones), or nations (e.g., Spengler's 'Western' style, or Bieberbach's 'Deutsche Mathematik'). The central idea here seems to be that style is subject to inherent fluctuations, which clearly de-emphasizes the universality of knowledge, while interpreted as universal in nature, on the contrary styles are given a more 'lasting' cognitive or epistemological depth. Actually, and presumably to be expected, most of the philosophical analyses fall within the latter category, with loads of terms having been proposed denoting different styles, often in opposing pairs: Euclidean (axiomatic, postulation), Cossic, operational, synthetic vs analytical, Cartesian (algebraic), vectorial, Platonic, poetic (!), geometric vs algebraic, archimedian, algorithmic, Baconian (experimental), formal vs informal, direct vs indirect, conceptual vs computational, conceptual vs problem-solving, hypothetical, taxonomical or classificatory, statistical, speculative vs rigorous, etc.

Apart from these definitional, dimensional and classificatory problems dealing with mathematical style, there is, of course, the deeper and tougher problem: is mathematical style something that cognitively contributes to the mathematical content (Hacking 1992), or is it just a mode of presentation among many others possible and thus a mere contextual residue (Granger 1968)? We do believe the former but at the same time we also believe that the distinctions and dimensional axes that have been mentioned are not necessarily the best ones to make that point. This negative belief of ours must of course be accompanied by a positive formulation and that will be the core thesis of this paper. All existing proposals so far, whether they talk about individuals, schools, or programs, tend to have a uniform look at what a mathematical style can be. Our suggestion is that even within the setting of one particular problem, to be solved by a specific individual mathematician, different styles can be at work. More precisely, we suggest that the notion of style should be linked to mathematical practices, seen as a problem-solving process.

There is first of all the simple and straightforward observation that a problem-solving process comes in different stages that require different methods and different ways of thinking and looking, and thus require, we claim, different styles. Through this connection mathematical style gets immediately related to problem-solving stages, and thus cognitively relevant, especially if it is the case that style considerations help to solve a problem or, at least, help to find a route towards its solution. In Sect. 12.4, we shall elaborate on this alternative perspective, invoking the work of Le Lionnais.

12.4 Another Conception of Mathematical Style

Partially inspired by a paper from François Le Lionnais, we propose here a first classification of different styles to be associated with stages of the problem-solving process. More precisely, we will propose three types: the discovery style, the generative style, and the proof style. Further on in this section, we will suggest an analogy, based on an overlapping contrast between what we propose to call the Columbus type of 'inventive' exploration (covering both the discovery and generative styles) and the Mercator type of 'systematic' exploration (covering the generative and the proof style).

- *Discovery style* we would hereby define as complexity-seeking, where one is in the business of formulating and starting to tackle new problems, covering new grounds, exploring 'uncharted' territory. Classic examples include the introduction of imaginary numbers, Cantor's theory of infinites or Riemann's hypothesis.
- Complementing this, *generative style* should be seen as solution-seeking, where fresh ideas are disseminated and further developed via notes, letters, bulletin boards, wiki's etc. A very nice example of the latter is the recent 'Polymath Project' coordinated by Timothy Gowers, i.e. a form of 'massively' collaborative and successful on line discussion on a new, elementary proof for the Density Hales-Jewett theorem, a central combinatorics result[2] (cfr. Gowers and Nielsen 2009).
- Contrasting with both these, *proof style* is explicitly unification-seeking. It is the best known and studied of mathematical styles, and also the most dominant one to date. Its perfect example is Bourbaki's setting of a format for producing mathematical texts.

[2] Since the time of writing of this paper, the 'Polymath Project' has grown into an extremely interesting phenomenon of its own, requiring the attention of philosophers, interested in mathematical practice. For a more recent state of affairs, see Allo et al. (2013).

Running ahead of an evaluation of this proposal, we wish to remark that one of its clearest advantages is that any discussion about the differences between micro- and macrolevels becomes a secondary issue. On the one hand, stylistic distinctions are made on the micro-level but this does not prevent on the other hand that some stylistic elements remain stable over longer periods of time. In this way it becomes possible to avoid an—in our view—false choice between aesthetical-rhetorical-contextual styles on the one hand and methodical-substantial qualities of particular styles on the other. Let us however first of all explain how we were inspired by Le Lionnais.

François Le Lionnais (1901–1984), in his paper *Beauty in mathematics* (Le Lionnais 1971a), distinguishes between two types of beauty in as many conceptions of mathematics, arriving at a fourfold classification: classical beauty in mathematical facts, romantic beauty in mathematical facts, classical beauty in mathematical methods, and romantic beauty in mathematical methods. Mathematical results, for Le Lionnais, have classical beauty when they impress by their sobriety, eliminating as many elements as possible foreign to the particular question they address. In contrast to this, romantic beauty in mathematics expresses eccentricity, appealing to notions that rather violently redirect inquiries at hand, and "in which the uninitiated are prone to see insane nightmares rather than the fruits of logical activity" (Le Lionnais 1971a, p. 132). As prototypes for these two categories, one might oppose the establishment of Euler's formula $e^{i\pi} + 1 = 0$ to the replacement of Euclid's parallel postulate by alternatives. Turning to mathematical methods, correspondingly, classical style refers to proceeding by moderate means to reach maximal effects. Good examples of the former are proof by recurrence, the positional notation, and geometrical unification through projection. Instead, romantic proof methods tend to complicate things, by approaching matters indirectly (e.g., proof by reductio) or ideosyncratically (e.g. Riemann's introduction of the zeta function), which often leads to confusion and discouragement among students, as it usually requires and exhibits genuine virtuosity.

However, it turns out that the notions thus developed by Le Lionnais are not really fit to serve any of the analyses of mathematical style as envisaged above (see Sect. 12.3), which explains why we insist that he was a source of inspiration. Our proposal is not a straightforward application of his approach because his notion of style is linked to the ways mathematics proceeds and develops and is thus of all times. Any mathematician, whether born in the sixteenth century or today, who explores new ideas is thereby a romantic, either topic-wise or method-wise. Any mathematician, whether born in the sixteenth century or today, that carefully brings together known results in a single, explanatory framework will be a classicist. Or, in other words, it is not clear how 'local' styles can be easily fitted in.

There is nevertheless a possibility to bring Le Lionnais' dichotomy closer to our three-fold proposal. Instead of the pair romantic-classicist, we propose the overlapping pair Columbus-Mercator:

- The *Columbus type* is the prototype of the explorer, who has (a) particular aim(s) (in Columbus' case, to find a shorter route to India), takes off into a high-risk undertaking, with a high degree of failure, but equipped with the best maps available, that need not necessarily be accurate and mutually compatible. To some extent, it appears that in these cases inaccuracy, if some way is found to use it in a creative manner, can be sometimes a help to discover new domains.
- The *Mercator type*, in contrast, need not leave his place of birth, as indeed was the case for Mercator. His aim concerns the maps themselves. Carefully scrutinizing all available information, solving inconsistencies between different maps (thus facing a similar task as Columbus, hence the indicated overlap), thinking about a general framework, makes it possible for this type of inquirer to produce an atlas, a map of all maps, usually with a high degree of accuracy.

For several reasons we do indeed believe that this contrasting pair is philosophically interesting, because ways in which explorers set about and cartographers drew their maps are truly culture-bound. To see this, think e.g. about the instruments explorers had available to them and about the mathematics of faithful projections in map making. To start elaborating the proposed analogy, let us take the Bourbaki project as a case in point.

To label what Bourbaki has undertaken as setting up a cartographical work, an atlas as it were, is a view that comes close, so we think, to their original intention. There is definitely no comparison with either Frege's two-volume *Grundgesetze der Arithmetik* or Russell's and Whitehead's three-volume *Principia Mathematica*. The term 'foundational' is not appropriate, rather the term 'encyclopedic' comes to mind (which, by the way, fits in nicely with the French tradition of the Encyclopédistes). We will not go into too much detail here, but quote André Weil, one of the founding fathers of Bourbaki, who in a paper on the future of mathematics famously stated: "If logic is the hygiene of the mathematician, it is not his source of food; the great problems furnish the daily bread on which he thrives" (Weil 1950).

In terms of the theme of this paper, Bourbaki's aim was (also) a stylistic one: if all of mathematics could be written in a single format or style, then comparability would be a lot easier and it would become much clearer how any piece of the mathematical building fits within the whole of the building. In so far as it is reasonable to attribute a structuralist view to the Bourbaki group, apart from the fact that they themselves made reference to it, one could claim that 'structuralist' is in no way synonymous with 'foundational'. Or, in architectural terms, it is about the design of the building, not its foundations. Although, of course, a good design will include some foundational considerations, witness the fact that the very first volume of the *Elements des mathématiques* was devoted to set theory. Note however that, in concord with Weil's quote, typically no reference is made to logical foundations. In the actual introduction the initial claim that "complete proofs are given" is relaxed and, in fact, reference is made to the importance of experience and the intuitions that result from it. In summary, what we are trying to claim here, is that, in principle, one should be able to recognize the Bourbaki style by simply looking at a page in one of the books.

There is another nice and important feature about the Columbus-Mercator distinction. The problem-solving process view tends to have an obvious, favoured direction: from the problem posed to its solution. It is less clear how the solution can generate new problems. Whereas in the other case, it is to be expected that an explorer returning home, will provide extremely useful information to the cartographer, not merely to describe previously unknown territory, but to make the necessary changes to existing maps.

Finally, an additional benefit of the problem solving process view is, we believe, that a comparison with literary styles becomes possible. This will allow us to introduce the third inspirational element of our presentation, next to Le Lionnais and the Bourbaki, namely, the OuLiPo, thus at the same time completing the (historical) picture.

12.5 From Mathematical to Literary Style, Through the OuLiPo

At first sight, it must seem a hopeless task to outline possible interactions and/or comparisons between mathematical style, an as yet not well described and understood concept, and literary style(s), where theories, views, opinions abound. Is this then not the worst possible strategy to support one's case? Indeed it would be, were it not for the fact that "bridge cases" do exist. Such bridges are formed by literary texts that are closely, if not directly inspired by mathematical techniques and thus these texts have stylistic properties that are somewhere in between literary and mathematical styles. Thus making it possible to make a transition from the one to the other. As we will show, in at least one case, related to proof style, there is a direct link to be found. But let us first present a few historical elements to situate the OuLiPo.

In 1960 a group of people consisting of mathematicians and writers (with a non-empty intersection) started the OuLiPo or the *Ouvroir de Littérature Potentielle*, the *Workshop of Potential Literature*. The two most famous founding members were the already often mentioned François Le Lionnais and mathematician-writer Raymond Queneau[3] (1903–1976). Among later adherents one should mention the

[3] It is not that well known that Queneau did in fact himself publish a number of mathematical papers. One among these carries the curious title: *Sur les conjectures fausses en théorie des nombres* (*On false conjectures in number theory*). Admitted, it is not a piece of mathematical research as such, but its topic is quite typical for Queneau, namely what conjectures did mathematicians initially believe to be true, but in the end turned out to be false. On the one hand, it is tempting to think that Queneau went along with the Bourbaki spirit: do no trust your intuitions, but accept only formalized proofs. However a different reading is that he felt challenged by such

Italian writer Italo Calvino (1923–1985), the enigmatic Georges Perec (1936–1982), co-founder of the Dada movement Marchel Duchamp (1887–1968), and the writer-mathematician Jacques Roubaud (1932). The OuLiPo has remained quite active ever since, as one can conclude from their website http://www.oulipo.net/ and from various paper publications (OuLiPo 1973, 1981, 2009). It has connections with the (in)famous Collège de 'Pataphysique, founded in 1948, to which Queneau also belonged, and the name of which is a reference to the work of the proto-Dadaist Alfred Jarry (1873–1907), best known for this plays featuring Ubu King, less known for his creation of Dr. Faustroll ('Faust drôle' or 'Funny Faust'). With the latter is associated the idea of 'Pataphysics, described in many different ways, one of which being "the systematic study of unique events", another one being that 'pataphysics stands to metaphysics as metaphysics stands to physics.

One of the main attractions of the OuLiPo is that they reflected explicitly about the discovery and generative aspects of the problem solving process, in this case the production or genesis of texts. In a few words, the basic idea of the OuLiPo was to produce literature in all its forms (novels, poems, essays, ...), using more or less systematic methods, rules and/or algorithms, and especially formal-mathematical tools.[4] Applying one or another of these schemes produces texts that are very unlikely to come up in the mind of a writer who is consciously thinking about what to write. Therefore the end products might or are even likely to have unexpected properties, something a writer can exploit in future work, in the very same way that the discovery and generative style in mathematical problem solving produce, e.g., new concepts that need not necessarily be directly useful in a proof but are nevertheless fruitful in the search for a proof.

Let us first briefly present a few examples what the concrete methods and tools are and, of course, what the possibly literary value of these products can be:

- *The $S + n$ method.* The object is to replace all nouns in any given text, using any dictionary. One simply looks up the substantive, searches for the n-th substantive that follows in the dictionary and replaces the former by the latter. The result is very often quite humorous, as is shown by the following mathematically inspired example[5]:

  1. A straight linguist selling can be drawn joining any two poles.
  2. Any straight linguist selling can be extended indefinitely in a straight linguist.

results. It proves that the mathematical building, to remain within the architectural metaphor, contains far more surprising elements than we imagine. Or, let us put it this way: whereas Bourbaki is in the business of restructuring the entire building, Queneau enjoys it as it is, further exploring one amazing room after another.

[4] There is an additional element that we will not discuss here, namely that the formal methods should be unrelated to the content. In that sense, there is not really a connection with the "écriture automatique" of the surrealists, where the method is seen as specific for the content generated, namely, the "secrets" of the unconscious.

[5] This example has been computer-generated, using the free facilities on the website http://www.spub.co.uk/n+7/.

3. Given any straight linguist selling, a circumstance can be drawn having the selling as rage and one endpoint as center.
4. All right annotations are congruent.
5. If two linguists are drawn which intersect a third in such a weapon that the sump of the inner annotations on one sideswipe is less than two right annotations, then the two linguists inevitably must intersect each other on that sideswipe if extended far enough. This pothole is escalator to what is known as the paraphrase pothole.

- *The art of combination, variation and permutation*. The most famous example is a collection of poems by Raymond Queneau, *Cent Mille Milliard de Poèmes*, see Queneau (1965). The book counts no more than 10 pages, each page containing just one sonnet, consisting (as it should) of 14 lines. However, the 10 pages are not single pieces of paper, but have been cut-up in such a way that every line of each sonnet can be turned separately. Thus every line of every sonnet can be combined with any other line to form new sonnets, in total 10^{14}, i.e., a hundred thousand billion or a hundred million million poems, as the French and English titles indicate.[6]
- *The non-use of letters or 'lipogram'*. Surely, the most famous example of this particular method is the novel-thriller *La disparition* by Georges Perec (1961). The amazing feature of the book is that the vowel 'e' is missing. Even more amazing is that the book has been translated in English[7] and recently also our own mother tongue, Dutch, the respective titles being *A Void* (1994) and *'t Manco* (2009).

Any mathematician will recognize that the techniques that have been used here as literary devices, are familiar in mathematics as well. The first corresponds to mappings between domains to transplant a structure in the other, the second

[6] As one might expect, several websites offer the possibility to explore this universe of poems electronically, both in French and in English (see, e.g., http://www.bevrowe.info/Poems/QueneauRandom.htm).

[7] We include a short excerpt from the English translation, in order for the reader to appreciate the complexity of the task (Perec 1961, p. 116): "Noon rings out. A wasp, making an ominous sound, a sound akin to a klaxon or a tocsin, its about. Augustus, who has had a bad night, sits up blinking and purblind. Oh what was that word (is his thought) that ran through my brain all night, that idiotic word that, hard as I'd try to pun it down, was always just an inch or two out of my grasp—fowl or foul or Vow or Voyal?—a word which, by association, brought into play an incongruous mass and magma of nouns, idioms, slogans and sayings, a confusing, amorphous outpouring which I sought in vain to control or turn of but which wound around my mind a whirlwind of a cord, a whiplash of a cord, a cord that would split again and again, would knit again and again, of words without communication or any possibility of combination, words without pronunciation, signification or transcription but out of which, notwithstanding, was brought forth a flux, a continuous, compact and lucid flow: an intuition, a vacillating frisson of illumination as if caught in a ash of lightning or in a mist abruptly rising to unshroud an obvious sign—but a sign, alas, that would last an instant only to vanish for good."

one seems obvious enough and the third one corresponds to (a form of) reverse mathematics, where one tries to show that a result can be obtained without using a particular axiom or concept.

In addition, there is also an important connection to be made on the level of proof style. Queneau published in 1947, a little book, entitled *Exercices de style* (Queneau 1947). The book contains 99 short stories, all recounting the same rather banal incident, but doing so in 99 different ways or styles (hence the title). It is of course extremely tempting to transfer this idea to mathematics itself. Take any theorem, write a proof for that theorem and think up 99 variations on that proof, 99 different styles to present the proof. One will surely remark that this is not at all new and, in fact, it is an important ingredient of mathematical practice, viz., the idea that different proofs for the same theorem strengthen our belief about the correctness of the proofs. Cannot Pythagoras' theorem boost some 300 different proofs? And to a lesser extent, does not the same go for the irrationality of $\sqrt{2}$ and the prime number theorem? We fully agree, and in fact this immediately shows the importance of stylistic considerations directly affecting mathematical practice itself. However, one might also read the assignment in a slightly different fashion. Given a proof, and given the specific concepts, ideas and methods involved in it, is it possible to generate different versions that will then differ exclusively in terms of style? We assume that readers of the proof will not remain neutral as to the quality of such variations, which provides another way of showing the importance of style.

Here is such an example. Three proofs for the same theorem. Note that the proofs are similar in the sense that they use the same proof strategy. It really boils down to the formulation or, if you like, the rhetorics of the proof.[8]

Theorem. *The sum of two primes, larger than* 2*, is never a prime.*

Proof. A prime larger than 2 is necessarily odd, thus of the form $2k+1$. Adding two such numbers, say $2k+1$ and $2k'+1$ produces a number of the form $2(k+k')+2$ which is obviously even. □

Proof. Odd and odd is even. □

Proof. Let p_1 and p_2 be two primes, larger than 2. Then $p_1 \equiv 1 \pmod 2$ and $p_2 \equiv 1 \pmod 2$, but $p_1 + p_2 \equiv 0 \pmod 2$ and so cannot be a prime. □

It is of course still a long stretch from this particular case to a general comparison between mathematical and literary styles, let alone to show that such comparison will be productive or intellectually stimulating. We do think however that we have shown that the enterprise is not a futile one and needs to be continued.

[8] One of us has published on this particular topic, see Van Bendegem (2008), and the two of us have published on the wider theme of argumentation and proofs, see Van Kerkhove and Van Bendegem (2009).

12.6 Conclusion

This paper has been a plea for a less monolithic interpretation of the style concept, in order for it to serve well as a methodological tool in the historiography of mathematics. Drawing inspiration from Le Lionnais, the Bourbaki movement and the French literary-mathematical OuLiPo movement, we have introduced an approach along the path of a 'problem solving' conception of mathematics, thus creating room for mathematical style to be significantly 'more' than a mere mode of presentation of immutable content, while obviously not claiming universality. No doubt our boldest claim is that this very same approach opens us the possibility for a fruitful comparison between mathematical and literary styles.

Acknowledgements The authors are affiliated to the Centre for Logic and Philosophy of Science at Vrije Universiteit Brussel (Belgium). Support for this paper was granted by the projects GOA49 and FWOAL525 of their university. The latter author is also indebted to the Alexander von Humboldt Foundation (Germany).

References

Allo, P., Van Kerkhove, B., & Van Bendegem, J. P. (2013). Mathematical arguments and distributed knowledge. In A. Aberdein & I. Dove (Eds.), *The argument of mathematics* (pp. 339–360). New York: Springer.

Chevalley, C. (1935). Variations du style mathémathique. *Revue de Métaphysique et de Morale, 3*, 375–384.

Corfield, D. (2003). *Towards a philosophy of real mathematics*. Cambridge/New York: Cambridge University Press.

de Lorenzo, J. (1971). *Introducción al estilo matematico*. Madrid: Editorial Tecnos.

Ferreirós, J., & Gray, J. (Eds.). (2006). *The architecture of modern mathematics. Essays in history and philosophy*. Oxford/New York: Oxford University Press.

Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. *Nature, 461*, 879–881.

Granger, G. G. (1968). *Essai d'une philosophie du style*. Paris: Librairie Armand Colin.

Hacking, I. (1992). 'Style' for historians and philosophers. *Studies in History and Philosophy of Science, 23*(1), 1–20.

Lakatos, I. (1976). *Proofs and refutations*. Cambridge/New York: Cambridge University Press.

Le Lionnais, F. (1971a). Beauty in mathematics. In F. Le Lionnais (Ed.), *Great currents of mathematical thought* (pp. 121–158). New York: Dover. Edited volume originally published in 1948 as *Les Grands Courants de la Pensée Mathématique*, Cahiers du Sud.

Le Lionnais, F. (Ed.). (1971b). *Great currents of mathematical thought*. New York: Dover. Edited volume originally published in 1948 as *Les Grands Courants de la Pensée Mathématique*, Cahiers du Sud.

Mancosu, P. (2010). Mathematical style. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Spring 2010 ed.). Http://plato.stanford.edu/entries/mathematical-style/.

OuLiPo. (1973). *La Littérature Potentielle (Créations, Re-créations, Recréations)*. Paris: Gallimard.

OuLiPo. (1981). *Atlas de Littérature Potentielle*. Paris: Gallimard.

OuLiPo. (2009). *Anthologie de l'OuLiPo*. Paris: Gallimard.

Perec, G. (1961). *La Disparition*. Paris: Gallimard. (English Trans. by Gilbert Adair: *A Void*, Harvill Press, 1994; Dutch Trans. by Guido van de Wiel: *'t Manco*, De Arbeiderspers, 2009).

Queneau, R. (1947). *Exercices de style*. Paris: Gallimard.

Queneau, R. (1965). *Cent Mille Millard de Poèmes*. Paris: Gallimard.

Van Bendegem, J. P. (2008). Elements for a rhetoric of mathematics: How proofs can be convincing. In C. Dégremont, L. Keiff, & H. Rückert (Eds.), *Dialogues, logics and other strange things: Essays in honour of Shahid Rahman* (pp. 437–454). London: King's College Publications.

Van Kerkhove, B., & Van Bendegem, J. P. (2009). Mathematical arguments in context. *Foundations of Science, 14*(1–2), 45–57.

Weil, A. (1950). The future of mathematics. *The American Mathematical Monthly, 57*(5), 295–306.

Chapter 13
Internalism Does Entail Scepticism

Jan Willem Wieland

13.1 Introduction

Consider this simple argument:

(1) Socrates is mortal, or he is unfindable.
(2) It is not so that he is unfindable.
(3) So, Socrates is mortal.

The upcoming discussion basically turns on the question: how is it that we are justified in believing things thanks to the fact that we know other things? Here are three positions:

Internalism.	You are justified in drawing (3) only if you know that the inference is valid by this or that logical rule.
Externalism.	You are justified in drawing (3) only if the inference is valid by this or that logical rule, and knowledge of this is not required.
Scepticism.	You are not justified in drawing (3).

Internalism and Externalism exclude one another, yet neither excludes Scepticism. Moreover, Boghossian (2001) has shown that Internalism as formulated entails Scepticism. Philie (2007) takes up this debate, and presents a proposal to block the entailment. Specifically, he invokes an assumption on rationality that is to do the trick. That Internalism might well entail Scepticism should be a worry for all non-sceptical Internalists, and interesting for all non-sceptical Externalists. For if Internalism is committed to Scepticism and the latter unacceptable, then this favours Externalism. Hence, the debate by Boghossian (2001), Philie (2007) and others is worth considering.

J.W. Wieland (✉)
VU University Amsterdam, Amsterdam, Netherlands
e-mail: jjwwieland@gmail.com

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_13,
© Springer Science+Business Media Dordrecht 2014

I will proceed as follows. In the first part, I summarize Philie's argument (Sect. 13.2). In the second part, I first enforce the entailment of Scepticism by Internalism, and then cast doubt on Philie's attempt to block the entailment (Sects. 13.3 and 13.4).

Some notes before turning to Philie's argument. First, this debate is not about whether we *in fact* draw (3) with or without logical knowledge. Some do, many do not, but that is not the point (cf. Philie 2007, p. 186). Instead, the debate is about justification and in particular the conditions under which our inferences are justified. To be sure, the more familiar Internalism/Externalism debate is on when our beliefs, rather than inferences, are justified. I am well aware that important further qualifications can be made regarding the three positions (for example, in the case of Externalism one might add the requirement that the inference is based on a reliable method). I will assume, though, that this will do for the following.

Lastly, in this paper I will restrict myself to the Scepticism objection, and put any other argument for or against Internalism and Externalism aside. Therefore, if Internalism does indeed entail Scepticism (as I will argue), then this should not save Externalism from its own worries.

13.2 Philie's Argument

In the following I summarize Philie's argument: (a) Internalism entails Scepticism, yet (b) a more sophisticated version of Internalism does not. Step (a) derives from Boghossian (Boghossian 2001, pp. 25–26; cf. also Wright in 2001, pp. 71–80), and (b) is Philie's follow-up on the discussion between Boghossian and Wright. I will refer to the positions of the latter authors in Sect. 13.4. Please note that I will not take issue with one of Philie's steps until Sect. 13.3.

13.2.1 Step (a)

Consider again:

(1) Socrates is mortal, or he is unfindable.
(2) It is not so that he is unfindable.
(3) So, Socrates is mortal.

According to Internalism, a justification for (1) and (2) and the step to (3) is insufficient for being justified in concluding to (3). What is required, next to this, is that you *know* that (3) follows from (1) and (2) by a rule which is valid, namely, that you know that the argument displays a valid logical form. In terms of our example, you should have knowledge of Disjunctive Syllogism (or something equivalent), that it is a valid rule (at least in classical contexts; I shall not repeat this qualification), and that (3) follows from (1) and (2) by this rule. Hence, you have to make the following inference in order to be justified in drawing (3)[1]:

(4) If an inference is based on Disjunctive Syllogism, it is valid.
(5) The inference of (3) from (1) and (2) based on Disjunctive Syllogism.
(6) So, the inference of (3) from (1) and (2) is valid.

Yet the step can be repeated. According to Internalism, to be justified in your inference of (6) from (4) and (5) you have to know that it is valid by this or that rule, and make at least the following inference:

(7) If an inference is based on Modus Ponens, it is valid.
(8) The inference of (6) from (4) and (5) is based on Modus Ponens.
(9) So, the inference of (6) from (4) and (5) is valid.

Indeed, the step can be repeated again. According to Internalism, to be justified in your inference of (9) from (7) and (8) you have to make at least the following inference:

(10) If an inference is based on Modus Ponens, it is valid.
(11) The inference of (9) from (7) and (8) is based on Modus Ponens.
(12) So, the inference of (9) from (7) and (8) is valid.

And so on. Internalism entails that we have to make an infinity of inferences in order to be justified in drawing (3). However, as we do not make so many inferences, we are not justified in drawing (3). As the same reasoning can be repeated for whatever inference, no matter how simple or complex, it follows that none of our inferences are justified. Hence: Internalism entails Scepticism.

In contrast, Externalism does not entail Scepticism because it does not state that we should know that our inferences are valid in order to be justified in our inferences. All that matters for Externalism is that the inferences are valid by this or that rule, irrespective of one's knowledge of this. In this sense, for Externalism the justification of an inference is no internal affair.

13.2.2 Step (b)

So far all inferences are treated on the same footing. If you are justified in your inference of (3) from (1) and (2) by making a further inference, then you are justified in your inference of (6) from (4) and (5) by making a further inference as well. Yet, if this uniformity policy is violated, then the regress could be stopped. Although Philie does not state it in these terms, this is precisely what he proposes.

[1] For those who suspect that the validity of the inference of (3) can be seen directly, that is, not via (4)–(6), see (***) in Sect. 13.4 below.

The basic idea is to make a distinction between two kinds of inferences. The inference of (3) from (1) and (2) is based on Disjunctive Syllogism and this logical form could well vary for other such inferences (call them ordinary or *first-order*). This is different for the inferences (6), (9), (12) and other such inferences (call them *higher-order*, as they are about first-order inferences), which all have the following form (cf. Philie 2007, p. 201):

- If an inference is based on a rule X, then it is valid (according to such and such a logic).
- Inference Y is based on X.
- So, Y is valid (according to such and such a logic).

So, all higher-order inferences of this form are based on at least Modus Ponens (or something equivalent, cf. Philie 2007, pp. 201–203). Furthermore, the proposal is that Internalism (i.e. the thesis that the justification of our inferences is sensitive to our knowledge of the relevant rules) applies to first-order, but not to higher-order inferences. If this is so, the regress stops at (6), and inferences such as (7)–(9) are no longer required.

Yet, this strategy is a no-go unless the distinction is properly motivated. Why should Internalism apply to first-order inferences and not to higher-order ones? Or in other words: why bother about inferences based on whatever rule, while higher-order inferences employing Modus Ponens are not to be justified by further inferences? To answer this query, Philie invokes an assumption on rationality plus the distinction between constitutive and non-constitutive rules. Let us look at the distinction first. The following principle seems to be assumed in the present debate:

Constitution. For any rule x and any practice y, x is constitutive for y only if y does not exist unless x is obeyed, and non-constitutive for y otherwise.

For example, the rule ‘whenever you write a philosophy paper, you should report extraordinary thoughts arrived at by non-empirical means’ is constitutive for writing a philosophy paper, for if you do not obey this rule you are just writing something else. By contrast, ‘whenever you write a philosophy paper, you should remain faithful to what Kant taught us’ is non-constitutive, because if you do not obey this rule you can still write a philosophy paper (or so one would like to think). At any rate, Philie exploits this distinction as follows:

> Not obeying [Modus Ponens] would mean that you are not engaging in inferential practice, i.e. that you do not reason in such a way so that your thoughts follow a logical pattern. It is irrational to engage in inferential practice without it. (Philie 2007, p. 206)

The idea is that Modus Ponens, rather than any other rule (not equivalent to it), is constitutive for ‘inferential practice’. Taken charitably, this does not mean that you cannot make any inference whatsoever without Modus Ponens (because you obviously can), but that you cannot engage in the practice of justifying ordinary inferences without obeying it.

As Modus Ponens is not usually stated in terms of obligations or permissions (but in terms of logical form, validity, or truth-preservation), to speak of 'obeying Modus Ponens' might sound somewhat odd. What seems to be meant, though, is that one obeys the rule from A and [B if A], you are allowed to infer B, but not allowed to infer not-B' (the qualification after 'but' is needed in order to be able to violate the rule).

As Philie notes (Philie 2007, p. 203, 206), this point even holds for logics where Modus Ponens is not valid unrestrictedly. For in that case, inferences are to be justified along the following lines: (i) if an inference is based on non-classical rule R, then it is valid (according to such and such non-classical logics XYZ); (ii) inference such and such is based on R; so, by Modus Ponens, (iii) inference such and such is valid (according to XYZ).

This is different for other inference rules. For example, Simplification or Addition are not constitutive for the practice of justifying inferences, because not obeying them does not entail that you cannot justify any ordinary inference.

From the text just cited it can be seen that Philie takes one further step where the assumption on rationality comes in (he takes his inspiration in this from Davidson's take on Akrasia, I will discuss this in Sect. 13.4):

Rationality. For any practice x and any of its constitutive rules y, you are rational with respect to x only if you obey y, and irrational with respect to x otherwise.

According to this, you are irrational if you violate Modus Ponens in the higher-order cases where you want to justify inferences like (3). If this is right, then the difference between first-order and higher-order inferences is motivated and might well be used to stop the regress and hence the entailment from Internalism to Scepticism.

So far, so good. In Sects. 13.3 and 13.4 I put both steps of Philie's argument to scrutiny.

13.3 Step (a): The Entailment of Scepticism

In the following I look at step (a) and argue that this step does not work unless a hidden assumption is explicitly motivated. In particular, I would like to take issue with Philie's step from regress to conclusion:

> It seems that this pattern could be repeated indefinitely; if so, it means that claiming knowledge of [(7)] when drawing inferences leads to a regress of warrants. This in turn would mean that knowledge of the validity of the Modus Ponens schema cannot act as a warrant for the conclusion [(3)]. (Philie 2007, pp. 185–186)

So the question is: why does it follow from the potential regress ("this pattern could be repeated indefinitely") that knowledge of Disjunctive Syllogism cannot be used to justify the inference of (3) from (1) and (2)? Or in other words: why does Internalism lead, via the regress, to Scepticism? My claim in the following is that Philie's regress argument is inconclusive as some hidden assumption is missing. Namely, this assumption is that you are justified in drawing (3) only if:

(i) You *draw* the additional inference that (3) is based on a valid rule, and
(ii) You *are justified in drawing* this additional inference that (3) is based on a valid rule.

Before I turn to the crucial, second clause, I would like to compare this with the initial case by Lewis Carroll (1895).[2] In Carroll's dialogue Achilles and the Tortoise consider the following simple argument:

(A) Things that are equal to the same are equal to each other.
(B) The two sides of this Triangle are things that are equal to the same.
(Z) The two sides of this Triangle are equal to each other.

As the story goes, the Tortoise is willing to accept (A) and (B), but not (Z) just because she denies that (Z) must be accepted if (A) and (B) are accepted. So to demonstrate that (Z) follows from (A) and (B), Achilles adds an extra premise to the argument:

(C) If (A) and (B) are true, (Z) must be true.

Still, the Tortoise is unsatisfied. This time she is willing to accept (A), (B) *and* (C), but not (Z) just because she denies that (Z) must be accepted if (A), (B) and (C) are accepted. So to demonstrate that (Z) follows from (A), (B) and (C), Achilles adds yet another premise to the argument:

(D) If (A), (B) and (C) are true, (Z) must be true.

And so on. The assumption here is that Achilles demonstrates that the Tortoise is forced to accept (Z) given the truth of (A) and (B) only if

(iii) Achilles adds the premise (C) 'if (A) and (B) are true, (Z) must be true' to the argument, and
(iv) Achilles demonstrates that the Tortoise is forced to accept (Z) given the truth of (A), (B) and (C).

Both clauses are needed for Carroll's argument to work. If (iv) would not hold, then even if Achilles comes up with (C), nothing follows. For in that case Achilles need not demonstrate that the Tortoise is forced to accept (Z) given the truth of (A), (B) and (C) in order to demonstrate that the Tortoise is forced to accept (Z) given the truth of (A) and (B), and so Achilles need not write down further premises in his notebook either.[3] Of course, in that case Achilles *may* demonstrate that the Tortoise

[2] At several points, Philie suggests he is discussing Carroll's argument itself, yet see Philie (2007, fn. 3) for a qualification on this. Indeed, it will become clear in a minute that there are important differences.

[3] Throughout the paper, I regard steps of the following form as uncontroversial: if you φ only if you ψ, then you have to ψ in order to φ.

is forced to accept (Z) given the truth of (A), (B) and (C) (and furthermore that the Tortoise is forced to accept (Z) given the truth of (A), (B), (C), (D), etc.), but it would be irrelevant to the initial case with (A), (B) and (Z).

The point carries over to Philie's case. If (ii) would not hold, then even if you draw the additional inference that (3) is based on a valid rule, nothing follows. For in that case you need not be justified in drawing the additional inference that (3) is based on a valid rule in order to be justified in drawing (3), and so you need not draw further inferences either. Again, you may be justified in drawing the additional inference that (3) is based on a valid rule, viz. line (6) (and furthermore in drawing the additional inference that (6) is based on a valid rule, etc.), yet it would be irrelevant to the initial inference of (3) from (1) and (2).

It is instructive to invoke further cases as the familiar Regress of Reasons. This regress is generated provided that it is assumed that you are justified in believing a proposition p only if (v) you appeal to a reason q for p, and (vi) you are justified in believing q. Without (vi) you need not be justified in believing q in order to be justified in believing p, and so you need not appeal to further reasons either. For all these cases, the general point is that *both* clauses are needed because they guarantee that a problematic regress is actually entailed (and not merely optional and irrelevant).[4]

Hence, the important issue is whether Internalism is committed to both (i) and (ii). (i) follows directly from Internalism. To recall, Internalism is the view that you are justified in drawing (3) only if you know that the inference is valid, and how to know this if you do not infer that (3) is valid by this or that rule? Matters are different for (ii) though: are you justified in drawing (3) only if you are justified in drawing the additional inference that (3) is based on a valid rule? Unfortunately, Philie does not argue for this, which means that so far the argument against Internalism is inconclusive (and consequently any defence from the side of the latter unnecessary).

Let us evaluate this point for other cases first, starting with the Regress of Reasons. So the situation is that you appeal to a reason q for p in order to be justified in believing p. For example, you appeal to the fact that Socrates is a man in order to be justified in believing that Socrates is mortal. The question is: why should you be justified in believing that Socrates is a man in this scenario? Well, suppose you are not. This would mean that your justification for the belief that Socrates is mortal is arbitrary, and hence it is no proper justification (for further discussion on this, cf. Klein 1999). If this is right, then if you have a reason q for p, you are justified in believing p only if you are justified in believing q.

Next consider Carroll's case. The situation is that Achilles appeals to (C) in order to demonstrate that the Tortoise is forced to accept (Z) given the truth of (A) and (B). Then why should Achilles demonstrate that the Tortoise is forced to accept (Z) given the truth of (A), (B) *and* (C) as well? Suppose he does not. A common reaction to Carroll's Regress is precisely that this should be all right. All Achilles has to do,

[4] For details on how regresses can be generated and how conclusions can be drawn from them, see Wieland (2014).

it goes, is to appeal to something like (C) which licenses the step to (Z) (be it not as an extra premise, but rather as a story on inference rules, and that would be it.[5] Of course, Achilles may demonstrate that (Z) follows from (A), (B) and (C) as well, but that would be just another matter. As Thomson puts it: "Why should this procedure [of requiring demonstrations by Achilles] be adopted in the first place?" (Thomson 1960, p. 96).

So in some cases the extra assumption (i.e. second clause) holds, in others it fails, but in many it is controversial (and in *all* it should explicitly be treated, I would say, as a point of controversy). Now how about Philie's case? To recall, the scenario was the following:

(1) Socrates is mortal, or he is unfindable.
(2) It is not so that he is unfindable.
(3) So, Socrates is mortal.

(4) If an inference is based on Disjunctive Syllogism, it is valid.
(5) The inference of (3) from (1) and (2) is based on Disjunctive Syllogism.
(6) So, the inference of (3) from (1) and (2) is valid.

The crucial question here is: should you be justified in drawing (6) in order to be justified in drawing (3)? Well, suppose you are not. On first sight, this looks unproblematic. The step to (6) is completely simple, as it relies on Modus Ponens plus indeed Universal Instantiation, but nothing else (and this is the same for each and every higher-order case). What matters is that you take this simple step to (6), though not that you are also justified in this. That is just a further issue.

Here is an argument that it does matter. If you are not justified in the inference step, then you are not justified in concluding to (6) (because, as all parties would agree, you need a justification for both the premises and the inference step). If you are not justified in concluding to (6), you are not justified in believing its content, i.e. that the inference of (3) from (1) and (2) is valid. By the widely accepted knowledge-requires-justified-belief assumption,[6] if you are not justified in believing that the inference of (3) from (1) and (2) is valid, you do not know that this inference is valid. According to Internalism, then, you are not justified in drawing (3). So, as it turns out, if you are not justified in drawing (6), then you are not justified in drawing (3). Or in other words, you are justified in drawing (3) only if you infer (6) and are justified in this: the missing assumption we were looking for.

Of course, the same reasoning holds for (6), (9), etc. Furthermore, if we take all these conditions together, then you are justified in concluding that Socrates is mortal only if [you are justified in believing (1) and (2) and in drawing (3) from them; you are justified in believing (4) and (5) and in drawing (6) from them; you are justified

[5] See Ryle (1950), Thomson (1960), Stroud (1979), and Smiley (1995).

[6] Standardly, beliefs need be true and degettierized as well. For present purposes, it is only needed to assume that knowledge requires justified belief. Indeed, a possibility to resist the argument is to defend that knowledge does not require this. This move may look implausible, but it is possible for Internalism as defined.

in believing (7) and (8) and in drawing (9) from them; and son on]. Now, as we do not make so many inferences, we are not justified in concluding that Socrates is mortal. At this point, the entailment from Internalism to Scepticism, which was not yet established by Philie's reconstruction of the argument, stands.

Let me conclude step (a) with two notes. First, and interestingly so, the analogy with Carroll's case breaks down if the crucial second clause holds in Philie's case (i.e. line (ii) above), but fails in Carroll's (i.e. (iv)).

Second, this argument should not be confused with the Argument from Rule-Circularity. Basically, the latter has it that you cannot use a certain rule to justify that very same rule. For example, you cannot rely on Modus Ponens in your justification of Modus Ponens. Whether or not this argument is sound depends on delicate issues discussed by Boghossian (2000, pp. 245–254, 2001, pp. 10–14) and Dogramaci (2010). This argument is different from the regress argument at issue. To see this, at least the following two problems have to be distinguished:

Problem A: Under what conditions are we justified in drawing a given inference?
Problem B: Under what conditions are we entitled to employ a given epistemic rule?

These problems are different because Internalism and Externalism (as formulated) form an answer to Problem A only. To recall, Internalism says that we are justified in our inferences only if we know that they are valid, whereas Externalism denies that this is required. The analogous debate for Problem B would take the form of 'we are entitled to employ certain epistemic rules, rather than others, only if such and such (e.g., we know that they are valid)'. I will not go into this debate here.

Importantly, the circularity argument is about Problem B and suggests that there is something fishy about justifying epistemic rules (such as Modus Ponens) by already relying on the rules in question. By contrast, the regress argument discussed in the present paper is about Problem A and suggests that there is something fishy about justifying particular inferences (rather than rules) by further inferences.[7]

13.4 Step (b): Blocking the Entailment?

In other words, Internalism entails Scepticism when the following is in place:

(*) For any inference x, one is justified in drawing x only if (i) one infers that x is valid by this or that logical rule, and (ii) one is justified in drawing the latter.

This is what I discussed in Sect. 13.3. The natural follow-up is: can Internalism modify this assumption to resist Scepticism? In the following I briefly identify

[7] See Wright (2001) for further qualifications on the regress argument.

Boghossian's and Wright's suggestions on this and compare these to Philie's. In Boghossian (Boghossian 2001, pp. 29–35, cf. Boghossian 2000, pp. 248–250) we find:

(**) For any inference x, one is justified in drawing x only if one is epistemically responsible and relies on a rule which is meaning-constituting (i.e. it constitutes the meaning of a logical constant, such as '$\supset$').

And Wright (2001, pp. 78–80) suggests:

(***) For any inference *x* (or at least for the simple cases), one is justified in drawing *x* only if one non-inferentially knows that *x* is valid (i.e. one directly sees that *x* is valid or, as it is traditionally put, one has 'rational insight' in this).

Both of these are weakenings of the overall ambitions of Internalism. Furthermore, neither is committed to the regress (and so to Scepticism), just because they do not say that knowledge of inferences is required for being justified in drawing them. Still, the question remains what epistemic responsibility, meaning constitution and non-inferential knowing consist in, provided we remain within the boundaries of Internalism.[8]

As I have presented it in Sect. 13.2, Philie explores a route which is somewhat different from (**) and (***). Namely, he blocks the entailment of Scepticism by holding onto the overall ambitions of Internalism but restricting the scope of (*) to ordinary inferences (and so denying that (*) applies to higher-order inferences). Philie's motivation for this restriction was basically, as we have seen, that ordinary inferences cannot be violated in the same way as higher-order inferences can be violated. Specifically, the proposal is that it would not be irrational to violate Disjunctive Syllogism as in, for example, the following scenario:

(1) Socrates is mortal, or he is unfindable.
(2) It is not so that he is unfindable.
(3*) So, Socrates is neither mortal nor unfindable.

By contrast, it would be irrational to violate Modus Ponens as in, for example, the following case:

(4) If an inference is based on Disjunctive Syllogism, it is valid.
(5) The inference of (3) from (1) and (2) is based on Disjunctive Syllogism.
(6*) So, the inference of (3) from (1) and (2) is invalid.

This is irrational because if you draw (6*) from (4) and (5) you just do not engage in the practice of justifying inferences.

In the following I would like to take issue with this motivation. I have two points. First, by the same reasoning it might be defended that it is irrational as well to violate Disjunctive Syllogism concerning argument (1)–(3*) above. True: it is not irrational with respect to the practice of justifying inferences (because the argument is simply

[8] I shall not go into this here, but cf. Boghossian (2003).

not about other inferences, but about Socrates). Still, it is irrational with respect to the practice of acquiring knowledge about Socrates. You cannot acquire knowledge about Socrates if you structurally violate Disjunctive Syllogism and other such rules.

My second and more pertinent objection is that Philie conflates two problems which need not go together: one about the *justification* of inferences, one about the *constitution* of that practice. Let me explain. To make his case, Philie draws the analogy between Modus Ponens and the so-called Principle of Continence from Davidson (1970):

(R1) If action x is what you regard as the best option, all-things-considered, then you ought to perform x.

Lazar (1999) has shown that it is possible to set-up a regress argument against the use of this principle. Consider the following example. If I take it to be the best option to buy fair trade products, all-things-considered, then by (R1) I ought to buy fair trade products. Now suppose I accept that I indeed take it that buying fair trade products is the best option, yet deny that I am obliged to buy them just because I deny (R1). One could appeal to the following meta-principle to show why (R1) must be accepted:

(R1*) If following (R1) is what you regard as the best option, all-things-considered, then you ought to follow (R1).

But of course I can resist this meta-principle in the same way, and a regress is off. It is not difficult to find more of such Carroll-style arguments in ethics. For instance, Dreier (2001) and Brunero (2005) present such an argument in terms of (R2) and (R3) respectively:

(R2) If you desire to perform action x and believe that only by performing y you will perform x, then you ought to perform y.

(R3) If you desire to comply with a rule R and R requires you to perform action x, then you ought to perform x.

These can be resisted in the same way as (R1). The moral in all these cases is *not* that (R1), (R2) and (R3) are useless. On the contrary, the moral is, it goes, that they are simply more basic than other rules, and should not be treated on the same footing. For example, (R3) above can be used to generate obligations from desires plus requirements, and so the obligation to comply with (R3) should not itself depend on anything (as itself or further meta-principles). The further idea is that (R1), (R2) and (R3) may be listed among the ‘principles of rationality’ (at least among the practical ones), and that violating such basic principles is irrational. This means that they are to be obeyed not because we have an obligation to do so, but because they are considered constitutive for all our obligations. That is to say, without them no obligations are possible in the first place.[9]

[9] To be sure, all the argument shows is that there should be one such principle, not three. For one is enough to generate obligations. But then which one should it be, if any of these? This calls for further argumentation, but let this pass here.

Now Philie suggests the analogy with the case against Internalism. The point is that we have to obey Modus Ponens, otherwise it is not possible within Internalism to justify any inference whatsoever.

My objection immediately follows. Even if the above is right, it does not affect the regress argument and the entailment of Scepticism. Namely, even if Modus Ponens is needed to make the higher-order inferences possible, it does not follow from this that the higher-order inferences are also justified. Yet, the latter is exactly what is needed to stop the regress. Simply put: making justification possible is not to be equated with making it actual.

Compare: even if (R3) above (to pick one of them) is needed to make obligations possible, it does not automatically follow that these obligations are also justified, that is, that the actions to be performed have the additional quality of being morally good.

Still, one might think that if it is rational to obey Modus Ponens, then inferences made on its basis should be justified as well. But this is flawed. According to the notion of rationality at issue (see Sect. 13.2), to say that it is rational to obey Modus Ponens in the practice of drawing high-order inferences is to say, simply and trivially, that Modus Ponens is constitutive for that practice and that we should obey Modus Ponens if we want to engage in it at all. But it does not mean, and importantly so, that we are also justified in our higher-order inferences. The justification question is just a further issue. And the answer to that, as Internalism has it, is that we are justified not depending on whether we obey Modus Ponens but on whether we *know* we obey it. And so the regress goes.

In short, there is a gap between the constitution of the practice of justifying inferences, and the justification itself, and it is not easy to see how the gap can be bridged. If this is right, then step (b) of Philie's argument, i.e. his attempt to block the entailment, does not succeed.

13.5 Coda

To sum up, I have argued that Internalism entails Scepticism when it is committed to the following:

(*) For any inference x, one is justified in drawing x only if (i) one infers that x is valid by this or that logical rule, and (ii) one is justified in drawing the latter.

From this, it can readily be seen that there are three ways in which the argument can be resisted (and all further options are combinations of these):

(a) Drop clause (i).
(b) Drop clause (ii).
(c) Restrict its scope.

From a metaphilosophical perspective, this is no uninteresting result, as this dialectical space can be generalized for other regress arguments. For example, all three strategies are open to Achilles in Carroll's dialogue[10] and the same goes for all arguments which are inspired by it, such as those in ethics (see Sect. 13.4). Surely, that strategies are always possible does not mean that they are always motivated. This is also why analogies between the Carroll-style arguments, even if helpful structure-wise, might well break down at certain points (see, for example, the analogy between Internalism's case and Carroll's case in Sect. 13.3, and the one between Internalism's case and the Carroll-style arguments from ethics in Sect. 13.4).

In this paper I have looked at strategies two and three in the case of the Internalism/Externalism debate regarding the justification of our inferences, and concluded that so far neither has the required support (that is to say, within Internalism). Strategy two, i.e. to hold that one can be justified in an inference x whether or not one is justified in inferring that x is valid, was rejected in Sect. 13.3. And strategy three, i.e. to hold that (*) does not apply to higher-order inferences, was rejected in Sect. 13.4. Therefore, pending further investigations (of, e.g., (**) and (***) listed in Sect. 13.4), the entailment claim stands by (*): Internalism does entail Scepticism.

Acknowledgements Thanks go to: Giuseppe Primiero, Dagmar Provijn and Erik Weber.

References

Boghossian, P. (2000). Knowledge of logic. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 229–254). Oxford: Oxford University Press.

Boghossian, P. (2001). How are objective epistemic reasons possible? *Philosophical Studies, 106*, 1–40.

Boghossian, P. (2003). Blind reasoning. *Proceedings of the Aristotelian Society, 77*, 225–248.

Brunero, J. (2005). Instrumental rationality and Carroll's tortoise. *Ethical Theory and Moral Practice, 8*, 557–569.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind, 4*, 278–280.

Davidson, D. (1970). How is weakness of the will possible? *Essays on actions and events* (pp. 21–42). Oxford: Clarendon.

Dogramaci, S. (2010). Knowledge of validity. *Noûs, 44*, 403–432.

Dreier, J. (2001). Humean doubts about categorical imperatives. In E. Millgram (Ed.), *Varieties of practical reasoning* (pp. 27–47). Cambridge, MA: MIT

Klein, P. (1999). Human knowledge and the infinite regress of reasons. *Philosophical Perspectives, 13*, 297–325.

Lazar, A. (1999). Akrasia and the principle of continence or what the tortoise would say to Achilles. In L. E. Hahn (Ed.), *The philosophy of Donald Davidson* (pp. 381–401). Chicago: Open Court.

Philie, P. (2007). Carroll's regress and the epistemology of logic. *Philosophical Studies, 134*, 183–210.

Rees, W. (1951). What Achilles said to the tortoise. *Mind, 60*, 241–246.

[10] 'What Achilles *could* have said to the Tortoise', to borrow yet another variation on this phrase, cf. Rees (1951), Thomson (1960), and Lazar (1999).

Ryle, G. (1950). ‘If’, ‘so’ and ‘because’. In M. Black (Ed.), *Philosophical analysis* (pp. 302–318). Ithaca: Cornell University Press.

Smiley, T. (1995). A tale of two tortoises. *Mind, 104*, 725–736.

Stroud, B. (1979). Inference, belief, and understanding. *Mind, 88*, 179–196.

Thomson, J. (1960). What Achilles should have said to the tortoise. *Ratio, 3*, 95–105.

Wieland, J. W. (2014). *Infinite regress arguments*. Dordrecht: Springer.

Wright, C. (2001). On basic logical knowledge. *Philosophical Studies, 106*, 41–85.

Chapter 14
Answering by Means of Questions in View of Inferential Erotetic Logic

Andrzej Wiśniewski

14.1 The Dyadic Perspective and Answering by Means of Questions

Logical theories of questions supply formalisms for questions as well as characteristics of the question-answer relation.[1] As long as question asking and question answering are concerned, they usually adopt a simple dyadic perspective. It is assumed that there are two parties, a questioner and an answerer. The former asks a question, whereas the role of the latter is to provide an answer to the question. Even eliciting information from Nature is modelled in this way.[2] The answer to be provided must not be a question, or, to be more precise, answers having the form of questions are permitted only if a clarification is needed.

The dyadic perspective, however, does not account for some phenomena which occur in real-life questioning. Generally speaking, these include: (a) replying with questions that are not clarification requests, and (b) question answering based on additional information actively sought for.

[1] For overviews see, e.g., Harrah (2002), Groenendijk and Stokhof (2011), and Ginzburg (2011). See also Wiśniewski (1995, Chap. 2).

[2] See Hintikka (1999).

A. Wiśniewski (✉)
Department of Logic and Cognitive Science, Institute of Psychology,
Adam Mickiewicz University, Poznań, Poland
e-mail: Andrzej.Wisniewski@amu.edu.pl

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic,
Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_14,
© Springer Science+Business Media Dordrecht 2014

14.1.1 Replying with Questions

When a question is replied with a question, the reply is most often a clarification request.[3] Yet there are cases in which a reply having the form of a question provides, though in an indirect manner, information of the required kind. For instance, let us consider:

Ann: *Is Andrew a genius?*
Bob: *Do penguins fly?*

or

Ann: *Is it true that you always answer with questions?*
Bob: *Really?*

The above examples can be accounted for in terms of "exploitation" of Grice's Conversational Maxims[4] (although the second example presumably requires more than that). Sometimes, however, an analysis of replying with questions carried on in a purely Gricean perspective is insufficient or even inadequate. The story presented below is instructive in this respect.

Example 1. Two parties, **A** and **B**, interact by exchanging questions and information in the following way.

A: *Where did Andrew leave for: Paris, London, or Moscow?*
B: *When did Andrew depart: in the morning, or in the evening?*
A: *In the morning.*
B: *Did Andrew take his famous umbrella?*
A: *No, he didn't.*
B: *[So] Andrew left for Paris.*

B's first move becomes pretty understandable if he knows that:

(1) *If Andrew left for Paris, London or Moscow, then he departed in the morning or in the evening.*
(2) *If Andrew departed in the morning, then he left for Paris or London.*
(3) *If Andrew departed in the evening, then he left for Moscow.*

Similarly, **B**'s second interrogative move is justified if he, in addition, knows that:

(4) *Andrew left for London if he took his famous umbrella; otherwise he did not leave for London.*

[3] *Added in 2013.* For a taxonomy of question-replies (QR) see Łupkowski and Ginzburg (2013). A corpus study reveals that clarification requests constitute about 80 % of QR, while about 10 % of QR falls into the category of 'dependent questions'. QR, in turn, constitute slightly more than 20 % of all responses to queries found in the spoken part of the British National Corpus.

[4] See Grice (1975).

Of course, **B**'s premises need not be known to **A**. So the dialogue might have continued as follows:

A: *How do you know?*
B: *Well, Andrew departed in the morning, and if he had done so, he left for Paris or London. But he did not take his famous umbrella, as he used to when travelling to London. So Andrew left for Paris.*

Note that the question-replies are asked by **B** and answered by **A**. The answer to the principal question is, finally, provided by **B**. In other words, **B**, an answerer, temporarily becomes a questioner in order to accomplish his main task, and **A**, a questioner, temporarily becomes an answerer.

14.1.2 Question Answering Based on Additional Information Actively Sought For

In view of the dyadic perspective a questioner requests information that she lacks (and needs for some reason(s)), whereas an answerer is supposed/obliged to provide information that fully satisfies the request. It often happens, however, that a questioner does not receive a satisfactory answer, but in order to arrive at such an answer has to assemble additional information, and this requires asking "good" auxiliary questions. The stories below illustrate this.

Example 2. As in Example 1, there are two parties, **A** and **B**. This time, however, they interact in a quite different way, viz.:

A: *Where did Andrew leave for: Paris, London, or Moscow?*
B: *Well, if these are the options, then he departed in the morning or in the evening. And if in the morning, he left for Paris or London; otherwise he left for Moscow. And, moreover, he takes his famous umbrella only when travelling to London.*
A: *When did Andrew depart: in the morning, or in the evening?*
B: *In the morning.*
A: *Did Andrew take his famous umbrella?*
B: *No, he didn't.*
A: *[So] Andrew left for Paris.*

Now questions are asked only by **A**. Observe that **B** does not reply **A**'s principal question with a satisfactory answer. Instead, he provides information which is relevant with respect to the question. Then **A** presses further by asking auxiliary questions and **B** answers them. It is **A** who concludes with the answer to the principal question: the answer is based on **B**'s consecutive answers.

Example 3. Now three parties, **A**, **B**, and **C**, are involved. The story goes as follows.

A: *Where did Andrew leave for: Paris, London, or Moscow?*
B: *Well, if these are the options, then he departed in the morning or in the evening. And if in the morning, he left for Paris or London; otherwise he left for Moscow.*
C: *Andrew takes his famous umbrella only when travelling to London.*
A: *When did Andrew depart: in the morning, or in the evening?*
C: *In the morning.*
A: *Did Andrew take his famous umbrella?*
B: *No, he didn't.*
A: *[So] Andrew left for Paris.*

Agents **B** and **C** do not violate the Maxim of Quantity, since none of them is able to provide a satisfactory answer to **A**'s principal question. Agent **B** supplies information that gives rise to **A**'s first auxiliary question, but the question is then answered by **C**. Similarly, **A**'s second auxiliary question arises from information initially provided by **C**, but is answered by **B**. Agent **A**, then, derives the answer to the principal question. The relevant items of information have been actively sought for by means of asking "good" auxiliary questions. Neither **B** nor **C** knew the answer. However, the answer constitutes an item of distributed knowledge of the group. It is **A** who elicits it—by means of asking auxiliary questions.

14.1.3 Alternative Courses of Events

Before we continue, let us observe that each of the above stories could have developed differently if different answers to the emerging questions had been received. The affirmative answer to the umbrella question would give "London" as the outcome. The answer "In the evening", in turn, would give "Moscow" and, what is more important, would cancel the umbrella question. In each case, however, an answer to the principal question will be found. Figure 14.1 displays possible courses of events.

14.2 Transitions from Questions to Questions: A Semantic Analysis

Transitions from questions to questions play an important role in any of the stories presented above. Are these transitions subjected to any logic? As we will see, the answer is "yes". In order to show why, let us, first, analyse the relation between the questions:

(5) *Where did Andrew leave for: Paris, London or Moscow?*
(6) *When did Andrew depart: in the morning, or in the evening?*

in semantic terms.

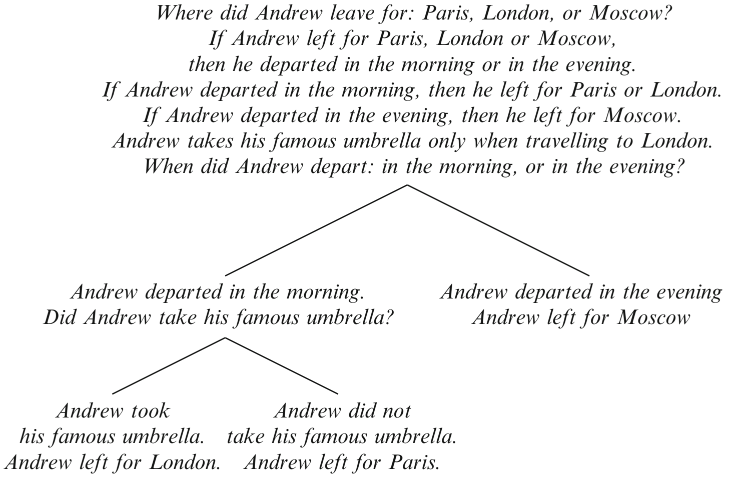

Where did Andrew leave for: Paris, London, or Moscow?
If Andrew left for Paris, London or Moscow,
then he departed in the morning or in the evening.
If Andrew departed in the morning, then he left for Paris or London.
If Andrew departed in the evening, then he left for Moscow.
Andrew takes his famous umbrella only when travelling to London.
When did Andrew depart: in the morning, or in the evening?
Andrew departed in the morning.
Did Andrew take his famous umbrella?
Andrew departed in the evening
Andrew left for Moscow
Andrew took
his famous umbrella.
Andrew left for London.
Andrew did not
take his famous umbrella.
Andrew left for Paris.

Fig. 14.1 Possible courses of events

Question (5) offers three "possibilities": Paris, London, and Moscow. The use of "leaves for" suggests that these possibilities are mutually exclusive. An erotetic logician would say that question (5) has three *direct answers*,[5] namely "Andrew left for Paris", "Andrew left for London", and "Andrew left for Moscow"; the short answers "Paris", "London" and "Moscow" have, respectively, the same meanings (in the current context) as the direct answers. The disjunction of all the direct answers is not a logical truth: it can happen that none of them is true. Andrew might have left for Rome, or stayed at home, and so forth. In terms of erotetic logic: question (5) is *risky*, that is, need not be sound (by a sound question we mean a question which has at least one *true* direct answer). Similarly, question (6) is risky: the direct answers are "Andrew departed in the morning" and "Andrew departed in the evening", and, again, it can happen that neither of them is true.

Now let us bring into the picture the relevant "premises", i.e.:

(1) *If Andrew left for Paris, London or Moscow, then he departed in the morning or in the evening.*
(2) *If Andrew departed in the morning, then he left for Paris or London.*
(3) *If Andrew departed in the evening, then he left for Moscow.*
(7) *Andrew takes his famous umbrella only when travelling to London.*

("only when" is construed here as a biconditional). For brevity, let us designate the set made up of all the above sentences by X^*. Again, the set X^* need not consist of truths. Even if it does, it is still possible that neither question (5) nor question (6)

[5] The logic of questions is also called *erotetic logic* (from Greek *erotema* that means "question"). Roughly, a *direct answer* is a possible just-sufficient answer, i.e. a possible answer which provides neither less nor more information than it is requested by the question.

is sound. But suppose that X^* consists of truths *and*, moreover, that question (5) is sound. Now question (6) must be sound: given the assumptions, either it is the case that Andrew departed in the morning, or it is the case that he departed in the evening.[6] In other words, the phenomenon of transmission of soundness and truth into soundness shows up.

Let us generalize. Suppose that Q_1 is a question that arises out of a question Q together with a set of declarative sentences/formulas X. The relevant phenomenon is described by the following statement:

(I) (TRANSMISSION OF SOUNDNESS/TRUTH INTO SOUNDNESS): *If Q is sound and X consists of truths, then Q_1 must be sound.*

Yet, this is not all. The question:

(6) *When did Andrew depart: in the morning, or in the evening?*

has two direct answers, namely:

(8) *Andrew departed in the morning.*
(9) *Andrew departed in the evening.*

Each of the answers is *cognitively useful*, on the basis of the (set of) premises, for finding an answer to question (5). Moreover, this holds due to some underlying semantic dependencies between the answers to question (6), the set X^*, and question (5). For suppose that answer (8) to question (6) is true and that X^* consists of truths. It follows that Andrew left for Paris or London. In other words, given the assumptions, a true direct answer to question (5) belongs to the following *proper subset* of the set of all the direct answers:

(10) {*Andrew left for Paris, Andrew left for London*}.

Now suppose that answer (9) to question (6) is true, and that all the elements of X^* are true. It follows that Andrew left for Moscow. In terms of sets: a true direct answer to question (5) belongs to the following proper subset of the set of all the direct answers:

(11) {*Andrew left for Moscow*}

which happens to be a unit set.

One can generalize this by:

(II) (OPEN-MINDED COGNITIVE USEFULNESS): *For each direct answer B to Q_1 there exists a non-empty proper subset Y of the set of direct answers to Q such that the following condition holds*:

- *If B is true and X consists of truths, then at least one direct answer (to Q) in Y must be true.*

[6] Clearly, it suffices to suppose that question (5) is sound and premise (1) is true. But a stronger assumption, according to which all the premises are true, does not make any harm. We aim at a semantic relation between a question, a set of declarative sentences/formulas, and a question.

So far we have analysed the semantic relation(s) between questions (5) and (6). But what about the questions:

(5) *Where did Andrew leave for: Paris, London or Moscow?*
(12) *Did Andrew take his famous umbrella?*

Recall that, in the above examples, question (12) occurred after receiving answer (8) to question (6). So the relevant set of "premises" is X^* enriched with sentence (8); let us designate this new set by X^{**}. Condition (**I**) is satisfied for trivial reasons. Condition (**II**) is fulfilled as well. For if the following is true:

(13) *Andrew took his famous umbrella.*

then, by (7) (and thus also by X^{**}), Andrew left for London. On the other hand, X^{**} yields:

(14) *If Andrew did not take his famous umbrella, then he left for Paris.*

which, together with the negative answer to question (12), i.e.:

(15) *Andrew did not take his famous umbrella.*

gives:

(16) *Andrew left for Paris.*

A remark is in order. Condition (**II**) is not satisfied by questions (5) and (12) with respect to the initial set of "premises" X^*. Answer (15) to question (12) only gives:

(17) *Andrew did not leave for London.*

The truth of all the sentences in X^* does not warrant, however, that the disjunction of all the direct answers to question (5) must be true, and thus, even if X^* consists of truths and (17) is true, it need not be the case that the following proper subset of the set of direct answers to question (5):

(18) {*Andrew left for Paris, Andrew left for Moscow*}

contains a true answer. This explains why it is wise to ask question (6) first: doing otherwise may produce a "dead end". We will come back to this issue later on. What is important for now is the following observation: consecutive questions are implied, in the sense of *Inferential Erotetic Logic*, by prior question(s) on the basis of the relevant items of information. This remark may sound mysterious. So let us turn to Inferential Erotetic Logic.

14.3 A Short Note on Inferential Erotetic Logic

By and large, Inferential Erotetic Logic (IEL for short) is a logic that analyses inferences in which questions perform the role of conclusions, and proposes criteria of validity for these inferences. IEL was developed in the 1990s as an alternative to the received view in the logic of questions, which situated the answerhood problem in the center of attention, and to the Interrogative Model of Inquiry, elaborated by Jaakko Hintikka.

In this section we briefly present only these elements of IEL which are needed for our purposes. A more detailed presentation of IEL can be found in Wiśniewski (2001); see also Wiśniewski (1995, 1996).

IEL starts with a trivial observation that before a question is asked or posed, a questioner must arrive at it. In many cases arriving at questions resembles coming to conclusions: there are premises involved and some inferential thought processes take place. If we admit that a conclusion need not be "conclusive", we can say that sometimes questions play the role of conclusions. But questions can also perform the role of premises: it often happens that an agent arrives at a question when looking for an answer to another question. Thus the concept of an *erotetic inference* is introduced. As a first approximation an erotetic inference may be defined as a thought process in which one arrives at a question on the basis of some previously accepted declarative sentence or sentences and/or a previously posed question. There are erotetic inferences of (at least) two kinds: the key difference between them lies in the type of premises involved. In the case of *erotetic inferences of the first kind* the set of premises consists of declarative sentence(s) only, and an agent passes from it to a question. For example:

Andrew always comes in time, but now he is late.

What has happened to him?

The premises of an *erotetic inference of the second kind* consist of a question and possibly some declarative sentence(s); erotetic inferences in which no declarative premise occurs can be regarded as a special case of erotetic inferences of the second kind. The stories presented in Sect. 14.1 involved erotetic inferences of the second kind with non-empty sets of declarative premises. Here is an example of an appropriate erotetic inference which does not rely on any declarative premise:

Is Andrew silly and ugly?

Is Andrew ugly?

An inference, even erotetic, is always someone's inference. In its general setting, however, IEL abstracts from this: erotetic inferences are construed syntactically. Erotetic inferences of the first kind are viewed as ordered pairs $\langle X, Q\rangle$, where X is a finite and non-empty set of declarative sentences and Q is a question. Similarly, an erotetic inference of the second kind is identified with an ordered triple $\langle Q, X, Q_1\rangle$, where Q, Q_1 are questions and X is a finite (possibly empty) set of declarative sentences. When formal languages enriched with questions are dealt with, X is a set of declarative well-formed formulas. Erotetic inferences construed syntactically are also called *erotetic arguments*.

IEL proposes some conditions of validity of erotetic inferences.

As long as we are concerned with inferences which have only declaratives as premises and conclusions, validity amounts to the transmission of truth: if the premises are all true, the conclusion must be true as well. However, it is doubtful whether it makes any sense to assign truth or falsity to questions and thus one cannot apply the above concept of validity to erotetic inferences. But in the case of questions the concept of soundness seems to play an equally important role as the concept of truth in the realm of declaratives. Recall that a question Q is *sound* if and only if at least one direct answer to Q is true, and unsound otherwise. This concept is extensively used in the analysis of validity proposed by IEL. Yet, some other concepts are needed as well.

There are erotetic inferences of (at least) two kinds, and the conditions of validity are distinct for each kind. In this paper we are interested only in erotetic inferences of the second kind. Let $\langle Q, X, Q_1 \rangle$ be such an inference. The message is: the relevant conditions of validity imposed by IEL are exactly the conditions (**I**) and (**II**) specified in Sect. 14.2. Let us recall them.

(I) (TRANSMISSION OF SOUNDNESS/TRUTH INTO SOUNDNESS): *If Q is sound and X consists of truths, then Q_1 must be sound.*

(II) (OPEN-MINDED COGNITIVE USEFULNESS): *For each direct answer B to Q_1 there exists a non-empty proper subset Y of the set of direct answers to Q such that the following condition holds*:

- *If B is true and X consists of truths, then at least one direct answer (to Q) in Y must be true.*

There is no space for a thorough discussion on these conditions, but it can be found elsewhere.[7] Let us only say the following. (**I**) is a natural generalization of the standard condition of validity. It is only a necessary condition of validity, however. If (**I**) had been sufficient, then, for instance, the following would have been valid inferences:

Is Andrew a logician?
Some philosophers are logicians, and some are not.
____
Is Andrew a philosopher?

Is Coco a human?
Humans are mammals.
____
Is Coco a mammal?

The problem here is that the questions which are conclusions have (direct) answers that are cognitively useless: these answers, if accepted, would not contribute to

[7] See Wiśniewski (1995, Chaps. 1 and 8); see also Wiśniewski (1996, 2001).

finding answers to initial questions.[8] On the other hand, an intuitive account of validity suggests that direct answers to the question which is the conclusion should be potentially useful, on the basis of the declarative premises, for finding an answer to the initial question. To secure this, IEL imposes (**II**) as the second necessary condition of validity. Let us stress that condition (**II**) is very demanding: it requires *each* direct answer to the question which is the conclusion to be (potentially) cognitively useful. One may argue that this is too much. However, as we will see, the universality of the claim of condition (**II**) makes the conceptual apparatus of IEL well suited for the analysis of the phenomena we are interested in here (and, needless to say, not only for this).

Conditions (**I**) and (**II**) are expressed somewhat loosely. Of course, IEL offers more than just formulating them. The semantic relation of (*erotetic*) *implication of a question by a question on the basis of a set of declarative formulas* is defined. The exact definition pertains to (a class of) formal languages enriched with questions and supplemented with an appropriate semantics. The details of the construction are presented in the Appendix. Let us stress, however, that erotetic implication is defined in such a way that when a question Q implies a question Q_1 on the basis of a set of declarative formulas X, the conditions (**I**) and (**II**) are satisfied. Then an erotetic inference, $\langle Q, X, Q_1 \rangle$, is said to be *valid* iff Q implies Q_1 on the basis of X; implies in the sense of IEL.

We will write $\mathbf{Im}(Q, X, Q_1)$ for "Q implies Q_1 on the basis of X".

IEL characterizes the properties of the relation **Im**. Again, there is no space for presenting them.[9] Let us only mention here, first, that **Im** is monotone with respect to sets of declaratives involved: if $\mathbf{Im}(Q, X, Q_1)$ and X is included in Y, then $\mathbf{Im}(Q, Y, Q_1)$. Second, **Im** is not "transitive" in the sense that if $\mathbf{Im}(Q, X, Q_1)$ and $\mathbf{Im}(Q_1, X, Q_2)$ hold, then $\mathbf{Im}(Q, X, Q_2)$ need not hold (although it holds in some cases). This is a consequence of the fact that **Im** has the property required by the condition (**II**). On the other hand, as we will see, this lack of "transitivity" makes chains of erotetic inferences non-trivial.

Im is defined in semantic terms. But a transition to the syntactic level is easy. IEL provides *question-implying rules*. These rules are grounded in (meta)theorems which say what questions are (erotetically) implied by what questions on the basis of what sets of declarative formulas.

Although the concept of erotetic implication serves as a tool for defining validity of the corresponding erotetic inferences, the area of applicability of the concept is wider. When $\mathbf{Im}(Q, X, Q_1)$ holds, then both transmission of soundness/truth into soundness takes place and the effect of open-minded cognitive usefulness shows up. These are, undoubtedly, desired properties in case Q_1 is an auxiliary question

[8] In the first case none of the answers is potentially useful. As for the second case, the negative answer is useful, whereas the affirmative answer is useless. Needless to say, in any of the above cases condition (**I**) is satisfied for a trivial reason, due to the structure of the "question-conclusion" only.

[9] See Wiśniewski (1994, 1995, 1996, 2001).

with regard to Q.[10] As we have seen in Sect. 14.2, the transitions from questions to questions that occur in the stories described in Sect. 14.1 display these properties; it is not difficult to show that the relevant items are linked by the relation **Im**. So the message is: there is more logic in the phenomena we are interested in here than one can expect at first sight. However, the logic involved is not specific to them, since erotetic inferences, including valid ones, occur in almost every process of inquiry. So in order to get a better account of the phenomena something more is needed. Our claim is: *erotetic search scenarios* may be of help.

14.4 E-Scenarios

In order to show what erotetic search scenarios[11] (e-scenarios for short) are and how they can be used in our enterprise, let us tell, again, a simple story first.

Suppose that one is looking for the (right) answer to the question of the form: *Which one of the following: p, q, r, holds?* Suppose further that it is known that p holds if s holds, and that either q or r holds if $\neg s$ holds. In this situation one arrives at the question: *Does s hold?*

What can happen next? It depends on the epistemic situation. If the request for information will be satisfied by s, the answer p to the initial question is found. If, however, the request will be satisfied by $\neg s$, the initial question transforms into the question: *Which one of the following: q, r, holds?*

Now suppose that it is also known that q holds if, and only if u holds. In this case one arrives at the question: *Does u hold?* If this request for information will be satisfied by u, one gets the answer q to the initial question. If the outcome will be $\neg u$, one gets r, since if u does not hold, q does not hold either, and, as $q \vee r$ holds, r must hold.

In each case an answer to the initial question is found.

We have told the story in epistemic terms. Let us now look, however, at the underlying structure displayed in Fig. 14.2. It is syntactic[12] and is of course completely *domain-unspecific*.

The exact definition of e-scenarios can be found elsewhere.[13] Here we only highlight some of their basic properties.

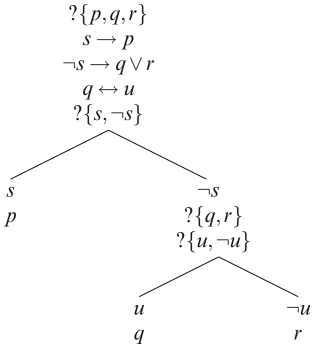
[10] In the case of dialogues, as Ginzburg (2010) observes, when query responses are erotetically implied, relevance (in a dialogue) is retained. Of course, relevance of query responses can be retained in other ways as well.

[11] The concept of erotetic search scenario was introduced in Wiśniewski (2003). See also Wiśniewski (2001).

[12] For simplicity, we operate on the propositional level. Letters $p, q, r, t, \ldots$ are propositional variables. $?\{A_1, \ldots, A_n\}$ is a question whose direct answers are exactly the explicitly listed formulas $A_1, \ldots, A_n$.

[13] See Wiśniewski (2001, 2003, 2004a). In the Appendix we present an equivalent definition in terms of (labelled) trees.

Fig. 14.2 An example of an e-scenario



An e-scenario is always *for* a given ("principal", "initial", "main") question and *relative to* a set of declarative sentences/formulas ("premises"). E-scenarios have a tree-like structure with a principal question as the root and possible (direct) answers to this question as leaves. Auxiliary questions enter e-scenarios on the condition of being erotetically implied (in the sense of IEL). Either an auxiliary question has another question as the immediate successor (cf. question $?\{q, r\}$ above) or an auxiliary question has all the direct answers to it as its immediate successors (cf. $?\{s, \neg s\}$ and $?\{u, \neg u\}$). In the latter case the immediate successors represent the possible ways in which the relevant request for information can be satisfied, and the structure of an e-scenario shows what further information requests (if any) are to be satisfied in order to arrive at an answer to the principal question. If an auxiliary question is a "branching point" of an e-scenario, it is a *query* of the e-scenario. However, an e-scenario can involve auxiliary questions which are not queries, but serve as ("erotetic") premises only. Finally, any declarative sentence/formula that occurs at a path of an e-scenario is either an initial premise, or a direct answer to a query, or a logical consequence of some initial premise(s) and/or answer(s) to queries that occur earlier at the path.

The e-scenarios approach transcends the common scheme of "production of a sequence of questions and affirmations." The fact that information requests can be satisfied in one way or another is taken seriously. An e-scenario shows what is desirable next in case the previous information request has been satisfied in such–and–such way, and does this with regard to any possible way of satisfying the request. In other words, (a diagram of) an e-scenario provides conditional instructions which tell what auxiliary questions should be asked and when they should be asked. Or, to put it differently, it shows "where to go" if such–and–such a direct answer to a query appears to be acceptable and does so with respect to any direct answer to each query.

Still, e-scenarios are abstract entities, defined accordingly in terms of IEL.

E-scenarios differ with respect to degree of complexity. Figures 14.3 and 14.4 present examples of relatively simple, yet quite useful e-scenarios.

Let us comment on Fig. 14.3. Either of q and r constitutes a sufficient condition for p, and their disjunction constitutes the necessary condition for p. So it is rational

Fig. 14.3 An example of an e-scenario

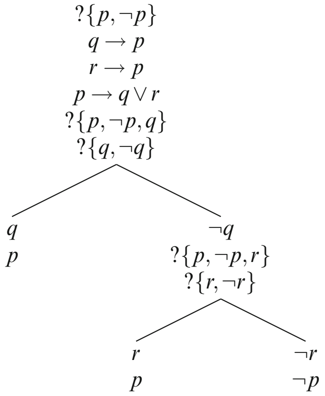
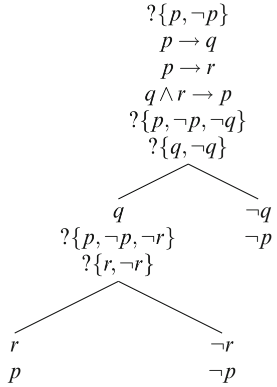


Fig. 14.4 An example of an e-scenario



to ask first if one of them holds. If the tested sufficient condition holds, there is no need for a further question, and the initial issue is solved affirmatively. If it does not hold, it is rational to ask whether the other holds. If it does, the initial issue is solved affirmatively. Otherwise the issue is, finally, solved negatively. Note that neither $?\{p, \neg p, q\}$ nor $?\{p, \neg p, r\}$ is a query. However, they are necessary in the IEL-grounded transitions which lead to queries.[14]

As for Fig. 14.4, this time both q and r are necessary conditions for p, and their conjunction constitutes a sufficient condition for p.[15]

[14] In IEL based on Classical Logic we do not have $\mathbf{Im}(?\{A, \neg A\}, B \rightarrow A, ?\{B, \neg B\})$. However, we have both $\mathbf{Im}(?\{A, \neg A\}, B \rightarrow A, ?\{A, \neg A, B\})$ and $\mathbf{Im}(?\{A, \neg A, B\}, ?\{B, \neg B\})$. So, although **Im** is not "transitive", it is possible to reach $?\{B, \neg B\}$ from $?\{A, \neg A\}$ and $B \rightarrow A$, but *in two steps* (recall that **Im** is monotone with respect to sets of declaratives).

[15] Again, there are non-queries involved. As long as Classical Logic constitutes the background, we do not have $\mathbf{Im}(?\{A, \neg A\}, A \rightarrow B, ?\{B, \neg B\})$. But we do have $\mathbf{Im}(?\{A, \neg A\}, A \rightarrow B, ?\{A, \neg A, \neg B\})$ and $\mathbf{Im}(?\{A, \neg A, \neg B\}, ?\{B, \neg B\})$.

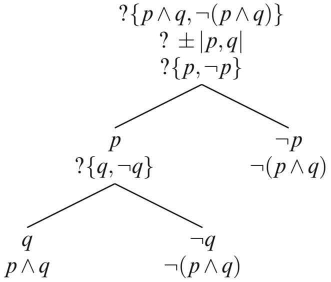


Fig. 14.5 An example of an e-scenario

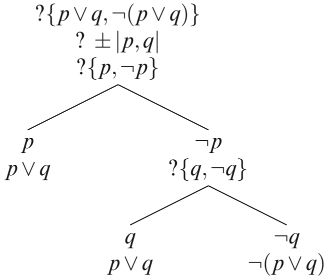


Fig. 14.6 An example of an e-scenario

There exist e-scenarios which do not involve any initial declarative premises (i.e. e-scenarios relative to the empty set). Figures 14.5 and 14.6 present simple examples of e-scenarios of this kind.[16]

For further examples of e-scenarios see, e.g., Wiśniewski (2001, 2003, 2004a).

There are e-scenarios of different kinds. But information-picking e-scenarios seem most useful in an analysis of answering by means of questions. In order to define them we have to introduce some technical concepts first.

Let Φ be an e-scenario for Q relative to X. For simplicity, let us construe Φ as a finite labelled tree (see the Appendix). The labels are either questions or declarative well-formed formulas (d-wffs for short). If γ is a node of Φ, we write ϕ_γ for the path of Φ whose last element/node is γ, and we use $\mathbf{dec}(\phi_\gamma)$ for the set of d-wffs which are labels of the nodes of ϕ_γ.[17] So when γ is a leaf, ϕ_γ is a branch (i.e. a maximal path from the root) and $\mathbf{dec}(\phi_\gamma)$ is the set of d-wffs that label the "declarative" nodes of the branch. By the definition of e-scenarios, the leaves of Φ are labelled by direct answers to Q. Let ℓ be the labelling function of Φ; $\ell(\gamma)$ is thus the expression (a d-wff or a question) that is the label of node γ.

By an *initial premise* of an e-scenario for Q relative to X we mean an element of X that labels a node of the e-scenario.

[16] $? \pm \mid p, q \mid$ abbreviates the *conjunctive question* $?\{p \wedge q, p \wedge \neg q, \neg p \wedge q, \neg p \wedge \neg q\}$.

[17] It can happen that a given question labels more than one node of an e-scenario. However, $\mathbf{dec}(\phi_\gamma)$ is always unique, since γ refers to a node.

We say that a question Q is *informative relative to* a set of d-wffs Z iff no direct answer to Q is entailed by Z.

An e-scenario Φ for Q relative to X is *information-picking* iff:

(a) Q is informative relative to the set of initial premises of Φ, and
(b) If γ is a node of Φ such that $\ell(\gamma)$ is a query of Φ, then $\ell(\gamma)$ is informative relative to $\mathbf{dec}(\phi_\gamma)$, and no immediate successor of γ is labelled by a direct answer to Q, moreover
(c) If γ is a leaf of Φ, then $\ell(\gamma)$ is entailed by $\mathbf{dec}(\phi_\gamma) \setminus \{\ell(\gamma)\}$.

Generally speaking, an information-picking e-scenario has the following features: (i) the principal question cannot be legitimately answered by deriving an answer to it from the relevant initial premises, (ii) all queries of the e-scenario are informative relative to the (sets of) d-wffs that "precede" them: each direct answer to each query brings in new information that is not provided by the (set of) d-wffs which "precede" the query at the relevant path of the e-scenario, (iii) no direct answer to a query is a direct answer to Q, and (iv) it shows that once all the queries that occur at a branch are truthfully answered in the way indicated by their possible (direct) answers occurring at the branch, and the initial premises are all true, the corresponding answer to the principal question (i.e. the direct answer that labels the leaf of the branch) is truthful and thus right.

The e-scenarios displayed in Figs. 14.2–14.6 are information-picking.

Finally, let us point at a certain feature of (all!) e-scenarios which is essential in view of possible applications. One can prove that e-scenarios have the *golden branch property* **GB**:

(GB) *If Φ is an e-scenario for Q relative to X, question Q is sound and the set X consists of truths, then there exists at least one branch of Φ such that each question which is (a label of) a node of the branch is sound, and each d-wff which is (a label of) a node of the branch is true.*

Needless to say, a golden branch "leads" to a true direct answer to the principal question.

14.5 E-Scenarios and Answering by Means of Questions

Let us come back to the examples presented in Sect. 14.1. The description of possible courses of events is shown in Fig. 14.1.

Now consider the e-scenario displayed in Fig. 14.7.

It is clearly visible that the e-scenario presented in Fig. 14.7 is exemplified by the content of Fig. 14.1.[18] Observe that the e-scenario is information-picking.

[18] For simplicity, we remain at the propositional level.

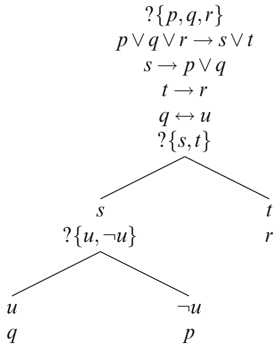


Fig. 14.7 An e-scenario that corresponds to Fig. 14.1

Let us consider Example 1 presented in Sect. 14.1. The key departure from the dyadic model was: an answerer, **B**, replies the questioner, **A**, with questions that are not clarification requests. How can this be explained?

One possible way of thinking is the following. **A**'s question is directed to **B** and hence becomes the principal question of **B**. Then **B** acts according to a *ramified plan*. The plan has the form of an e-scenario *for* the principal question *relative to* what **B**, at the moment, knows about the case. This is, pragmatically, a good plan. First, it is potentially executable due to the golden branch property of e-scenarios—of course assuming that the principal question is sound. Second, the plan copes with any of **A**'s reactions of the expected kind (by this we mean here providing a direct answer to a query). This effect is two-fold: the plan always shows what to do next regardless of which direct answer to a query would be provided by **A**, and the order in which possible queries occur warrants that no direct answer to any query would create a "dead end". Coming back to the example itself. **B** replies with "When did Andrew depart: in the morning, or in the evening?", because this is a query of his ramified plan based on the e-scenario, and replies with this question first because it is the first query of the e-scenario. Then **B** replies with the umbrella question because this is a query of the relevant e-scenario that should be asked if the previous query is answered with "In the morning", and the previous query has been answered this way.

Observe that once the queries are answered, **B** is able to reach his main objective, that is, to provide a satisfactory answer to **A**'s principal question.

Let us now turn to Example 2 of Sect. 14.1. It dealt with the phenomenon of question answering based on additional information actively sought for. This time the "interrogative" roles of **A**, the questioner, and **B**, the answerer, were standard and stable. Yet, **B**'s first answer had not been a satisfactory one, and **A** pressed further by asking consecutive auxiliary questions. Intuitively, each of the auxiliary questions seems "good" with respect to what has happened before. Moreover, the answers received, jointly, yield a satisfactory answer to the initial question. It is **A** who assembles them and finally arrives at the answer. How can all this be accounted for?

As before, one of the possible ways of thinking is the following. After receiving an unsatisfactory answer to her principal question, **A** acts according to a ramified plan. This plan, again, is based on an e-scenario for the principal question, but this time the e-scenario is relative to the item(s) of information received in reply to the question (and, possibly, **A**'s knowledge relevant to the case). The plan has all the advantages specified above. By executing it accordingly, **A** seeks for additional information which, if received, enables her to find, at the end, a satisfactory answer to her principal question. The answer is satisfactory because it displays two desired features. First, it is a direct answer, that is, as Belnap puts it, an answer which is "directly and precisely responsive to the question, giving neither more nor less information than what is called for."[19] Second, the answer is true assuming that the consecutive answers provided by **B** are true.

Now let us consider Example 3. It differs from Example 2 in that it illustrates how a satisfactory answer can be elicited from pieces of knowledge which are scattered over a group (from distributed knowledge of a group). A possible model is this. The questioner, **A**, acts according to a ramified plan based on replies received first. Again, the plan has the form of an e-scenario for the principal question, relative to the replies received first and, possibly, items of **A**'s knowledge (relevant to the case). The details are analogous as above.

So far so good. But three problems arise: (1) where do e-scenarios come from? (2) are the explanations relying on ramified plans (based, in turn, on e-scenarios) empirically adequate?, and (3) do they provide a complete account of the phenomenon of answering by means of questions? The last question is the easiest one, and the answer is obviously "No". As for the first question, the simplest answer is: "E-scenarios just exist, as other logical objects do". If one is not pleased with this Platonic answer, a cautious yet true answer is: IEL enables to construct them. Concerning the problem of empirical adequacy, it can be resolved only by empirical means. My guess is that the explanations are sometimes empirically adequate. But even if this is not right, the aspects of question answering analysed in this paper are real and one may be interested in designing AI agents—or "communities" of them—which act as the models sketched above show, especially when faced with a problem solving task.

14.6 Final Remarks

The concept of e-scenario was initially introduced in order to model some aspects of effective problem solving. One of the crucial principles which govern effective problem solving is the following[20]:

(DP) (DECOMPOSITION PRINCIPLE): *Decompose a principal problem (PP) into simpler sub-problems (SPs) in such a way that solutions to SPs can be assembled into an overall solution to PP.*

[19] See Belnap (1969, p. 124). For direct answers in Belnap's sense see Belnap and Steel (1976).

[20] I owe this formulation of **DP** to Mariusz Urbański.

This principle, viewed from the standpoint of the logic of questions, reduces to:

(EDP) (EROTETIC DECOMPOSITION PRINCIPLE): *Transform a principal question (PQ) into auxiliary questions (AQs) in such a way that: (a) consecutive AQs are dependent upon previous questions and, possibly, answers to previous AQs, and (b) once AQs are resolved, PQ is resolved as well.*

Clearly, e-scenarios enable precisely this. An e-scenario determines a ramified search plan which has all the nice properties we have pointed out above. The more e-scenarios are at hand, the more efficacious an agent who solves problems can be.[21]

As for problem-solving, e-scenarios have been used in computational settings as well (cf., e.g., Bolotov et al. 2006; Łupkowski 2010b). The concept of e-scenario, however, was also applied in some areas outside problem solving proper: in proof theory (cf., e.g., Urbański 2001), in cooperative answering (cf. Łupkowski 2010a), and in the modelling of interrogator's hidden agenda (cf. Urbański and Łupkowski 2010; Łupkowski 2011). As we have seen in this paper, the area of applicability of the concept is even wider.

It is an open question still whether, and if yes, how, the e-scenario approach can be combined with the "issue management" logic of van Benthem and Minică (cf. van Benthem and Minică 2009). A similar question pertains to the multi-agent propositional epistemic logic with questions introduced in Peliš and Majer (cf. Peliš and Majer 2010) and substantially developed in Peliš (2011).[22] Recall that Example 3 and its analysis deal with the process of answer mining among agents (but directed by a "main" agent). Moreover, a branch of an information-picking e-scenario can be viewed as showing after what sequence of truthful public announcements a direct answer to a principal question becomes a piece of common knowledge.

⋆ ⋆ ⋆

[21] It may be of interest that there also exists a second IEL-based approach to problem solving. This time the underlying idea is: transform a question into consecutive questions until a question which can be answered in only one rational way is arrived at. This is modelled by means of the so-called *erotetic calculi*. Rules of these calculi operate on questions only; a rule transforms a question into a further question. A *Socratic transformation* is a sequence of questions, starting with a question about entailment/derivability/theoremhood. This question is then transformed, step by step, into consecutive questions according to the rules of a calculus. Answers play no role in the process. There are successful and unsuccessful transformations; a successful transformation ends with a question of a required final form (the details depend on the logic under consideration). A successful transformation is a *Socratic proof*. The rules are designed in such a way that once a successful transformation is accomplished, the initial issue is affirmatively resolved and there is no need for performing any further deductive moves. Moreover, each step in a Socratic transformation is an IEL-valid inference from a question to a question. So far erotetic calculi have been developed for Classical Logic (see Wiśniewski 2004b; Wiśniewski and Shangin 2006), some paraconsistent propositional logics (see Wiśniewski et al. 2005), and normal modal propositional logics (Leszczyńska 2004, 2007, 2008, 2009). An approach to Intuitionistic Propositional Logic based on a similar idea can be found in Skura (2005).

[22] Cf. also Peliš and Majer (2011).

Formal modelling of problem-solving processes has been one of the subjects of interest of Diderik Batens.[23] This turned his attention to the logic of questions. We have never fully agreed as to what the final solutions should be. Nevertheless, Diderik's criticism and his remarks were always inspiring to me. I dare to dedicate this paper to him. Needless to say, if you blame Diderik for the weak points of this paper in particular, and of IEL in general, you are absolutely wrong. Moreover, there is no adaptive logic that justifies the withdrawal of this conclusion.

Added in 2013. This paper was written in 2011. The reader can find more information about IEL and e-scenarios in: Andrzej Wiśniewski, *Questions, Inferences, and Scenarios*, College Publications, London 2013.

Appendix

Erotetic Implication

Let L be an arbitrary but fixed formal language such that the following conditions are satisfied:

(a) The set $\mathbf{D}_L$ of *declarative well-formed formulas* (d-wffs) of L is defined;
(b) The set Ψ_L of *questions* of L is defined, where $\mathbf{D}_L \cap \Psi_L = \emptyset$;
(c) If Q is a question of L, then there exists an at least two-element set $\mathbf{d}Q \subseteq \mathbf{D}_L$ of *direct answers* to Q;
(d) (The declarative part of) L is supplemented with a semantics rich enough to define the concept of *truth* for d-wffs, and the class of *admissible partitions*.

A *partition* of $\mathbf{D}_L$ is an ordered pair $\mathrm{P} = \langle \mathbf{T}_\mathrm{P}, \mathbf{U}_\mathrm{P} \rangle$, where $\mathbf{T}_\mathrm{P} \cap \mathbf{U}_\mathrm{P} = \emptyset$, and $\mathbf{T}_\mathrm{P} \cup \mathbf{U}_\mathrm{P} = \mathbf{D}_L$. Intuitively, $\mathbf{T}_P$ consists of all the d-wffs which are true in P, and $\mathbf{U}_\mathrm{P}$ is made up of all the d-wffs which are untrue in P. For brevity, we will be speaking about truths and untruths of a partition.

By a partition of L we simply mean a partition of $\mathbf{D}_L$.

Note that we have used the term "partition" as pertaining to the set of d-wffs only. What is "partitioned" is neither the "logical space" nor the set of questions. Recall that $\mathbf{D}_L \cap \Psi_L = \emptyset$. Thus when we have a partition $\langle \mathbf{T}_\mathrm{P}, \mathbf{U}_\mathrm{P} \rangle$ of L and a question of L, the question is neither in $\mathbf{T}_\mathrm{P}$ nor in $\mathbf{U}_\mathrm{P}$.

A question Q is *sound* in a partition $\langle \mathbf{T}_\mathrm{P}, \mathbf{U}_\mathrm{P} \rangle$ iff $\mathbf{d}Q \cap \mathbf{T}_\mathrm{P} \neq \emptyset$.

The concept of partition is very wide and admits partitions which are rather odd from the intuitive point of view. For example, there are partitions in which $\mathbf{T}_\mathrm{P}$ is a singleton set, or in which $\mathbf{U}_\mathrm{P}$ is the empty set. In order to avoid oddity on the one hand, and to reflect some basic semantic facts about the language just considered on the other, we should distinguish a class of *admissible partitions*, being a non-empty subclass of the class of all partitions of the language.

[23] See, e.g., Batens (2003, 2006, 2007, 2014).

Admissible partitions are defined either directly or indirectly. In the former case one imposes some explicit conditions on the class of all partitions. In the latter case one uses a previously given semantics of d-wffs. For example, when $\mathbf{D}_L$ is the set of well-formed formulas of Classical Propositional Calculus (CPC), a partition $\langle \mathbf{T}_{\mathrm{P}}, \mathbf{U}_{\mathrm{P}} \rangle$ is called admissible iff for some CPC-valuation v, $\mathbf{T}_{\mathrm{P}} = \{A \in \mathbf{D}_L : v(A) = \mathbf{1}\}$, and $\mathbf{U}_{\mathrm{P}} = \{A \in \mathbf{D}_L : v(A) = \mathbf{0}\}$.

In what follows it is assumed that we are dealing with expressions of L and admissible partitions of L. For brevity, the specifications "in L" and "of L" are omitted.

Let X stand for a set of d-wffs and let A be a d-wff. *Entailment*, symbolized by $\models$, is defined by:

Definition 1. $X \models A$ iff there is no admissible partition $\langle \mathbf{T}_{\mathrm{P}}, \mathbf{U}_{\mathrm{P}} \rangle$ such that $X \subseteq \mathbf{T}_{\mathrm{P}}$ and $A \in \mathbf{U}_{\mathrm{P}}$.

We also need *multiple-conclusion entailment* (mc-entailment for short).[24] This is a relation between sets of d-wffs. Mc-entailment, $||\!\!=$, is defined as follows:

Definition 2. $X \mid\!\models Y$ iff there is no admissible partition $\langle \mathbf{T}_{\mathrm{P}}, \mathbf{U}_{\mathrm{P}} \rangle$ such that $X \subseteq \mathbf{T}_{\mathrm{P}}$ and $Y \subseteq \mathbf{U}_{\mathrm{P}}$.

Thus X mc-entails Y if there is no admissible partition in which X consists of truths and Y consists of untruths. In other words, mc-entailment between X and Y holds just in case the truth of all the d-wffs in X warrants the presence of some truth(s) among the elements of Y: whenever all the d-wffs in X are true in an admissible partition P, then at least one d-wff in Y is true in the partition P.

Erotetic implication is defined by:

Definition 3. A question Q implies a question Q_1 on the basis of a set of d-wffs X (in symbols: $\mathbf{Im}(Q, X, Q_1)$) iff

1. For each $A \in \mathbf{d}Q : X \cup \{A\} \mid\!\models \mathbf{d}Q_1$, and
2. For each $B \in \mathbf{d}Q_1$ there exists a non-empty proper subset Y of $\mathbf{d}Q$ such that $X \cup \{B\} \mid\!\models Y$.

E-Scenarios as Labelled Trees

E-scenarios have been defined in Wiśniewski (2003) (cf. also Wiśniewski 2001, 2004a) as families of interconnected sequences of questions and d-wffs, the so-called *erotetic derivations*. In this paper, however, we give an equivalent definition in terms of trees. E-scenarios will be defined here as *labelled trees*, where the labels are d-wffs and questions.

[24] For this concept see Shoesmith and Smiley (1978).

Definition 4. A finite labelled tree Φ is an erotetic search scenario for a question Q relative to a set of d-wffs X iff

1. The nodes of Φ are labelled by questions and d-wffs; they are called e-nodes and d-nodes, respectively;
2. Q labels the root of Φ;
3. Each leaf of Φ is labelled by a direct answer to Q;
4. $\mathbf{d}Q \cap X = \emptyset$;
5. For each d-node γ_d of Φ: if A is the label of γ_d, then
   - $A \in X$, or
   - $A \in \mathbf{d}Q^*$, where $Q^* \neq Q$ and Q^* labels the immediate predecessor of Φ, or
   - $\{B_1, \ldots, B_n\} \models A$, where B_i $(1 \leq i \leq n)$ labels a d-node of Φ that precedes the d-node γ_d in Φ;
6. Each d-node of Φ has at most one immediate successor;
7. There exists at least one e-node of Φ which is different from the root;
8. For each e-node γ_e of Φ different from the root: if Q^* is the label of γ_e, then $\mathbf{d}Q^* \neq \mathbf{d}Q$ and
   - $\mathbf{Im}(Q^{**}, Q^*)$ or $\mathbf{Im}(Q^{**}, \{B_1, \ldots, B_n\}, Q^*)$, where Q^{**} labels an e-node of Φ that precedes γ_e in Φ and B_i $(1 \leq i \leq n)$ labels a d-node of Φ that precedes γ_e in Φ, and
   - An immediate successor of γ_e is either an e-node or is a d-node labelled by a direct answer to the question that labels γ_e, moreover
     - If an immediate successor of γ_e is an e-node, it is the only immediate successor of γ_e,
     - If an immediate successor of γ_e is not an e-node, then for each direct answer to the question that labels γ_e there exists exactly one immediate successor of γ_e labelled by the answer.

A *query* of an e-scenario Φ can be defined as a question that labels an e-node of Φ which is different from the root and whose immediate successor is not an e-node. Paths of e-scenarios are construed in the standard manner; a branch is a maximal path which originates from the root. By d-wffs *of* a branch we mean the d-wffs which are labels of d-nodes of the branch, and similarly for questions.

The following holds:

Theorem 1 (Golden Path Theorem). *Let Φ be an e-scenario for Q relative to X. Let* $\mathrm{P} = \langle \boldsymbol{T}_{\mathrm{P}}, \boldsymbol{U}_{\mathrm{P}} \rangle$ *be an admissible partition such that $X \subseteq \boldsymbol{T}_{\mathrm{P}}$ and Q is sound in* P. *Then the e-scenario Φ has at least one branch ϕ such that:*

1. *Each d-wff of ϕ is in $\boldsymbol{T}_{\mathrm{P}}$, and*
2. *Each question of ϕ is sound in* P, *and*
3. *The leaf of ϕ is (labelled by) a direct answer to Q which is in $\boldsymbol{T}_{\mathrm{P}}$.*

The core of the proof lies in the following observations. Erotetic implication preserves soundness given that the relevant declarative premises are true. Needless to say, entailment preserves truth. On the other hand, by the clause (8) of Definition 4, a query has all the direct answers "as" immediate successors, and thus also the true answer(s).

Although the theorem speaks about a ("golden") branch, it is called a Golden Path Theorem because in the original setting (see Wiśniewski 2003) e-scenarios are not defined as trees and what is called a path of an e-scenario in the "old" setting corresponds to a branch of an e-scenario in the current setting.

References

Batens, D. (2003). A formal approach to problem solving. In C. Delrieux & J. Legris (Eds.), *Computer modeling of scientific reasoning* (pp. 15–26). Bahia Blanca: Universidad Nacional del Sur.

Batens, D. (2006). A diagrammatic proof search procedure as part of a formal approach to problem solving. In L. Magnani (Ed.), *Model based reasoning in science and engineering: Cognitive science, epistemology, logic* (pp. 265–284). London: King's College Publications.

Batens, D. (2007). Content guidance in formal problem solving processes. In O. Pombo & A. Gerner (Eds.), *Abduction and the process of scientific discovery* (pp. 121–156). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Batens, D. (2014). *Adaptive logics and dynamic proofs. Mastering the dynamics of reasoning*. In E. Weber, D. Wouters, & J. Meheus (Eds.), *Logic, reasoning, and rationality* (Vol. 5). Dordrecht: Springer. (In progress, parts available at http://logica.ugent.be/adlog/book.html).

Belnap, N. (1969). Åqvist's corrections-accumulating question sequences. In J. Davis, D. Hockney, & W. Wilson (Eds.), *Philosophical logic* (pp. 122–134). Dordrecht: Reidel.

Belnap, N., & Steel, T. (1976). *The logic of questions and answers*. New Haven: Yale University Press.

Bolotov, A., Łupkowski, P., & Urbański, M. (2006). Search and check. Problem solving by problem reduction. In A. Cader (Ed.), *Artificial intelligence and soft computing* (pp. 505–510). Warsaw: Academic Publishing House EXIT.

Ginzburg, J. (2010). Relevance for dialogue. In P. Łupkowski & M. Purver (Eds.), *Aspects of semantics and pragmatics of dialogue. SemDial 2010, 14th workshop on the semantics and pragmatics of dialogue*, Poznań (pp. 121–129). Poznań: Polish Society for Cognitive Science.

Ginzburg, J. (2011). Questions: Logic and interactions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 1133–1146). Amsterdam: Elsevier.

Grice, H. (1975). Logic and conversation. In D. Davidson & G. Harman (Eds.), *The logic of grammar* (pp. 64–75). Encino: Dickenson.

Groenendijk, J., & Stokhof, M. (2011). Questions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 1059–1132). Amsterdam: Elsevier.

Harrah, D. (2002). The logic of questions. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (2nd ed., pp. 1–60). Dordrecht: Kluwer.

Hintikka, J. (1999). *A logic of scientific discovery*. Dordrecht: Kluwer.

Leszczyńska, D. (2004). Socratic proofs for some normal modal propositional logics. *Logique et Analyse, 185–188*, 259–285. Appeared 2005.

Leszczyńska, D. (2007). *The method of Socratic proofs for normal modal propositional logics*. Poznań: Adam Mickiewicz University Press.

Leszczyńska, D. (2008). The method of Socratic proofs for normal modal propositional logics: K5, S4.2, S4.3, S4M, S4F, S4R and G. *Studia Logica, 89*, 371–405.

Leszczyńska, D. (2009). A loop-free decision procedure for modal propositional logics K4, S4 and S5. *Journal of Philosophical Logic, 38*, 151–177.

Łupkowski, P. (2010a). Cooperative answering and inferential erotetic logic. In P. Łupkowski & M. Purver (Eds.), *Aspects of semantics and pragmatics of dialogue. SemDial 2010, 14th workshop on the semantics and pragmatics of dialogue*, Poznań (pp. 75–82). Poznań: Polish Society for Cognitive Science.

Łupkowski, P. (2010b). Erotetic search scenarios and problem decomposition. In D. Rutkowska (Ed.), *Some new ideas and research results in computer science* (pp. 202–214). Warsaw: Academic Publishing House EXIT.

Łupkowski, P. (2011). A formal approach to exploring the interrogator's perspective in the Turing test. *Logic and Logical Philosophy, 20*, 139–158.

Łupkowski, P., & Ginzburg, J. (2013). A corpus-based taxonomy of question responses. In *Proceedings of the 10th international conference on computational semantics (IWCS 2013) – Short papers* (pp. 354–361). Potsdam: Association for Computational Linguistics. http://www.aclweb.org/anthology/W13-0209.

Peliš, M. (2011). *Logic of questions*. Ph.D. thesis, Faculty of Arts, Department of Logic, Charles University in Prague, Prague.

Peliš, M., & Majer, O. (2010). Logic of questions from the viewpoint of dynamic epistemic logic. In M. Peliš (Ed.), *The logica yearbook 2009* (pp. 157–172). London: College Publications.

Peliš, M., & Majer, O. (2011). Logic of questions and public announcements. In N. Bezhanishvili, S. Lo̎bner, K. Schwabe, & L. Spada (Eds.), *Logic, language, and computation 8th international Tbilisi symposium on logic, language, and computation*, Tbilisi (Lecture notes in computer science, Vol. 6618, pp. 145–157). Berlin/Heidelberg: Springer.

Shoesmith, D., & Smiley, T. (1978). *Multiple-conclusion logic*. Cambridge: Cambridge University Press.

Skura, T. (2005). Intuitionistic Socratic procedures. *Journal of Applied Non-Classical Logics, 15*, 453–464.

Urbański, M. (2001). Synthetic tableaux and erotetic search scenarios: Extension and extraction. *Logique et Analyse, 173–175*, 69–91. Appeared 2003.

Urbański, M., & Łupkowski, P. (2010). Erotetic search scenarios: Revealing interrogator's hidden agenda. In P. Łupkowski & M. Purver (Eds.), *Aspects of semantics and pragmatics of dialogue. SemDial 2010, 14th workshop on the semantics and pragmatics of dialogue*, Poznań (pp. 67–74). Poznań: Polish Society for Cognitive Science.

van Benthem, J., & Minică, Ş. (2009). Toward a dynamic logic of questions. In X. He, J. Horty, E. Pacuit (Eds.), *Proceedings of second international workshop on logic, rationality and interaction, LORI II*, Chongging (Lecture notes in computer science, Vol. 5834, pp. 27–41). Springer.

Wiśniewski, A. (1994). Erotetic implications. *Journal of Philosophical Logic, 23*, 174–195.

Wiśniewski, A. (1995). The posing of questions: Logical foundations of erotetic inferences. Dordrecht: Kluwer.

Wiśniewski, A. (1996). The logic of questions as a theory of erotetic arguments. *Synthese, 109*, 1–25.

Wiśniewski, A. (2001). Questions and inferences. *Logique et Analyse, 173–175*, 5–43. Appeared 2003.

Wiśniewski, A. (2003). Erotetic search scenarios. *Synthese, 134*, 389–427.

Wiśniewski, A. (2004a). Erotetic search scenarios, problem-solving and deduction. *Logique et Analyse, 185–188*, 139–166. Appeared 2005.

Wiśniewski, A. (2004b). Socratic proofs. *Journal of Philosophical Logic, 33*, 299–326.

Wiśniewski, A., & Shangin, V. (2006). Socratic proofs for quantifiers. *Journal of Philosophical Logic, 35*, 147–178.

Wiśniewski, A., Vanackere, G., & Leszczyńska, D. (2005). Socratic proofs and paraconsistency: A case study. *Studia Logica, 80*, 433–468.