David Makinson
Jacek Malinowski
Heinrich Wansing
Editors

Trends in Logic 28

Towards Mathematical Philosophy

Papers from the Studia Logica conference Trends in Logic IV

Springer

Towards Mathematical Philosophy

TRENDS IN LOGIC

Studia Logica Library

VOLUME 28

SCOPE OF THE SERIES

Trends in Logic is a bookseries covering essentially the same area as the journal *Studia Logica* – that is, contemporary formal logic and its applications and relations to other disciplines. These include artificial intelligence, informatics, cognitive science, philosophy of science, and the philosophy of language. However, this list is not exhaustive, moreover, the range of applications, comparisons and sources of inspiration is open and evolves over time.

Volume Editor
Ryszard Wójcicki

For other titles published in this series, go to www.springer.com/series/6645

David Makinson • Jacek Malinowski • Heinrich Wansing
Editors

Towards Mathematical Philosophy

Papers from the Studia Logica conference
Trends in Logic IV

Springer

David Makinson
London School of Economics
Department of Philosophy
Houghton Street
London WC2A 2AE
United Kingdom
e-mail: david.makinson@googlemail.com

Jacek Malinowski
Polish Academy of Sciences
Institute of Philosophy and Sociology
Nowy Swiat 72
00-330 Warszawa
Poland
e-mail: jacekm@uni.torun.pl

Heinrich Wansing
Dresden University of Technology
Institute of Philosophy
01062 Dresden
Germany
e-mail: Heinrich.Wansing@tu-dresden.de

ISBN 978-1-4020-9083-7 e-ISBN 978-1-4020-9084-4

Library of Congress Control Number: 2008934403

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Contents

List of Contributors

Arnon Avron, School of Computer Science, Tel-Aviv University, Ramat Aviv 69978, Israel, e-mail: aa@post.tau.ac.il

Thomas Ågotnes, Bergen University College, PO Box 7030, N-5020 Bergen, Norway, e-mail: tag@hib.no

Jonathan Ben-Naim, CNRS Researcher in Computer Science Université Paul Sabatier, IRIT, LILaC 118 route de Narbonne F-31062 Toulouse Cedex 9, France, e-mail: bennaim@irit.fr

Janusz Czelakowski, Institute of Mathematics and Informatics, Opole University, Oleska 48, 45-052 Opole, Poland and
Institute of Mathematics and Physics, Opole University of Technology, Luboszycka 3, 45-036 Opole, e-mail: jczel@math.uni.opole.pl

Melvin Fitting, The Graduate School & University Center, CUNY, Department of Computer Science, 365 Fifth Avenue New York City, NY 10016-4309, USA, e-mail: melvin.fitting@lehman.cuny.edu

Norman Y. Foo, School of Computer Science and Engineering, University of New South Wales, Sydney, NSW2032, Australia,
e-mail: norman@cse.unsw.edu.au

Robert Goldblatt, Centre for Logic, Language and Computation, Victoria University, P.O. Box 600, Wellington, New Zealand,
e-mail: rob.goldblatt@mcs.vuw.ac.nz

Ian Hodkinson, Department of Computing, Imperial College London, London, SW7 2AZ, UK, e-mail: imh@doc.ic.ac.uk

Wiebe van der Hoek, Department of Computer Science, University of Liverpool, Liverpool, L69 3BX United Kingdom, e-mail: wiebe@csc.liv.ac.uk

Beata Konikowska, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, e-mail: beatak@ipipan.waw.pl

Tomasz Kowalski, Research School of Information Sciences and Engineering, Australian National University, bldg 115, cnr North and Daley Rds, Canberra ACT, Australia, e-mail: tomasz.kowalski@anu.edu.au

Rex Bing Hung Kwok, School of Computer Science and Engineering, University of New South Wales, Sydney, NSW2032, Australia,
e-mail: rkwok@cse.unsw.edu.au

Jui-Lin Lee, Department of Philosophy, National Chung-Cheng University, Chia-Yi, Taiwan, e-mail: jlleelogician@gmail.com

Jason Jingshi Li, Research School of Information Science and Engineering, Australian National University, Canberra, ACT0200, Australia,
e-mail: jason.li@anu.edu.au

Piotr Łukowski, Łódź University, Department of Logic and Methodology of Science, ul. Kopcińskiego 16/18, 90–232 Łódź, Poland,
e-mail: lukowski@filozof.uni.lodz.pl

David Makinson, London School of Economics & Political Science, Department of Philosophy, Logic, Houghton Street, London, WC2A 2AE, UK,
e-mail: david.makinson@googlemail.com

Jacek Malinowski, Polish Academy of Sciences, Institute of Philosophy and Sociology, Nowy Swiat 72, 00-330 Warszawa, Poland,
e-mail: jacekm@uni.torun.pl

Yutaka Miyazaki, Meme Media Laboratory, Hokkaido University, Kita 13, Nishi 8, Kita-ku, Sapporo 060-8628, Japan,
e-mail: y-miya@meme.hokudai.ac.jp

Daniele Mundici, Department of Mathematics "Ulisse Dini", University of Florence, Viale Morgagni 67/A, 50134 Florence, Italy,
e-mail: mundici@math.unifi.it

Francesca Poggiolesi, Center for Logic and Philosophy of Science (CLWF), Vrije Universiteit Brussel, 5B457, Pleinlaan 2, 1050 Brussels, Belgium,
e-mail: poggiolesi@gmail.com

Juan A. Rodríguez-Aguilar, Artificial Intelligence Research Institute IIIA, Spanish Council for Scientific Research CSIC, Barcelona, Spain, e-mail: jar@iiia.csic.es

Hans Rott, Institute of Philosophy, University of Regensburg, 93040 Regensburg, Germany, e-mail: Hans.Rott@psk.uni-regensburg.de

Carles Sierra, Artificial Intelligence Research Institute IIIA, Spanish Council for Scientific Research CSIC, Barcelona, Spain, e-mail: sierra@iiia.csic.es

Wojciech Zielonka, Faculty of Mathematics and Computer Science, University of Warmia and Mazury, Żołnierska 14a, 10-561 Olsztyn, Poland, e-mail: zielonka@uwm.edu.pl

Stephan van der Waart van Gulik, Centre for Logic and Philosophy of Science, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium, e-mail: Stephan.vanderWaartvanGulik@UGent.be

Heinrich Wansing, Dresden University of Technology, Institute of Philosophy, 01062 Dresden, Germany, e-mail: HeinrichWansing@tu-dresden.de

Michael Wooldridge, Department of Computer Science, University of Liverpool, Liverpool, L69 3BX United Kingdom, e-mail: mjw@csc.liv.ac.uk

Urszula Wybraniec-Skardowska, Autonomous Section of Applied Logic and Methodology of Sciences, Poznań School of Banking, Department in Chorzów, Wandy 66, 41–500 Chorzów, Poland, e-mail: uws@uni.opole.pl

David Makinson
Jacek Malinowski
Heinrich Wansing

From Logic to Mathematical Philosophy

Introduction

A scientific discipline earns its keep by solving problems. They can be of three broad kinds. First, there are those that are generated by the world around us and our participation in it. In the long run, these are the most decisive ones — a discipline that is not able to make significant contributions to this kind of problem, directly or indirectly, is bound to die out. Then there are problems that are internal to the discipline itself: formulating new concepts, answering new questions, and on a more global level organizing the evolving whole into coherent forms. If such internal development ceases, the theory ossifies, even when it is still a source of useful practical applications. Finally, there is the challenge of contributing towards the solution of problems that take their origin in other disciplines. That, indeed, is a path to glory.

Through its long history, logic has at times shone more brightly in one or another of these directions. For two thousand years, from the death of Aristotle to the early nineteenth century, it seemed to be a body of knowledge perfect and complete in itself, but essentially static, capable of only minor improvements and reorganizations. However, with its mathematization from the time of Boole, then Frege, through the early twentieth century, logic saw an extraordinary growth in its internal development and its connections with the remainder of pure mathematics.

What about its application to the world around us and to problems raised by neighbouring disciplines, other than pure mathematics? By the middle of the twentieth century, these appeared to be falling behind. It is revealing to recall the following remark, made by John Venn in 1866 about the probability theory of his time, in the preface to the first edition of *The Logic of Chance.* By 1966 there were critics who felt much the same about logic:

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28 © Springer Science + Business Media B.V. 2009

> The science of Probability occupies at present a somewhat anomalous position. It is impossible, I think, not to observe in it some of the marks and consequent disadvantages of a *sectional* study. By a small body of ardent students it has been cultivated with great assiduity, and the results they have obtained will always be reckoned among the most extraordinary products of mathematical genius. But by the general body of thinking men its principles seem to be regarded with indifference or suspicion. Such persons may admire the ingenuity displayed, and be struck with the profundity of many of the calculations, but there seems to them, if I may so express it, an *unreality* about the whole treatment of the subject. To many persons the mention of Probability suggests little else than the notion of a set of rules, very ingenious and profound rules no doubt, with which mathematicians amuse themselves by setting and solving puzzles.

How probability theory has changed its image since then! Now deemed one of the most practical of the pure parts of mathematics, and one of the more useful to other sciences from biology through meteorology to physics, it looks outwards as much as inwards.

In recent decades, logic too began deepening its role as a tool for neighbouring scientific disciplines: in computer science, most visibly, but also in theoretical linguistics, cognitive science, game theory and theoretical economics. In at least two cases, game theory and preference aggregation in economics, a reverse flow has also been evident.

In philosophy, the role of logic as an instrument of analysis had been cultivated much earlier under the influence of Russell in the early twentieth century and subsequently the so-called logical positivists, its fortunes rising and falling as excessive hopes and claims were followed by disappointments. However, with the recent development of formal epistemology, a more balanced connection seems here to stay.

The conference *Towards Mathematical Philosophy: Trends in Logic IV* (details below) was organized in 2006 with these three facets of the subject in mind. The viewpoint was that while it is natural, indeed inevitable, that the bulk of the work done in a living and growing discipline is essentially internal, the importance of external connections with the world around us and other disciplines must not be forgotten.

The present volume, which is based on the conference, continues the same triple perspective. Its contributions have been grouped under three main headings. The first group consists of investigations in *modal logic*, which is currently a very active field in the internal development of the area and in applications to linguistics, formal epistemology, and the study of norms. The second contains papers on *non-classical and many-valued logics*, with an eye on applications in computer science and through it to engineering. The third concerns the *logic of belief management*, which is likewise closely connected with recent work in computer science but also links directly with epistemology, the philosophy of science, the study of legal and other normative systems, and cognitive science. The grouping is of course rough, for there are contributions to the volume that lie astride a boundary; at least one of them is relevant, from a very abstract perspective, to all three areas.

We say a few words about each of the individual chapters, to relate them to each other and the general outlook of the volume.

Modal Logics

The first bundle of papers in this volume contains contribution to modal logic. Three of them examine general problems that arise for all kinds of modal logics. The first paper is essentially semantical in its approach, the second proof-theoretic, the third semantical again:

- *Commutativity of quantifiers in varying-domain Kripke models*, by R. Goldblatt and I. Hodkinson, investigates the possibility of commutation (i.e. reversing the order) for quantifiers in first-order modal logics interpreted over relational models with varying domains. The authors study a possible-worlds style structural model theory that does not validate commutation, but satisfies all the axioms originally presented by Kripke for his familiar semantics for first-order modal logic.
- *The method of tree-hypersequents for modal propositional logic*, by F. Poggiolesi, introduces generalised sequent systems for the well-known modal logics **K**, **K4**, **KD**, and **KD4**. The 'tree hypersequents' of these calculi are trees whose nodes carry sequents. The admissibility of weakening and contraction, the invertibility of all rules, soundness and completeness for the systems with cut, and finally, cut elimination are all established. The central proofs are entirely syntactic.
- *All splitting logics in the lattice* NExt(**KTB**), by T. Kowalski and Y. Miyazaki, explores the lattice of all normal extensions of the modal logic **KTB**. It shows that only two logics split this lattice: a surprising negative result, as there are many splitting logics both in a larger lattice NExt(**K**) and in a smaller lattice NExt(**S5**).

The other two papers in this section deal more specifically with temporal, deontic and epistemic modalities. One focuses on temporal logic of computation, the other on epistemic logic of belief justification.

- *On the temporal logic of normative systems*, by T. Ågotnes, W. van der Hoek, J. Rodríguez-Aguilar, C. Sierra and M. Wooldridge, concerns a conservative extension of computational tree logic, where the path quantifiers are replaced by deontic operators. Called NTL, this normative/temporal logic is given a sound and complete axiomatisation, some of its model-checking problems are studied, and complexity results are obtained.
- *Reasoning with justifications*, by M. Fitting, has been positioned in the section on modal logics in view of its techniques but, given its subject matter, could equally well have been grouped with the contributions on belief representation. It surveys what are known as 'justification logics'. In these logics, which have evolved from work of S. Artemov, the usual representation of knowledge claims taking as components the knowledge operator, the item known, and possibly the agent bearing the knowledge, is expanded to include also the justification for the claim as part of the object language.

Non-Classical and Many-Valued Logics

Many-valued logics have always been tempting to those interested in applications, but at the foundations one finds difficult philosophical questions concerning the nature of their logical values. This is evident, for example, in the case of fuzzy logic, which began as a tool for engineers without a clear logical basis or even much mathematical rigour, but which in recent years has been systematized as a logic of residuated structures, thereby locating it in the area of substructural logic as much as that of many-valued logic. Other substructural logics include the syntactic calculi of categorial grammar emanating from work of Ajdukiewicz and Lambek. Quite different approaches to many-valued logic have emerged from logical investigations into information processing, giving rise to the so-called 'useful four-valued logic' of Dunn and Belnap, which have developed into general theories of bilattices and, most recently, trilattices.

- The first paper in this section, *Monotone relations, fixed points, and recursive definitions*, by J. Czelakowski, really cuts across all of the boundaries of this volume. Fixed points play a crucial role in several areas of computer science, mathematics, and logic including domain theory, database theory, finite model theory, nonmonotonic reasoning in the style of Reiter, logic programming with negation, the logic of common knowledge, and the modal μ-calculus. The relationship of fixed point constructions to recursive definitions is both important and subtle. The paper is concerned with reflexive points of relations. The notions of monotone and chain σ-continuous relations are introduced and their significance is revealed in the context of indeterminate recursion principles.

- *Processing information from a set of sources*, by A. Avron, J. Ben-Naim and B. Konikowska, analyses ways in which we can integrate and process information collected by a computer from different directions, in a spirit that derives ultimately from the early work of Dunn and Belnap. It defines the notion of a source-processor structure and, from it, the consequence relation induced by a class of source-processor structures. Certain assumptions about the behaviour of the sources together with an existential strategy for collecting information give rise to many-valued logics defined using the notion of a non-deterministic matrix. Strongly sound and complete sequent calculi for the logics under consideration are presented.
- *The classical model existence theorem in sub-classical predicate logics*, by J.-L. Lee, presents a 'resource-aware' proof of the well-known classical version of the theorem. Four sub-classical predicate logics satisfying the theorem are extracted, and a suitable completeness result is established.
- *Weak implicational logics related to the Lambek calculus — Gentzen versus Hilbert formalisms*, by W. Zielonka, investigates questions of finite axiomatisability. It shows that sequent systems for the implicational fragments of the associative and the non-associative Lambek calculus (with possibly empty antecedents) cannot be axiomatized by any finite number of axioms (closed under substitution) with the cut rule as the only rule of inference. This implies similar results for the corresponding Hilbert-style calculi.
- *Faithful and invariant conditional probability in Łukasiewicz logic*, by D. Mundici, explores one aspect of the connection between logic and probability. Specifically, it shows that we may define the notion of a conditional probability function in the context of Łukasiewicz infinite-valued propositional logic, in a manner that succeeds in avoiding a Dutch book in the sense of de Finetti.
- *A fuzzy logic approach to non-scalar hedges*, by S. van der Waart van Gulik, presents a formal framework for handling so-called 'non-scalar

hedges' of everyday language, that is, expressions such as 'strictly speaking' and 'loosely speaking'. Taking earlier work of George Lakoff as its point of departure, it introduces a new first-order fuzzy logic for reasoning with non-scalar hedges, both semantically and proof-theoretically. The logic makes use of a set of selection functions to provide conceptual information that is critical for the semantics of hedged predicates.

Belief Management

Logic is not just about inference. It concerns, more generally, reasoning and what we might call belief management — the multiple processes of obtaining, assimilating, organizing, prioritizing, and discarding beliefs and conjectures. These operations evidently bring us into close connection with those of data modification in the information sciences, as well as the intellectual operations of human beings studied by cognitive science. At the centre of attention in the contemporary study of belief management is the theory of belief change — contraction, revision, and update — with its surprisingly close relations to the qualitative study of uncertain inference, also known as non-monotonic logic, as well as to other global dimensions of belief sets such as their coherence. The final section of this volume contains four papers from this area.

- *The procedures for belief revision*, by P. Łukowski, presents an approach to belief revision using both the familiar Tarskian notion of a consequence operation and the much less familiar one of an elimination operation. Counterparts of the concepts of structurality and finitariness for elimination operations are defined in this setting in order to address questions of belief change and non-monotonic reasoning.
- *Shifting priorities: simple representations for twenty-seven iterated theory change operators*, by Hans Rott, proposes the use of prioritized belief bases (understood as weakly ordered sets of sentences) as a tool for representing iterated belief change operators, alternative to the usual approaches for closed theories, carried out using systems of spheres or related semantic constructions. To model the iterative aspect of belief change, attention is focused on ways in which the order of prioritization may itself change following a contraction or revision.
- *The coherence of theories — dependencies and weights*, by J. J. Li, R. B. H. Kwok and N. Foo, studies the problem of evaluating competing, but possibly incompatible theories that account for the same set of

empirical observations. To this end, a quantitative notion of theory coherence is developed, with weights assigned to theory components. The concepts are illustrated by an application to game theory and the iterated prisoner's dilemma.

- *On meta-knowledge and truth*, by U. Wybraniec-Skardowska, is devoted to investigating the interplay between knowledge and meta-knowledge. More philosophical in tone than some of the other contributions, it presents a formal theory of language, meaning, ontology, and reality, in terms of which the notion of truth is conceptualized.

Acknowledgements

The present volume originated from the *Studia Logica* International Conference *Towards Mathematical Philosophy: Trends in Logic IV*, which was held in Toruń, Poland, September 1–4, 2006. The editors would like to thank both the Institute of Philosophy and Sociology of the Polish Academy of Sciences and the Department of Logic at Nicolaus Copernicus University Toruń, for providing excellent material conditions for the conference and a warm and stimulating atmosphere. Moreover, we gratefully acknowledge generous financial support from Volkswagen Stiftung (Hannover), the Evert Willem Beth Foundation (Amsterdam) and Springer Verlag (Dordrecht). Last but not least, we would like to express our gratitude to the colleagues who gave their time and energy as anonymous referees for this volume.

David Makinson, Jacek Malinowski, Heinrich Wansing

Robert Goldblatt
Ian Hodkinson

Commutativity of Quantifiers in Varying-Domain Kripke Models

Abstract. A possible-worlds semantics is defined that validates the main axioms of Kripke's original system for first-order modal logic over varying-domain structures. The novelty of this semantics is that it does not validate the commutative quantification schema $\forall x \forall y \varphi \to \forall y \forall x \varphi$, as we show by constructing a counter-model.

Keywords: possible-worlds semantics, commutative quantification, premodel, model, Kripkean model.

Introduction and Overview

Kripke's model theory for first-order modal logic [4] assigns to each world w a set Dw thought of as the domain of individuals that exist in w. The quantifier $\forall x$ is interpreted at a world as meaning "for all existing x". This semantics does not validate the Universal Instantiation schema

UI $\forall x \varphi \to \varphi(y/x)$, where y is free for x in φ,[1]

because the value of variable y may not exist in a particular world. It does however validate the variant

UI° $\forall y(\forall x \varphi \to \varphi(y/x))$, where y is free for x in φ,

along with the schemata

UD $\forall x(\varphi \to \psi) \to (\forall x \varphi \to \forall x \psi)$,

VQ $\varphi \to \forall x \varphi$, where x is not free in φ,

of Universal Distribution, and Vacuous Quantification, as well as being sound for the Universal Generalisation rule

UG from φ infer $\forall x \varphi$.

---

[1] $\varphi(\tau/x)$ is the formula obtained by uniform substitution of term τ in place of free x in φ; the side condition is the usual proviso that no variable of τ becomes bound in $\varphi(\tau/x)$ as a result.

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

In addition this semantics validates the schema

CQ $\forall x \forall y \varphi \to \forall y \forall x \varphi$

of Commutative Quantification, which was shown by Fine [1] not to be derivable from UI°, UD and VQ by using UG and valid Boolean reasoning. Fine's method involves a syntactic transformation that provides a nonstandard definition of the substitution of a constant for free x in φ. Recently Grant Reaber (personal communication) has noted that reading "$\forall x$" as "for cofinitely many values of x" in the structure $(\omega, <)$ gives an interpretation that falsifies CQ while verifying UI°, UD, VQ and UG.

These observations raise the question of whether there is some plausible, "possible-worlds style", structural model theory for systems that have the axioms UI°, UD and VQ, but perhaps not CQ.[2] In this paper such a semantics is presented, and a model constructed that falsifies CQ while validating the other three quantificational axioms, along with the axioms for any specified normal propositional modal logic. The approach has been used previously in [6] and [3] to give a complete semantics for the quantified relevant logic RQ and for a range of first-order modal logics that are incomplete for their standard possible-worlds models.

There are two basic ideas involved. The first, already long exploited in propositional modal logic, is that not every set of worlds need count as a proposition. Instead we take a collection *Prop* of sets of worlds, the *admissible propositions*, that forms a Boolean set algebra closed also under the operation that interprets the modality $\Box$. The "truth value" of any formula must then be a member of *Prop*.

The second notion has long been exploited in algebraic logic: the universal quantifier $\forall x$ is interpreted as a *greatest lower bound* in the lattice of propositions, this being the natural interpretation of arbitrary *conjunctions*. To illustrate this, suppose we have the set W of worlds, and a universe U of individuals that serves as the range of the quantifier $\forall x$. If φ is a formula in which x is the only free variable, let $\varphi(a)$ be the result of replacing free x in φ by the individual a, viewed as a constant. Let $|\forall x\varphi|$ and $|\varphi(a)|$ be the sets of worlds (subsets of W) at which these sentences are true, respectively. Intuitively, $\forall x\varphi$ is semantically equivalent to the conjunction of the $\varphi(a)$'s

---

[2] The axiomatisation of [4] took as axioms the *closures* of all instances of UI°, UD, VQ, tautologies and appropriate modal schemata, with detachment for material implication as the only inference rule. UG and Necessitation (from φ infer $\Box\varphi$) are then derivable rules. Here a closure of φ is any sentence obtained by prefixing universal quantifiers and copies of $\Box$ to φ in any order.

for all $a \in U$. So

$$|\forall x\varphi| = \bigcap_{a \in U} |\varphi(a)|,$$

where $\bigcap$ is set-theoretic intersection. This makes $|\forall x\varphi|$ the greatest lower bound of the $|\varphi(a)|$'s in the lattice of *all* subsets of W, i.e. the largest/weakest proposition that implies all of the propositions $|\varphi(a)|$. But if we are constrained to use the set *Prop* of *admissible* propositions, which may not be the full powerset $\wp W$ of W, then instead we should take

$$|\forall x\varphi| = \bigsqcap_{a \in U} |\varphi(a)|,$$

where $\bigsqcap$ is the greatest lower bound operation in the ordered set $(\mathit{Prop}, \subseteq)$. The definition of "model" should require that $\bigsqcap_{a \in U} |\varphi(a)|$ always exists in *Prop*. It will be the weakest *admissible* proposition that implies all of the $|\varphi(a)|$'s. *But it may not be equal to* $\bigcap_{a \in U} |\varphi(a)|$!

This interpretation, as developed in [3], has the quantifiers ranging over a fixed domain of possible individuals. But here we have the varying domains $Dw \subseteq U$ of existing individuals, with $\forall x\varphi$ being equivalent to the conjunction of the assertions "if a exists then $\varphi(a)$" for all $a \in U$. To formalise this, let $Ea = \{w \in W : a \in Dw\}$, so that Ea represents the proposition "a exists". Then we want

$$|\forall x\varphi| = \bigsqcap_{a \in U} Ea \Rightarrow |\varphi(a)|, \tag{0.1}$$

where $\Rightarrow$ is the Boolean set implication operation: $X \Rightarrow Y = (W \setminus X) \cup Y$. When $\bigsqcap = \bigcap$, equation (0.1) reproduces the Kripkean semantics of [4] for the quantifier $\forall x$.

In working with greatest lower bounds we put

$$\bigsqcap S = \bigcup \{X \in \mathit{Prop} : X \subseteq \bigcap S\},$$

so that $\bigsqcap S$ is defined for an arbitrary $S \subseteq \wp W$. When $S \subseteq \mathit{Prop}$ and $\bigsqcap S \in \mathit{Prop}$, then $\bigsqcap S$ is indeed the greatest lower bound of S in *Prop*. Also, if $\bigcap S \in \mathit{Prop}$, then $\bigsqcap S = \bigcap S$. But by making $\bigsqcap$ a totally defined operation we ensure that $|\forall x\varphi|$ is always defined, regardless of whether it is admissible. We will see that admissibility of $|\forall x\varphi|$ is not required for the validity of a number of principles, including UI°, UD and UG, but is required for VQ.

We will show that if all of the Ea's are admissible (i.e. $Ea \in \mathit{Prop}$), then the definition (0.1) of $|\forall x\varphi|$ validates CQ. The same conclusion holds if U is finite, or if the Boolean algebra *Prop* is atomic, hence if *Prop* is finite, and hence if W is finite. Moreover, validity of CQ follows if equality is definable in the model in the sense that there is a formula "$x \approx y$" such that when instantiated with any two elements a, b in the domain we obtain

$$|a \approx b| = \begin{cases} W, & \text{if } a = b, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Thus the construction of a falsifying model for CQ is not a simple matter.

In Sections 1–3 we define *model structures, premodels* (in which $|\forall x\varphi|$ need not be admissible) and *models* (in which it is), and prove several soundness results. Section 4 gives sufficient criteria for validity of CQ, and Section 5 constructs its falsifying model. The final Section 6 briefly states completeness results for various logics relative to the given semantics, and points out some interesting relationships between CQ and the Barcan formula.

1. Model Structures

A *model structure* is a system $\mathcal{S} = (W, R, \mathit{Prop}, U, D)$ such that

- W is a set, and R is a binary relation on W;
- *Prop* is a Boolean subalgebra of the powerset algebra $\wp W$;
- *Prop* is closed under the operation $[R]$ defined by

$$[R]X = \{w \in W : \forall v \in W(wRv \text{ implies } v \in X)\};$$

- U is a set, and D is a function assigning to each $w \in W$ a subset $Dw \subseteq U$.

Members of *Prop* are called the *admissible* sets of $\mathcal{S}$. For each $a \in U$ we define $Ea = \{w \in W : a \in Dw\}$. Sets of the form Ea may be referred to as "existence sets". They are not required to be admissible.

Using *Prop* we define, for each $X \subseteq W$,

$$X{\downarrow} = \bigcup\{Y \in \mathit{Prop} : Y \subseteq X\},$$
$$X{\uparrow} = \bigcap\{Y \in \mathit{Prop} : X \subseteq Y\},$$

giving $X{\downarrow} \subseteq X \subseteq X{\uparrow}$. The sets $X{\downarrow}$ and $X{\uparrow}$ need not belong to *Prop*, but if they do, then $X{\downarrow}$ is the largest admissible subset of X, and $X{\uparrow}$ the smallest admissible superset. So if $X \in \mathit{Prop}$, then $X{\downarrow} = X{\uparrow} = X$. Operations $\bigsqcap$ and $\bigsqcup$ on $\wp\wp W$ are defined by putting, for all $S \subseteq \wp W$,

$$\bigsqcap S = (\bigcap S){\downarrow}, \qquad \bigsqcup S = (\bigcup S){\uparrow}.$$

Then any *admissible* X has $X \subseteq \bigsqcap S$ iff $X \subseteq \bigcap S$. If $S \subseteq \mathit{Prop}$ and $\bigsqcap S \in \mathit{Prop}$, then $\bigsqcap S$ is the *greatest lower bound* of S in the partially-ordered set (*Prop*, $\subseteq$), i.e. the largest admissible set included in every member of S. Dual statements hold concerning the role of $\bigsqcup S$ as the *least upper bound* of $S \subseteq Prop$.

It is quite possible that $\bigsqcap S$ is admissible while $\bigcap S$ is not. However, if $\bigcap S \in Prop$ then $\bigsqcap S = \bigcap S$.

We now record some useful facts about $\bigsqcap$, some of which involve the Boolean set "implication" operation $\Rightarrow$, defined by $X \Rightarrow Y = (W \setminus X) \cup Y$. Its main property is that $Z \subseteq X \Rightarrow Y$ iff $Z \cap X \subseteq Y$.

In the following Lemma, X_i, Y_i, X_{ij} are subsets of W, S is a subset of $\wp W$, and $\bigsqcap_{i \in I} X_i$ is $\bigsqcap\{X_i : i \in I\}$.

Lemma 1.1.

(1) *If* $X_i \subseteq Y_i$ *for all* $i \in I$, *then* $\bigsqcap_{i \in I} X_i \subseteq \bigsqcap_{i \in I} Y_i$.

(2) $\bigsqcap_{i \in I} \bigsqcap_{j \in J} X_{ij} = \bigsqcap_{j \in J} \bigsqcap_{i \in I} X_{ij}$.

(3) *If* $X \in Prop$, *then* $X \Rightarrow \bigsqcap S = \bigsqcap_{Y \in S}(X \Rightarrow Y)$.

(4) *If* $\{Y_i : i \in I\} \subseteq Prop$, *then* $\bigsqcap_{i \in I}(X_i \Rightarrow Y_i) = \bigsqcap_{i \in I}(X_i\uparrow \Rightarrow Y_i)$.

Proof.

(1) $\bigcap_{i \in I} X_i \subseteq \bigcap_{i \in I} Y_i$, and the operation $\downarrow$ is $\subseteq$-monotonic.

(2) (N.B: the X_{ij}'s need not be admissible here.)
Let X be an admissible subset of $\bigsqcap_{i \in I} \bigsqcap_{j \in J} X_{ij}$. Then $X \subseteq X_{ij}$ for all $(i, j) \in I \times J$. So, for a given $j_0 \in J$ we have $X \subseteq X_{ij_0}$ for all $i \in I$, hence $X \subseteq \bigsqcap_{i \in I} X_{ij_0}$ because $X \in Prop$. Since this holds for every $j_0 \in J$, $X \subseteq \bigsqcap_{j \in J} \bigsqcap_{i \in I} X_{ij}$, again as X is admissible. Symmetrically, each admissible subset of $\bigsqcap_{j \in J} \bigsqcap_{i \in I} X_{ij}$ is a subset of $\bigsqcap_{i \in I} \bigsqcap_{j \in J} X_{ij}$. Hence $\bigsqcap_{i \in I} \bigsqcap_{j \in J} X_{ij} = \bigsqcap_{j \in J} \bigsqcap_{i \in I} X_{ij}$, since both are unions of admissible subsets.

(3) (N.B: the members of S need not be admissible.)
Since $Y \subseteq (X \Rightarrow Y)$, $\bigsqcap S \subseteq \bigsqcap_{Y \in S}(X \Rightarrow Y)$ by (1). Also, as $W \setminus X \subseteq (X \Rightarrow Y)$, and $W \setminus X \in Prop$ because $X \in Prop$, we have $W \setminus X \subseteq \bigsqcap_{Y \in S}(X \Rightarrow Y)$. Altogether then,

$$X \Rightarrow \bigsqcap S = W \setminus X \cup \bigsqcap S \subseteq \bigsqcap_{Y \in S} (X \Rightarrow Y).$$

For the converse inclusion it is enough to show that any admissible subset of $\bigcap_{Y \in S}(X \Rightarrow Y)$ is a subset of $X \Rightarrow \bigsqcap S$. But if $Z \in Prop$ has $Z \subseteq \bigcap_{Y \in S}(X \Rightarrow Y)$, then for all $Y \in S$, $Z \subseteq (X \Rightarrow Y)$, so $Z \cap X \subseteq Y$. Hence $Z \cap X \subseteq \bigsqcap S$ as $Z \cap X \in Prop$. Therefore $Z \subseteq X \Rightarrow \bigsqcap S$.

(4) (N.B: the X_i need not be admissible.)
First, since $X_i \subseteq X_i{\uparrow}$, we have $(X_i{\uparrow} \Rightarrow Y_i) \subseteq (X_i \Rightarrow Y_i)$, for all $i \in I$. Hence $\bigsqcap_{i\in I}(X_i{\uparrow} \Rightarrow Y_i) \subseteq \bigsqcap_{i\in I}(X_i \Rightarrow Y_i)$ by (1).
For the converse inclusion, let Z be any admissible subset of $\bigsqcap_{i\in I}(X_i \Rightarrow Y_i)$. Then for all $i \in I$, $Z \subseteq X_i \Rightarrow Y_i$, hence $X_i \subseteq Z \Rightarrow Y_i$. But $Z \Rightarrow Y_i$ is admissible (by admissibility of Z and Y_i), and so $X_i{\uparrow} \subseteq Z \Rightarrow Y_i$, implying that $Z \subseteq X_i{\uparrow} \Rightarrow Y_i$. Hence $Z \subseteq \bigsqcap_{i\in I}(X_i{\uparrow} \Rightarrow Y_i)$. ■

2. Premodels and Models

Let $\mathcal{L}$ be a set of relation and function symbols and individual constants. A *premodel* $\mathcal{M} = (\mathcal{S}, |\cdot|^{\mathcal{M}})$ for $\mathcal{L}$, based on a model structure $\mathcal{S}$, is given by an interpretation function $|\cdot|^{\mathcal{M}}$ on $\mathcal{L}$ that assigns

- to each n-ary relation symbol P a function $|P|^{\mathcal{M}} : U^n \to \mathit{Prop}$,
- to each individual constant c an element $|\mathsf{c}|^{\mathcal{M}} \in U$, and
- to each n-ary function symbol F a function $|F|^{\mathcal{M}} : U^n \to U$.

We emphasise that the language is not assumed to have an equality symbol, by which we would mean a binary relation symbol P interpreted in $\mathcal{M}$ by

$$|P|^{\mathcal{M}}(a,b) = \begin{cases} W, & \text{if } a = b, \\ \emptyset, & \text{otherwise.} \end{cases}$$

As indicated in the Introduction, and as will be proved in Corollary 4.5, the presence of a relation symbol thus interpreted, or even the definability of the equality relation at all in the model, entails the validity of CQ.

We deal with first-order modal $\mathcal{L}$-formulas, without equality, generated using a set $\{x_n : n < \omega\}$ of first-order variables, but often regard this set simply as ω by identifying x_n with n. A variable-assignment is then a map $f \in {}^{\omega}U$. Any $\mathcal{L}$-term τ can be interpreted via f as an element $\tau^{\mathcal{M}}f \in U$ in the usual way. We use the letters $x, y, z, \cdots$ for variables, and define $f[a/x]$ to be the function that "updates" f by assigning the value $a \in U$ to x and otherwise acting as f.

A premodel gives an interpretation $|\varphi|^{\mathcal{M}} : {}^{\omega}U \to \wp W$ to each $\mathcal{L}$-formula. For each assignment f, $|\varphi|^{\mathcal{M}}f$ is thought of as the set of worlds at which φ is true under f. This is defined by induction on the formation of φ:

- $|P\tau_1 \cdots \tau_n|^{\mathcal{M}}f = |P|^{\mathcal{M}}(\tau_1^{\mathcal{M}}f, \ldots, \tau_n^{\mathcal{M}}f) \in \mathit{Prop}$,
- $|\top|^{\mathcal{M}}f = W$ and $|\bot|^{\mathcal{M}}f = \emptyset$,
- $|\neg\varphi|^{\mathcal{M}}f = W \setminus |\varphi|^{\mathcal{M}}f$, and $|\varphi \wedge \psi|^{\mathcal{M}}f = |\varphi|^{\mathcal{M}}f \cap |\psi|^{\mathcal{M}}f$,

- $|\Box\varphi|^{\mathcal{M}}f = [R]|\varphi|^{\mathcal{M}}f$,
- $|\forall x\varphi|^{\mathcal{M}}f = \bigsqcap_{a\in U}\big(Ea \Rightarrow |\varphi|^{\mathcal{M}}f[a/x]\big)$.

Thus if $X \in \mathit{Prop}$, then $X \subseteq |\forall x\varphi|^{\mathcal{M}}f$ iff $X \subseteq Ea \Rightarrow |\varphi|^{\mathcal{M}}f[a/x]$ for all $a \in U$. We have

$$\begin{aligned}|\forall x\varphi|^{\mathcal{M}}f &= \Big[\bigcap_{a\in U} Ea \Rightarrow |\varphi|^{\mathcal{M}}f[a/x]\Big]{\downarrow}.\\ &= \Big[\bigcap_{a\in U}(W\setminus Ea)\cup|\varphi|^{\mathcal{M}}f[a/x]\Big]{\downarrow}.\end{aligned}$$

Identifying $\exists$ with $\neg\forall\neg$ gives

$$\begin{aligned}|\exists x\varphi|^{\mathcal{M}}f &= \bigsqcup_{a\in U} Ea\cap|\varphi|^{\mathcal{M}}f[a/x]\\ &= \Big[\bigcup_{a\in U} Ea\cap|\varphi|^{\mathcal{M}}f[a/x]\Big]{\uparrow}.\end{aligned}$$

Remark 2.1. The semantics of [4] interprets an n-ary relation symbol P as a function

$$\Phi(P,\cdot) : W \to \wp(U^n)$$

assigning to each world w an n-ary relation $\Phi(P,w) \subseteq U^n$. From such a Φ we can define $|P| : U^n \to \wp W$ by

$$w \in |P|(a_1,\ldots,a_n) \quad\text{iff}\quad \langle a_1,\ldots,a_n\rangle \in \Phi(P,w).$$

Alternatively, this can be viewed as a definition of Φ, given $|P|$, so the two methods are equivalent. We find that use of the "proposition-valued" functions $|\varphi|$ provides a convenient way of handling the restriction to admissible propositions.

It is worth emphasising that this kind of model theory allows relations and properties to hold of non-existent objects (e.g. Pegasus has wings). Thus it is not required that $\Phi(P,w) \subseteq (Dw)^n$; equivalently, it is not required that

$$|P|(a_1,\ldots,a_n) \subseteq Ea_1\cap\cdots\cap Ea_n.$$

In fact there are numerous ways to set up a model theory for the language of first-order modal logic, depending on a whole range of potential requirements like this, including whether terms are allowed to be non-rigid (i.e. world-dependent), whether they are interpreted locally at a world (i.e. as a member of the domain of that world), whether predicates are taken to be extensional or intensional, whether domains are fixed or variable or nested, etc. The "quantified modal logic roadmap" of [2, Figure 1] gives some impression of the complexity of this range of possibilities. ■

Writing $\mathcal{M}, w, f \models \varphi$ to mean that $w \in |\varphi|^{\mathcal{M}} f$, we get the following clauses for this satisfaction relation $\models$, with all except that for $\forall$ being familiar:

- $\mathcal{M}, w, f \models P\tau_1 \cdots \tau_n$ iff $w \in |P\tau_1 \ldots \tau_n|^{\mathcal{M}} f$,
- $\mathcal{M}, w, f \models \top$ and $\mathcal{M}, w, f \not\models \bot$,
- $\mathcal{M}, w, f \models \neg\varphi$ iff $\mathcal{M}, w, f \not\models \varphi$,
- $\mathcal{M}, w, f \models \varphi \wedge \psi$ iff $\mathcal{M}, w, f \models \varphi$ and $\mathcal{M}, w, f \models \psi$,
- $\mathcal{M}, w, f \models \Box\varphi$ iff for all $v \in W(wRv$ implies $\mathcal{M}, v, f \models \varphi)$.
- $\mathcal{M}, w, f \models \forall x\varphi$ iff there is an $X \in \mathit{Prop}$ such that $w \in X$ and $X \subseteq \bigcap_{a \in U} \big(Ea \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\big)$.

A formula φ is *valid in* premodel $\mathcal{M}$, written $\mathcal{M} \models \varphi$, if $|\varphi|^{\mathcal{M}} f = W$ for all f, i.e. if $\mathcal{M}, w, f \models \varphi$ for all $w \in W$ and $f \in {}^{\omega}U$.

As with standard semantics, satisfaction of a formula depends only on value-assignment to *free* variables:

Lemma 2.2. *In any premodel $\mathcal{M}$, for any formula φ, if assignments $f, g \in {}^{\omega}U$ agree on all free variables of φ, then $|\varphi|^{\mathcal{M}} f = |\varphi|^{\mathcal{M}} g$.*

Proof. The only departure from the standard proof is the inductive case that φ is $\forall x\psi$. Then if f and g agree on all free variables of φ, then for each $a \in U$, $f[a/x]$ and $g[a/x]$ agree on all free variables of ψ, so $|\psi|^{\mathcal{M}} f[a/x] = |\psi|^{\mathcal{M}} g[a/x]$ by induction hypothesis. Hence

$$|\varphi|^{\mathcal{M}} f = \prod_{a \in U} \big(Ea \Rightarrow |\psi|^{\mathcal{M}} f[a/x]\big) = \prod_{a \in U} \big(Ea \Rightarrow |\psi|^{\mathcal{M}} g[a/x]\big) = |\varphi|^{\mathcal{M}} g. \quad \blacksquare$$

This result can be used to establish the usual relationship between syntactic substitution of terms for variables and updating of evaluations:

Lemma 2.3. *Let φ be any formula, and τ a term that is free for x in φ. Then in any premodel $\mathcal{M}$, for any $f \in {}^{\omega}U$, $|\varphi(\tau/x)|^{\mathcal{M}} f = |\varphi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x]$.*

Proof. Again the only nonstandard case is when φ is of the form $\forall y\psi$. First, when x is not free in φ then f and $f[\tau^{\mathcal{M}} f/x]$ agree on all free variables of φ, and $\varphi(\tau/x)$ is just φ, so the result is given by Lemma 2.2.

Otherwise, x is free in φ, so $x \neq y$ and $\varphi(\tau/x) = \forall y(\psi(\tau/x))$ with τ free for x in ψ, so y does not occur in τ. Then

$$|\varphi(\tau/x)|^{\mathcal{M}} f = \prod_{a \in U} Ea \Rightarrow |\psi(\tau/x)|^{\mathcal{M}} f[a/y], \quad \text{and}$$

$$|\varphi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x] = \prod_{a \in U} Ea \Rightarrow |\psi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x][a/y].$$

But for any $a \in U$, the induction hypothesis on ψ gives

$$|\psi(\tau/x)|^{\mathcal{M}} f[a/y] = |\psi|^{\mathcal{M}} f[a/y][\tau^{\mathcal{M}} f[a/y]/x],$$

and $\tau^{\mathcal{M}} f[a/y] = \tau^{\mathcal{M}} f$ because y is not in τ, while

$$f[a/y][\tau^{\mathcal{M}} f/x] = f[\tau^{\mathcal{M}} f/x][a/y]$$

as $y \neq x$. So altogether

$$|\psi(\tau/x)|^{\mathcal{M}} f[a/y] = |\psi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x][a/y],$$

and hence $|\varphi(\tau/x)|^{\mathcal{M}} f = |\varphi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x]$ in this case. ■

Corollary 2.4. *If $\mathcal{M} \models \varphi$, then $\mathcal{M} \models \varphi(\tau/x)$ whenever τ is free for x in φ.*

Proof. If $\mathcal{M} \models \varphi$, then for any f, $|\varphi(\tau/x)|^{\mathcal{M}} f = |\varphi|^{\mathcal{M}} f[\tau^{\mathcal{M}} f/x] = W$. ■

We will say that a formula φ is *admissible in* $\mathcal{M}$ if the function $|\varphi|^{\mathcal{M}}$ has the form ${}^{\omega}U \to \mathit{Prop}$, i.e. $|\varphi|^{\mathcal{M}} f \in \mathit{Prop}$ for all $f \in {}^{\omega}U$. Every atomic formula $P\tau_1 \cdots \tau_n$ is admissible. Given the closure properties of *Prop* it is evident that the set of admissible formulas is closed under the Boolean connectives and $\Box$. In particular, every *quantifier-free* formula is admissible.

A *model* for $\mathcal{L}$ is a premodel in which every $\mathcal{L}$-formula is admissible.

Lemma 2.5. *In any model $\mathcal{M}$, $|\forall x\varphi|^{\mathcal{M}} f = \bigsqcap_{a \in U} \big(Ea{\uparrow} \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\big)$.*

Proof. As φ is admissible in $\mathcal{M}$, $\{|\varphi|^{\mathcal{M}} f[a/x] : a \in U\} \subseteq \mathit{Prop}$. Hence by Lemma 1.1(4),

$$\bigsqcap_{a \in U} \big(Ea \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\big) = \bigsqcap_{a \in U} \big(Ea{\uparrow} \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\big).$$ ■

3. Soundness and $\mathcal{M}$-Equivalence

We now fix a premodel $\mathcal{M}$, and examine the validity of various principles in it, identifying some whose validity requires $\mathcal{M}$ to be a model. From now on, the $\mathcal{M}$-superscript will often be dropped from the notation $|\varphi|^{\mathcal{M}} f$.

Proposition 3.1. *The schemata UI° and UD are valid in $\mathcal{M}$, and the rule UG is sound for validity in $\mathcal{M}$.*

Proof. UG is dealt with first, as it is simplest. If $\mathcal{M} \models \varphi$, then for any f and a, $Ea \Rightarrow |\varphi| f[a/x] = Ea \Rightarrow W = W$, so $|\forall x\varphi| f = \bigsqcap\{W\} = W$. Hence $\mathcal{M} \models \forall x\varphi$.

For UD, suppose that $\mathcal{M}, w, f \models \forall x(\varphi \to \psi)$ and $\mathcal{M}, w, f \models \forall x\varphi$. Then there exist $X, Y \in \mathit{Prop}$ such that

$$w \in X \subseteq \bigcap_{a \in U} Ea \Rightarrow |\varphi \to \psi| f[a/x], \quad \text{and}$$
$$w \in Y \subseteq \bigcap_{a \in U} Ea \Rightarrow |\varphi| f[a/x].$$

Then $w \in X \cap Y \in \mathit{Prop}$, and for all a,

$$X \cap Y \cap Ea \subseteq |\varphi \to \psi| f[a/x] \cap |\varphi| f[a/x] \subseteq |\psi| f[a/x],$$

hence $X \cap Y \subseteq Ea \Rightarrow |\psi| f[a/x]$. This shows $\mathcal{M}, w, f \models \forall x\psi$.

For UI°, let y be free for x in φ. It suffices to show that for any f and a,

$$Ea \subseteq |\forall x\varphi \to \varphi(y/x)| f[a/y]. \tag{3.1}$$

For then $Ea \Rightarrow |\forall x\varphi \to \varphi(y/x)| f[a/y] = W$ for all $a \in U$, so

$$|\forall y(\forall x\varphi \to \varphi(y/x))| f = \textstyle\prod\{W\} = W,$$

and hence $\mathcal{M} \models \forall y(\forall x\varphi \to \varphi(y/x))$.

To prove (3.1), let $w \in Ea$. Then if $w \in |\forall x\varphi| f[a/y]$, there exists $X \in \mathit{Prop}$ with

$$w \in X \subseteq \bigcap_{b \in U} Eb \Rightarrow |\varphi| f[a/y][b/x].$$

In particular, when $b = a$, since $w \in Ea$ we get $w \in |\varphi| f[a/y][a/x]$. But by Lemma 2.3, $|\varphi| f[a/y][a/x] = |\varphi(y/x)| f[a/y]$ because $y^{\mathcal{M}} f[a/y] = a$. Thus

$$w \in |\forall x\varphi| f[a/y] \Rightarrow |\varphi(y/x)| f[a/y] = |\forall x\varphi \to \varphi(y/x)| f[a/y]. \qquad \blacksquare$$

Next we consider the validity of VQ:

Proposition 3.2. *Suppose that x has no free occurrence in φ. If φ is admissible in $\mathcal{M}$, then $\mathcal{M} \models \varphi \to \forall x\varphi$.*

Proof. For any $f \in {}^{\omega}U$ and $a \in U$, the assignments f and $f[a/x]$ agree on all free variables of φ, so by Lemma 2.2,

$$|\varphi| f = |\varphi| f[a/x] \subseteq Ea \Rightarrow |\varphi| f[a/x].$$

But $|\varphi| f \in \mathit{Prop}$ by $\mathcal{M}$-admissibility of φ, so

$$|\varphi| f \subseteq \prod_{a \in U} \big(Ea \Rightarrow |\varphi| f[a/x]\big) = |\forall x\varphi| f.$$

Hence $|\varphi| f \Rightarrow |\forall x\varphi| f = W$ for all f. $\blacksquare$

Corollary 3.3. *Every model validates VQ.*

Proof. In a model, every φ is admissible. ■

We say that formulas φ and ψ are *$\mathcal{M}$-equivalent* if $|\varphi|^{\mathcal{M}} = |\psi|^{\mathcal{M}}$. The following properties of this equivalence relation are left to the reader to check.

Proposition 3.4. *In any premodel $\mathcal{M}$:*

(1) *φ is $\mathcal{M}$-equivalent to ψ iff $\mathcal{M} \models \varphi \leftrightarrow \psi$.*

(2) *If φ is tautologically equivalent to ψ (i.e. $\varphi \leftrightarrow \psi$ is a tautology), then φ and ψ are $\mathcal{M}$-equivalent.*

(3) *$\mathcal{M}$-equivalence is a congruence on the algebra of $\mathcal{L}$-formulas, i.e. if the pair φ, ψ are $\mathcal{M}$-equivalent, then so are the pairs $\neg\varphi, \neg\psi$ and $\varphi \wedge \theta, \psi \wedge \theta$ and $\Box\varphi, \Box\psi$ and $\forall x\varphi, \forall x\psi$ and $\exists x\varphi, \exists x\psi$ etc.*

(4) *If ψ is obtained from φ by replacing some subformula by an $\mathcal{M}$-equivalent formula, then ψ is $\mathcal{M}$-equivalent to φ.* ■

The next result will be used in a model construction in Section 5.

Proposition 3.5. *In any premodel $\mathcal{M}$:*

(1) *$\exists x(\varphi \vee \psi)$ and $\exists x\varphi \vee \exists x\psi$ are $\mathcal{M}$-equivalent.*

(2) *$\exists x(\varphi \wedge \psi)$ and $\varphi \wedge \exists x\psi$ are $\mathcal{M}$-equivalent if φ is admissible in $\mathcal{M}$ and has no free occurrences of x.*

Proof. (1) It is enough to show that the formula

$$\exists x(\varphi \vee \psi) \leftrightarrow \exists x\varphi \vee \exists x\psi$$

is valid in $\mathcal{M}$. But, as the reader can check, this formula is derivable from tautologies and instances of UD using the rule UG and valid Boolean reasoning. Hence it is valid in $\mathcal{M}$ by Proposition 3.1.

(2) If φ is $\mathcal{M}$-admissible and without free x, then $\neg\varphi$ is $\mathcal{M}$-admissible and without free x, so by Lemma 3.2 the formulas $\varphi \to \forall x\varphi$ and $\neg\varphi \to \forall x\neg\varphi$ are valid in $\mathcal{M}$. But from these two, using tautologies, UD, UG and valid Boolean reasoning we can derive

$$\exists x(\varphi \wedge \psi) \leftrightarrow \varphi \wedge \exists x\psi,$$

which is therefore valid in $\mathcal{M}$. ■

4. Validating CQ

We now give some conditions under which the formulas $\forall x \forall y \varphi$ and $\forall y \forall x \varphi$ are $\mathcal{M}$-equivalent in a model. Of course we can assume $x \neq y$ here, for otherwise there is no work to do. Then assignments $f[a/x][b/y]$ and $f[b/y][a/x]$ are identical, and may be written $f[a/x, b/y]$ or $f[b/y, a/x]$.

Lemma 4.1. *In a premodel $\mathcal{M}$, let $f \in {}^{\omega}U$ and let $\mathcal{B}$ be any Boolean subalgebra of Prop that contains $|\varphi|^{\mathcal{M}} f[a/x, b/y]$, $|\forall x \varphi| f[b/y]$, and $|\forall y \varphi| f[a/x]$ for all $a, b \in U$. Then exactly the same atoms of $\mathcal{B}$ are included in the sets $|\forall x \forall y \varphi|^{\mathcal{M}} f$ and $|\forall y \forall x \varphi|^{\mathcal{M}} f$.*

Proof. Let X be an atom of $\mathcal{B}$ with $X \not\subseteq |\forall x \forall y \varphi| f$. Then as $X \in \mathit{Prop}$, there exists $a_0 \in U$ such that

$$X \not\subseteq Ea_0 \Rightarrow |\forall y \varphi| f[a_0/x]. \tag{4.1}$$

Hence $X \not\subseteq |\forall y \varphi| f[a_0/x]$, so again as $X \in \mathit{Prop}$ there exists $b_0 \in U$ such that

$$X \not\subseteq Eb_0 \Rightarrow |\varphi| f[a_0/x, b_0/y]. \tag{4.2}$$

Hence $X \not\subseteq |\varphi| f[a_0/x, b_0/y]$. But X is a $\mathcal{B}$-atom and $|\varphi| f[a_0/x, b_0/y] \in \mathcal{B}$ as given, so X must be *disjoint* from $|\varphi| f[a_0/x, b_0/y] = |\varphi| f[b_0/y, a_0/x]$. Since $X \cap Ea_0 \neq \emptyset$ by (4.1), this implies

$$X \not\subseteq Ea_0 \Rightarrow |\varphi| f[b_0/y, a_0/x].$$

Hence

$$X \not\subseteq \bigsqcap_{a \in U} Ea \Rightarrow |\varphi| f[b_0/y, a/x] = |\forall x \varphi| f[b_0/y].$$

Again the atomicity of X then makes X disjoint from $|\forall x \varphi| f[b_0/y] \in \mathcal{B}$. Since $X \cap Eb_0 \neq \emptyset$ by (4.2),

$$X \not\subseteq Eb_0 \Rightarrow |\forall x \varphi| f[b_0/y].$$

Hence

$$X \not\subseteq \bigsqcap_{b \in U} Eb \Rightarrow |\varphi| f[b/y] = |\forall y \forall x \varphi| f.$$

Conversely, interchanging x and y in this argument shows that if $X \not\subseteq |\forall y \forall x \varphi| f$, then $X \not\subseteq |\forall x \forall y \varphi| f$. ■

Proposition 4.2. *A* **model** *validates CQ if any of the following hold:*

(1) *Prop is an atomic Boolean algebra.*

(2) *Prop is finite.*

(3) *The universe U is finite.*

Proof. (1) Put $\mathcal{B} = \mathit{Prop}$. For any f, all sets $|\varphi|f[a/x, b/y]$, $|\forall x\varphi|f[b/y]$, $|\forall y\varphi|f[a/x]$ are in $\mathcal{B}$ by admissibility. But likewise the sets $|\forall x\forall y\varphi|f$ and $|\forall y\forall x\varphi|f$ are in $\mathcal{B}$, and include the same atoms of $\mathcal{B}$ by Lemma 4.1, hence as $\mathcal{B}$ is atomic this makes $|\forall x\forall y\varphi|f = |\forall y\forall x\varphi|f$.

(2) By (1), as any finite Boolean algebra is atomic.

(3) If U is finite, then for any f,

$$\begin{aligned}&\{|\forall x\forall y\varphi|f, |\forall y\forall x\varphi|f\}\\ &\cup \{|\varphi|f[a/x, b/y], |\forall x\varphi|f[b/y], |\forall y\varphi|f[a/x] : a, b \in U\}\end{aligned}$$

is a finite subset of *Prop*, so it generates a Boolean subalgebra $\mathcal{B}$ of *Prop* that is finite, hence atomic. The proof that $|\forall x\forall y\varphi|f = |\forall y\forall x\varphi|f$ in $\mathcal{B}$ then follows by the argument of (1). ■

Next we consider consequences of admissibility of the "existence sets" Ea and $Ea{\uparrow}$.

Proposition 4.3. *If a model has $Ea{\uparrow} \in \mathit{Prop}$ for all $a \in U$, then it validates CQ.*

Proof. Since we are working in a model, we can use Lemma 2.5 to replace Ea by $Ea{\uparrow}$ in the definition of $|\forall x\varphi|$. Thus

$$\begin{aligned}&|\forall x\forall y\varphi|f\\ &= \textstyle\bigsqcap_{a\in U}\big(Ea{\uparrow} \Rightarrow \bigsqcap_{b\in U}(Eb{\uparrow} \Rightarrow |\varphi|f[a/x, b/y])\big)\\ &= \textstyle\bigsqcap_{a\in U}\bigsqcap_{b\in U}\big(Ea{\uparrow} \Rightarrow (Eb{\uparrow} \Rightarrow |\varphi|f[a/x, b/y])\big) && \text{by Lemma 1.1(3) as } Ea{\uparrow} \in \mathit{Prop},\\ &= \textstyle\bigsqcap_{a\in U}\bigsqcap_{b\in U}\big(Ea{\uparrow} \cap Eb{\uparrow} \Rightarrow |\varphi|f[a/x, b/y]\big) && \text{by set theory.}\end{aligned}$$

Similarly, $|\forall y\forall x\varphi|f = \bigsqcap_{b\in U}\bigsqcap_{a\in U}\big(Eb{\uparrow} \cap Ea{\uparrow} \Rightarrow |\varphi|f[b/y, a/x]\big)$.

But $Eb{\uparrow} \cap Ea{\uparrow} \Rightarrow |\varphi|f[b/y, a/x] = Ea{\uparrow} \cap Eb{\uparrow} \Rightarrow |\varphi|f[a/x, b/y]$, so the $\bigsqcap$-commutation result of Lemma 1.1(2) applies to give $|\forall x\forall y\varphi|f = |\forall y\forall x\varphi|f$. ■

Corollary 4.4. *If a model has $Ea \in \mathit{Prop}$ for all $a \in U$, then it validates CQ.*

PROOF. If $Ea \in Prop$, then $Ea = Ea\!\uparrow$. ■

We say that *equality is definable in* $\mathcal{M}$ if for any distinct variables x, y, there is an $\mathcal{L}$-formula "$x \approx y$" such that

$$|x \approx y|^{\mathcal{M}} f = \begin{cases} W, & \text{if } fx = fy, \\ \emptyset, & \text{otherwise.} \end{cases}$$

COROLLARY 4.5. *If equality is definable in a model, then it validates CQ.*

PROOF. Let $a \in U$ be arbitrary, and suppose $f \in {}^{\omega}U$ satisfies $fx = a$. Then $|\exists y(x \approx y)|f = [\bigcup_{b\in U} Eb \cap |x \approx y| f[b/y]]\!\uparrow = Ea\!\uparrow$. Hence $Ea\!\uparrow \in Prop$ as every formula is admissible in $\mathcal{M}$. By Proposition 4.3, CQ is valid in $\mathcal{M}$.[3] ■

A premodel $\mathcal{M}$ will be called *Kripkean* if it always has

$$|\forall x\varphi|^{\mathcal{M}} f = \bigcap_{a\in U} \big(Ea \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\big).$$

This means that $\forall$ gets the varying-domain semantics of Kripke [4]:

$$\mathcal{M}, w, f \models \forall x\varphi \ \text{ iff } \ \text{for all } a \in Dw,\ \mathcal{M}, w, f[a/x] \models \varphi. \tag{4.3}$$

A Kripkean *model* has

$$\Big[\bigcap_{a\in U} Ea \Rightarrow |\varphi|^{\mathcal{M}} f[a/x]\Big] \in Prop$$

by admissibility of formula $\forall x\varphi$, and conversely this last condition implies that a model is Kripkean.

PROPOSITION 4.6. *Every Kripkean premodel validates CQ.*

PROOF. This is straightforward, essentially because the quantifiers *for all existing ...* commute in the metalanguage. A more formal proof can be given by repeating the proof of Proposition 4.3 with $\bigcap$ in place of $\bigsqcap$ (and Ea in place of $Ea\!\uparrow$). Instead of parts (2) and (3) of Lemma 1.1, the results

$$\bigcap_{i\in I}\bigcap_{j\in J} X_{ij} = \bigcap_{j\in J}\bigcap_{i\in I} X_{ij}, \qquad X \Rightarrow \bigcap S = \bigcap_{Y\in S} (X \Rightarrow Y),$$

are used. These are laws of set theory that hold independently of any admissibility constraints. ■

---

[3] For this proof to work it suffices in fact that $|x \approx y|^{\mathcal{M}} f \supseteq Efx$ when $fx = fy$, and $|x \approx y|^{\mathcal{M}} f = \emptyset$ otherwise.

Remark 4.7. If a model structure $\mathcal{S}$ has *all* sets admissible, i.e. $Prop = \wp W$, then in general $\bigsqcap S = \bigcap S$, so all models on $\mathcal{S}$ are Kripkean and must validate CQ by Proposition 4.6. This shows that a falsifying model for CQ must restrict the admissible sets.

Increasing the expressivity of the language may force more sets to be admissible. This is simply because all formulas have to be interpreted in models as admissible sets. In particular, if equality is definable, all sets $Ea{\uparrow}$ are of the form $|\exists y(x \approx y)|f$ and so become admissible, which is enough to entail CQ (Corollary 4.5).

5. A Countermodel to CQ

This section exhibits a model that falsifies an instance of CQ. It is not so hard to construct a premodel that does this, but we wish to ensure that every formula is admissible in $\mathcal{M}$, so that it validates VQ as well as UI° and UD. From what has been shown in the last Section, our model must have infinite sets for U and *Prop*, and hence for W. Also *Prop* cannot be atomic, and cannot contain every Ea, or every $Ea{\uparrow}$. Moreover, the model cannot be Kripkean, or permit the definability of equality.

Let $\sim$ denote a fixed (but arbitrary) equivalence relation on $\mathbb{Q}$ (the rationals) with infinitely many equivalence classes, each of which is dense in $\mathbb{Q}$: so each interval (a, b) for $a < b$ in $\mathbb{Q}$ contains a point from each equivalence class. Such a relation is easy to construct. Let $b/{\sim}$ denote the $\sim$-equivalence class containing b.

We define a model structure $\mathcal{S} = (W, R, Prop, U, D)$, where

- $W = U = \mathbb{Q}$;
- either $R = \emptyset$, or $R = \{(a, a) : a \in \mathbb{Q}\}$;
- *Prop* is the Boolean subalgebra of $\wp(\mathbb{Q})$ generated by the set of all half-open intervals $[a, b) = \{x \in \mathbb{Q} : a \leq x < b\}$, where $a, b \in \mathbb{Q}$ and $a < b$;
- $Da = \{a\}$ for each $a \in \mathbb{Q}$. Hence $Ea = \{a\}$.

We have actually defined two model structures, depending on the choice of R. In the first case with $R = \emptyset$, $[R]X = W$ for all $X \subseteq W$. In the second case with R the identity relation, $[R]X = X$. Hence in both cases *Prop* is $[R]$-closed. In the first case (W, R) (and hence $(W, R, Prop)$) validates the smallest normal propositional modal logic containing $\Box\bot$, while in the second case it validates the smallest normal logic containing the schema $\Box\varphi \leftrightarrow \varphi$. But each normal propositional modal logic is a sublogic of one of these two [5], so is validated by one of these structures. We will make use of that fact in Section 6.

Each non-empty $X \in \mathit{Prop}$ is a finite union of intervals of the form $(-\infty, a)$, $[b, c)$, and $[d, +\infty)$. *Prop* is atomless, and $Ea{\uparrow} = Ea = \{a\} \notin \mathit{Prop}$ for all $a \in \mathbb{Q}$.

Lemma 5.1. *Write* $\mathbb{Q}/{\sim}$ *for the set of all* $\sim$*-classes, and let* $\mathcal{E} \subseteq \mathbb{Q}/{\sim}$*. Then* $(\bigcup \mathcal{E}){\uparrow}$ *and* $(\bigcup \mathcal{E}){\downarrow}$ *are admissible, with*

$$(\bigcup \mathcal{E}){\uparrow} = \begin{cases} \emptyset, & \text{if } \mathcal{E} = \emptyset, \\ \mathbb{Q}, & \text{otherwise,} \end{cases} \qquad (\bigcup \mathcal{E}){\downarrow} = \begin{cases} \mathbb{Q}, & \text{if } \mathcal{E} = \mathbb{Q}/{\sim}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Proof. If $\mathcal{E} = \emptyset$ then $\bigcup \mathcal{E} = \emptyset$, and clearly $\emptyset{\uparrow} = \emptyset$. Otherwise, by density, any non-empty $X \in \mathit{Prop}$ intersects $\bigcup \mathcal{E}$, and so $(\bigcup \mathcal{E}){\uparrow} = \mathbb{Q}$. The case of $\downarrow$ is similar (or it can be derived from the $\uparrow$ case, using the equation $S{\downarrow} = \mathbb{Q} \setminus ((\mathbb{Q} \setminus S){\uparrow})$ for $S \subseteq \mathbb{Q}$). ■

Now let $\mathcal{L}$ consist of two binary relation symbols, P and $\sim$. (The two uses of $\sim$ will be distinguished by context.) We define an $\mathcal{L}$-premodel on $\mathcal{S}$ by putting, for each $a, b \in \mathbb{Q}$,

- $|{\sim}|^{\mathcal{M}}(a, b) = \begin{cases} \mathbb{Q}, & \text{if } a \sim b, \\ \emptyset, & \text{otherwise;} \end{cases}$
- $|P|^{\mathcal{M}}(a, b) = \begin{cases} \mathbb{Q}, & \text{if } a \sim b, \\ \text{some non-empty interval} & \\ [b, c) \text{ not containing } a, & \text{otherwise.} \end{cases}$

Note that *Prop* contains $|{\sim}|^{\mathcal{M}}(a, b)$ and $|P|^{\mathcal{M}}(a, b)$ for all $a, b \in \mathbb{Q}$, as required. The definition ensures that $b \in |P|^{\mathcal{M}}(a, b)$ for all b, while $a \in |P|^{\mathcal{M}}(a, b)$ iff $a \sim b$.

Proposition 5.2. $\mathcal{M}$ *does not validate* $\forall x \forall y Pxy \to \forall y \forall x Pxy$.

Proof. We show that for any $f \in {}^{\omega}U$,

$$|\forall x \forall y Pxy| f = \mathbb{Q} \quad \text{while} \quad |\forall y \forall x Pxy| f = \emptyset.$$

Now $|\forall y Pxy| f = \big[\bigcap_{b \in \mathbb{Q}} Eb \Rightarrow |P|(fx, b)\big]{\downarrow}$. But for any b,

$$Eb \Rightarrow |P|(fx, b) = \{b\} \Rightarrow |P|(fx, b) = \mathbb{Q},$$

since $b \in |P|(fx, b)$. Hence $|\forall y Pxy| f = \mathbb{Q}{\downarrow} = \mathbb{Q}$. It follows that for any f, $|\forall x \forall y Pxy| f = [\bigcap_{a \in \mathbb{Q}} Ea \Rightarrow \mathbb{Q}]{\downarrow} = \mathbb{Q}$ as well.

On the other hand, $|\forall x Pxy|f = \big[\bigcap_{a\in\mathbb{Q}} Ea \Rightarrow |P|(a, fy)\big]{\downarrow}$. But

$$Ea \Rightarrow |P|(a, fy) = \mathbb{Q} \setminus \{a\} \cup |P|(a, fy) = \begin{cases} \mathbb{Q}, & \text{if } a \sim fy, \\ \mathbb{Q} \setminus \{a\}, & \text{otherwise,} \end{cases}$$

so $|\forall x Pxy|f = \big[\bigcap_{a\not\sim fy} \mathbb{Q} \setminus \{a\}\big]{\downarrow} = (fy/{\sim}){\downarrow} = \emptyset$ by Lemma 5.1.

It follows that for any f, $|\forall y \forall x Pxy|f = [\bigcap_{b\in\mathbb{Q}} \mathbb{Q} \setminus \{b\} \cup \emptyset]{\downarrow} = \emptyset{\downarrow} = \emptyset$ as well. ■

Notice that this proof shows that $\mathcal{M}$ is *non-Kripkean*: since $\emptyset \neq fy/{\sim}$, we have

$$|\forall x Pxy|f \neq \bigcap_{a\in\mathbb{Q}} Ea \Rightarrow |P|(a, fy).$$

We now have to show that the premodel $\mathcal{M}$ is actually a *model*, i.e. $|\varphi|^{\mathcal{M}} f$ is always an admissible set. For this we recall that formulas φ, ψ are $\mathcal{M}$-*equivalent* if $|\varphi| = |\psi|$ in this $\mathcal{M}$. The above proof of Proposition 5.2 shows that the formulas $\forall x \forall y Pxy$ and $\forall y \forall x Pxy$ are not only admissible, they are $\mathcal{M}$-equivalent to the formulas $\top$ and $\bot$, respectively. It turns out that every formula φ is $\mathcal{M}$-equivalent to some formula ψ that lacks quantifiers. But as observed earlier, all quantifier-free formulas are admissible, so we get $|\varphi| = |\psi| \in \mathit{Prop}$.

The proof of this "quantifier-eliminability" will be given next. A naive approach would require us to express some basic equality types of tuples, by which we mean conjunctions of formulas of the form $x \approx y$ and $x \not\approx y$. This has to be handled carefully as we also need equality not to be definable. So in the example, $\sim$ plays the role of equality. It is a "weak equality", with only $|\exists y(x \sim y)|^{\mathcal{M}} f = \mathbb{Q}$, rather than $|\exists y(x \sim y)|^{\mathcal{M}} f = Efx{\uparrow}$, so it does not add to the admissible sets (cf. Remark 4.7). But it is still expressive enough to allow quantifier elimination, as we now show.

PROPOSITION 5.3. *Every formula is $\mathcal{M}$-equivalent to a quantifier-free formula, and hence is admissible in $\mathcal{M}$.*

PROOF. Let us say that a formula φ is *reducible* if it is $\mathcal{M}$-equivalent to a quantifier-free formula. We show that every φ is reducible, by induction on φ. In the proof, we write '$\mathcal{M}$-equivalent' simply as 'equivalent'.

Note that *any formula that is equivalent to a reducible one is itself reducible*, a fact that will be used repeatedly. To begin with, any formula is equivalent to one formed from atomic formulas by the propositional connectives (including $\Box$) and the quantifier $\exists$, so we can suppose without loss of generality that φ has this form.

If φ is atomic, we are given the reducibility. The set of reducible formulas is clearly closed under the Boolean connectives. It is also closed under $\Box$, since $\Box\varphi$ is equivalent to the reducible $\top$ when $R = \emptyset$, and equivalent to φ itself when R is the identity relation.

Assume that φ is reducible. We will prove that $\exists x\varphi$ is reducible. Inductively, there is a quantifier-free formula ψ equivalent to φ, and so $\exists x\varphi$ is reducible if the equivalent $\exists x\psi$ is reducible. Thus we can suppose that φ is quantifier-free. But then there is a quantifier-free ψ in disjunctive normal form that is tautologically equivalent to φ, and hence equivalent to φ in $\mathcal{M}$. Again, $\exists x\varphi$ will be reducible if the equivalent $\exists x\psi$ is. Thus we can suppose that φ is in disjunctive normal form.

So, suppose that φ is $\varphi_1 \vee \cdots \vee \varphi_n$, where each φ_i is a conjunction of *literals*, i.e. atomic and negated-atomic formulas. If each $\exists x\varphi_i$ is reducible, then so is $\exists x\varphi_1 \vee \cdots \vee \exists x\varphi_n$, which is equivalent to $\exists x(\varphi_1 \vee \cdots \vee \varphi_n)$ by Lemma 3.5(1), so $\exists x\varphi$ will be reducible. Hence we can suppose that φ is a conjunction of literals.

Next we can split off the conjuncts of φ in which x does not occur. For, if φ is equivalent to $\psi\wedge\theta$ with ψ a literal not containing x, and $\exists x\theta$ is reducible, then so is $\psi \wedge \exists x\theta$, which is equivalent to $\exists x(\psi \wedge \theta)$ by admissibility of ψ and Lemma 3.5(2), hence equivalent to $\exists x\varphi$. So we can suppose that x occurs in each conjunct of φ.

Similarly, we can delete $P(x,x)$ and $x \sim x$ if they occur as conjuncts of φ, since each is equivalent to $\top$ by the definitions of $|{\sim}|^{\mathcal{M}}$ and $|P|^{\mathcal{M}}$, and $\exists x(\top \wedge \theta)$ is equivalent to $\exists x\theta$. Moreover, if the negation of $P(x,x)$ or $x \sim x$ occurs in φ then we are done, since $\exists x(\bot \wedge \theta)$ is equivalent to the reducible $\bot$. Finally, $y \sim x$ with y different to x can be replaced by the equivalent $x \sim y$. So altogether we can suppose that we are dealing with a formula of the form $\exists x\varphi$, where

$$\begin{aligned}\varphi \quad = \quad & \bigwedge_i P(x,y_i) \wedge \bigwedge_j P(z_j,x) \wedge \bigwedge_k \neg P(x,u_k) \wedge \bigwedge_l \neg P(v_l,x) \\ & \qquad\qquad\qquad\qquad \wedge \bigwedge_m (x \sim s_m) \wedge \bigwedge_n \neg(x \sim t_n),\end{aligned}$$

all variables y_i, z_j, etc are distinct from x, and each $\bigwedge$ could be empty. Now for any $f \in {}^{\omega}U$, we have

$$\begin{aligned}|\exists x\varphi|f \quad = \quad & \Big[\bigcup_{a\in\mathbb{Q}} \Big(Ea \cap \bigcap_i |P|(a, fy_i) \cap \bigcap_j |P|(fz_j, a) \\ & \cap \bigcap_k (\mathbb{Q} \setminus |P|(a, fu_k)) \cap \bigcap_l (\mathbb{Q} \setminus |P|(fv_l, a)) \\ & \cap \bigcap_m |{\sim}|(a, fs_m) \cap \bigcap_n (\mathbb{Q} \setminus |{\sim}|(a, ft_n)) \Big)\Big]{\uparrow}.\end{aligned}$$

Any empty intersection here is interpreted as $\mathbb{Q}$. Now $Ea = \{a\}$ for any $a \in \mathbb{Q}$. So

$$\begin{aligned}|\exists x\varphi|f = \Big\{a \in \mathbb{Q} :\ & a \in \bigcap_i |P|(a, fy_i) \cap \bigcap_j |P|(fz_j, a) \\ & \cap \bigcap_k \big(\mathbb{Q} \setminus |P|(a, fu_k)\big) \cap \bigcap_l \big(\mathbb{Q} \setminus |P|(fv_l, a)\big) \\ & \cap \bigcap_m |{\sim}|(a, fs_m) \cap \bigcap_n \big(\mathbb{Q} \setminus |{\sim}|(a, ft_n)\big)\Big\}{\uparrow}.\end{aligned}$$

Observe now that

- $\{a \in \mathbb{Q} : a \in |P|^{\mathcal{M}}(a, b)\} = \{a \in \mathbb{Q} : a \in |{\sim}|^{\mathcal{M}}(a, b)\} = b/{\sim}$ for any $b \in \mathbb{Q}$,
- $\{b \in \mathbb{Q} : b \in |P|^{\mathcal{M}}(a, b)\} = \mathbb{Q}$ for any $a \in \mathbb{Q}$.

So the set $|\exists x\varphi|f$ above is

$$\begin{aligned}\Big[\bigcap_i (fy_i/{\sim})\ \cap \bigcap_j \mathbb{Q}\ \cap \bigcap_k (\mathbb{Q} \setminus (fu_k/{\sim}))\ \cap \bigcap_l \emptyset \\ \cap \bigcap_m (fs_m/{\sim})\ \cap \bigcap_n (\mathbb{Q} \setminus (ft_n/{\sim}))\Big]{\uparrow}.\end{aligned}$$

If the l-conjunction is non-empty — a condition determined by φ and independent of f — this set is $\emptyset$, and so $\exists x\varphi$ is equivalent to $\bot$. We are done. Otherwise, write Y for the set of all variables y_i, s_m above, and write Z for the set of all variables u_k, t_n. Then

$$\begin{aligned}|\exists x\varphi|f &= \Big[\bigcap_{y \in Y} (fy/{\sim})\ \cap \bigcap_{z \in Z} (\mathbb{Q} \setminus (fz/{\sim}))\Big]{\uparrow} \\ &= \Big[\bigcap_{y \in Y} (fy/{\sim})\ \setminus \bigcup_{z \in Z} (fz/{\sim})\Big]{\uparrow}.\end{aligned}$$

The set in square brackets here is a Boolean combination of $\sim$-equivalence classes. It is therefore of the form $\bigcup \mathcal{E}$ for some set $\mathcal{E}$ of $\sim$-classes. So by Lemma 5.1, the $\uparrow$ of the set belongs to *Prop*. This shows that $\exists x\varphi$ is admissible. The proof that it is reducible involves two cases, syntactically determined by φ:

- If $Y = \emptyset$, then $|\exists x\varphi|f = \mathbb{Q}$ for all f, because there are infinitely many $\sim$-classes in $\mathbb{Q}$ and only finitely many of them are eliminated by the Z-term. So $\exists x\varphi$ is equivalent to $\top$ in this case.

- if $Y \neq \emptyset$, then $|\exists x\varphi|f$ is $\mathbb{Q}$ if all the fy are $\sim$-equivalent and no fz is $\sim$-equivalent to them: for then, the set inside the square brackets is a single
$\sim$-equivalence class, so its $\uparrow$ is $\mathbb{Q}$. Otherwise, $|\exists x\varphi|f$ is $\emptyset$. Thus, for any $f \in {}^{\omega}U$,
$$|\exists x\varphi|f = \Big| \bigwedge_{y,y' \in Y} y \sim y' \wedge \bigwedge_{y \in Y, z \in Z} \neg(y \sim z) \Big| f.$$
So $\exists x\varphi$ is equivalent to this quantifier-free formula if $Y \neq \emptyset$ (and, as one can see, if $Y = \emptyset$ as well).

This completes the proof of Proposition 5.3, and hence the proof that $\mathcal{M}$ is a model. ■

6. Completeness and the Barcan Formulas

Let L be any (consistent) normal propositional modal logic. For a given signature $\mathcal{L}$, let $\mathrm{Q}^{-}\mathrm{L}$ be the smallest set of $\mathcal{L}$-formulas that includes

- all tautologies,
- all $\mathcal{L}$-substitution-instances of L-theorems,
- the schemata UI°, UD and VQ,

and is closed under

- detachment for material implication,
- the rule of Necessitation: from φ infer $\Box\varphi$, and
- the rule UG.

Now in the last section we defined two models for $\mathcal{L} = \{P, \sim\}$, call them $\mathcal{M}_0$ and $\mathcal{M}_1$, with $R = \emptyset$ and $R =$ the identity relation, respectively. We noted that the underlying propositional frame (W, R) of one of these models validates L, by the result of [5]. But then this model itself validates all $\mathcal{L}$-substitution-instances of L-theorems, by an argument given in the proof of [3, Theorem 2]. From the soundness results we have proved, and the evident soundness of Necessitation in any premodel, it then follows that this model validates $\mathrm{Q}^{-}\mathrm{L}$, while falsifying CQ.

It is notable that both the "Barcan formula"

BF $\forall x\Box\varphi \to \Box\forall x\varphi$

and its converse

CBF $\Box\forall x\varphi \to \forall x\Box\varphi$

are valid in $\mathcal{M}_0$ and $\mathcal{M}_1$. This follows from the fact that $\Box\psi$ is equivalent to $\top$ in $\mathcal{M}_0$, and to ψ in $\mathcal{M}_1$.

It turns out that for any $\mathcal{L}$, the logic Q$^-$L is complete for the class of all $\mathcal{L}$-models validating L (i.e. validating all $\mathcal{L}$-substitution-instances of L-theorems). This can be shown by a Henkin-model construction which reveals that the axioms UI°, UD and VQ, together with the rule UG, exactly capture the $\forall$-semantics

$$|\forall x\varphi| = \bigsqcap_{a\in U} Ea \Rightarrow |\varphi(a)|$$

of the $\mathcal{L}$-models we have used.

The converse Barcan formula is valid in any $\mathcal{L}$-model satisfying the *expanding domains* condition

$$wRv \quad \text{implies} \quad Dw \subseteq Dv, \tag{6.1}$$

equivalent to the requirement that $Ea \subseteq [R]Ea$ for all $a \in U$.

The logic Q$^-$L+CBF is complete for the class of its expanding domain models. But it is also complete for the class of its models that have *constant domains*:

$$wRv \quad \text{implies} \quad Dw = Dv. \tag{6.2}$$

This last claim may raise the eyebrows of some readers who are used to thinking of (6.2) as a condition that also validates the Barcan formula, which is typically not derivable in Q$^-$L+CBF. But the point is that BF can only be shown to be valid in the presence of (6.2) when the model is *Kripkean* in the sense of (4.3), in which case it also validates CQ.

The schema CQ is not a theorem of Q$^-$L+CBF+BF, as the models $\mathcal{M}_0$ and $\mathcal{M}_1$ show. The logic Q$^-$L+CBF+BF+CQ can be shown to be complete for its class of constant-domain *Kripkean* models. These results indicate that the main role of the Barcan formula in possible-worlds model theory is not to provide models that have constant domains, but rather to ensure that in a Henkin-style construction, the quantifier $\forall$ can be given the Kripkean interpretation via $\bigsqcap$.

Justification of all these claims will be presented elsewhere.

Acknowledgements. We thank Ed Mares for discussions on this and related topics of admissible-set semantics, and the referees for helpful comments. This research was supported by a grant from the Marsden Fund of the Royal Society of New Zealand.

References

[1] Fine, Kit, 'The permutation principle in quantificational logic', *Journal of Philosophical Logic* 12: 33–37, 1983.

[2] Garson, James W., 'Quantification in modal logic', in Gabbay, D.M., Guenthner, F. (eds.), *Handbook of Philosophical Logic*, Vol. 3, 2nd edn., Kluwer Academic Publishers, Dordrecht, 2001, pp. 267–323.

[3] Goldblatt, Robert, Mares, Edwin D., 'A general semantics for quantified modal logic', in Governatori, Guido, Hodkinson, Ian, Venema, Yde (eds.), *Advances in Modal Logic*, Vol. 6, College Publications, London, 2006, pp. 227–246. `http://www.aiml.net/volumes/volume6/`.

[4] Kripke, Saul A., 'Semantical considerations on modal logic', *Acta Philosophica Fennica*, 16: 83–94, 1963.

[5] Makinson, D.C., 'Some embedding theorems for modal logic', *Notre Dame Journal of Formal Logic*, 12: 252–254, 1971.

[6] Mares, Edwin D., Goldblatt, Robert, 'An alternative semantics for quantified relevant logic', *The Journal of Symbolic Logic*, 71(1): 163–187, 2006.

Robert Goldblatt
Centre for Logic, Language and Computation,
Victoria University, P.O. Box 600, Wellington, New Zealand
`rob.goldblatt@mcs.vuw.ac.nz`

Ian Hodkinson
Department of Computing,
Imperial College London,
London, SW7 2AZ, U.K.
`imh@doc.ic.ac.uk`

F. Poggiolesi

The Method of Tree-hypersequents for Modal Propositional Logic

Abstract. In this paper we present a method, that we call the tree-hypersequent method, for generating contraction-free and cut-free sequent calculi for modal propositional logics. We show how this method works for the systems K, KD, K4 and KD4, by giving a sequent calculus for these systems which are normally presented in the Hilbert style, and by proving all the main results in a purely syntactical way.

Keywords: Contraction-free, Cut-free, Hypersequents, Modal logic, Sequent Calculus, Tree-hypersequents.

1. Introduction

One of the open problems of modal propositional logic consists in the lack of a *good* sequent calculus for (at least) its main systems, where we understand a good sequent calculus to be one that satisfies certain requirements, mainly listed by [1], [4] and [12], the principal ones being[1]:

Subformula Property: we should be able to associate to every proof d of the sequent calculus, a proof d^* of the same final sequent, in which each formula is a subformula of the formulas occurring in the final sequent.

Semantic Purity: the sequent calculus should not make any use of explicit semantic elements, such as possible words or truth values.

Explicitness: logical rules should exhibit the constant they introduce only in the conclusion.

Separation: logical rules should not exhibit any constant other than the one they introduce.

Symmetry: each constant of the language of the sequent calculus should have at least two logical rules: one which introduces it on the left side of the sequent, one which introduces it on the right.

Invertibility: for each of the rules of the calculus it should hold that not only the conclusion is derivable from the premise(s), but also the premise(s) from the conclusion.

---

[1] The list is by no means exhaustive but our aim here is not to discuss the properties which define a good sequent calculus.

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

The first attempts made at solving such a problem use the standard sequent calculus (see, for example, [3], [6] and [11]), but they are generally not good sequent calculi, as they do not satisfy the subformula property and their rules are not explicit, nor invertible, nor separate nor symmetric.

More recently research has been oriented towards finding methods which can generate extensions of the standard sequent calculus. These methods can be divided in two groups: in one group there are methods which generate purely syntactic sequent calculi, the most important of which are the method of hypersequents [1] and the method of display logic [12]. In the other group there are methods which extend the standard sequent calculus by adding explicit semantic parameters. Some examples of this group are the calculus given by G. Mints [7] and the calculus given by S. Negri [8]. None of the methods proposed so far can generate a calculus which satisfies all the desired properties: the calculi of the second group for the very reason that they include explicit semantic parameters, and the calculi of the first group because they lack other properties normally required.

When this work was essentially completed, we were informed of the existence of a further method, created by R. Kashima [5], and then developed by K. Brünnler [2], which is called nested-sequents method (by Kashima), or deep-sequents method (by Brünnler), and which makes use of the same notion of sequent that we will present in this article, though using a different notation. The main difference between the nested (or deep) — sequents method and the one that we will introduce below, consists in the fact that here the proof of cut-elimination is developed in a purely syntactic way (see section 5), as well as the proof of admissibility of the several structural rules (see section 3), while Kashima and Brünnler use semantics instead (Brünnler sketches a syntactic cut-elimination procedure for the system K in his paper). Moreover we apply the method for capturing the D — axiom, while Kashima and Brünnler do not, and the cut-rule, as well as the related notions of equivalence position and product, are introduced here for the first time. Finally, although Kashima's calculi are not contraction-free and invertible, ours (and those of Brünnler) are.

We want also to point out that although the nested (deep, tree-hyper) — sequents method has doubtless several common points with the methods of Negri and Mints, it also has one important difference: it does not use any labels. A clarification about the property of "not using labels" and the advantages of having such property can be found in [2].

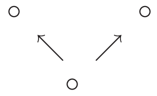
Having clarified the relationships with the recent related works, we can finally introduce the method that we will call tree-hypersequent method. In order to do it, we begin by informally explaining what a tree-hypersequent is and we do this by constructing such an object step-by-step. Let us, then, recall, first of all, the simple notion of empty *hypersequent*: an empty hypersequent is a syntactic object of the following form:

$$\overbrace{-/-/-}^{n}$$

which is to say: n slashes which separate $n+1$ dashes. If the order of the dashes is taken on account (as is not standardly done), we can look to this entire structure as a tree-frame in Kripke semantics, where the dashes are meant to be the worlds of the tree-frame and the slashes the relations between worlds in the tree-frame. Following this analogy the dash at distance one in an empty hypersequent denotes a world at distance one in the corresponding tree-frame, a dash at distance two denotes a world a distance two in the corresponding tree-frame, and so on.

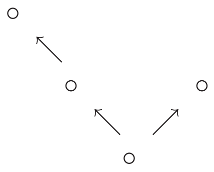
In a tree-frame, at every distance, except the first one, we may find n different possible worlds: how can we express this fact in our syntactic object? We separate different dashes with a semi-colon and obtain, in this way, the notion of empty tree-hypersequent. So an example of an empty tree-hypersequent is an object of the following form (the figure on the left):

$$-\,/\,-\,;\,- \qquad\qquad \rightsquigarrow$$



which corresponds to a tree-frame (the figure on the right) with a world at distance one related with two different worlds at distance two. Another example of an empty tree-hypersequent is an object of the following form (the figure on the left):

$$-/(-/-);- \qquad\qquad \rightsquigarrow$$



which corresponds to a tree-frame (the figure on the right) with a world at distance one related with two different worlds at distance two, each of which is, in turn, related with another world at distance three. Finally, in order to obtain a tree-hypersequent we fill the dashes with sequents which are objects of the form $M \Rightarrow N$, where M and N are multisets of formulas.

In the next section we will show how to apply this method in order to obtain calculi for: (i) the basic system K, which is valid and complete in all the frames, (ii) the extensions of K which contain one or both the D axiom and the 4 axiom, namely KD, $K4$ and $KD4$. We remind the reader that the D axiom has the form: $\Box\alpha \rightarrow \Diamond\alpha$, and its characteristic frame property is seriality; the 4 axiom, instead, has the form: $\Box\alpha \rightarrow \Box\Box\alpha$, and its characteristic frame property is transitivity. In the third section we will show which rules are admissible in these calculi, in the fourth section we will prove that they are valid and complete in the corresponding Hilbert-style system, and in the fifth section we will finally prove the cut-elimination theorem for all of them.

2. The Calculi CSK*

We define the modal propositional language $\mathcal{L}^{\Box}$ in the following way:

atoms: p_0, p_1, ...
logical constant: $\Box$
connectives: $\neg$, $\vee$

The other classic connectives can be defined as usual, as well as the constant $\Diamond$ and the formulas of the modal language $\mathcal{L}^{\Box}$.

`Syntactic Conventions`:

α, β, ...: formulae,
M, N, ...: finite multisets of formulae,
Γ, Δ, ...: sequents (SEQ). The empty sequent ($\Rightarrow$) is included.
G, H, ...: tree-hypersequents (THS).
$\underline{X}$, $\underline{Y}$, ...: finite multisets of tree-hypersequents (MTHS), $\emptyset$ included.

We point out that for the sake of brevity we will use the following notation: given $\Gamma \equiv M \Rightarrow N$ and $\Pi \equiv P \Rightarrow Q$, we will write α, Γ instead of $\alpha, M \Rightarrow N$ and Γ, α instead of $M \Rightarrow N, \alpha$, as well as $\Gamma \centerdot \Pi$ instead of $M, P \Rightarrow N, Q$.

Definition 2.1. The notion of tree-hypersequent is inductively defined in the following way:

- if $\Gamma \in$ SEQ, then $\Gamma \in$ THS,
- if $\Gamma \in$ SEQ and $\underline{X} \in$ MTHS, then $\Gamma / \underline{X} \in$ THS.

Given the above definition, an example of tree-hypersequent is the following one:

$$\Delta/(\Gamma/\Sigma);(\Gamma_1/(\Sigma_1/\Theta);\Sigma_2)$$

DEFINITION 2.2. The intended interpretation of a tree-hypersequent is:

- $(M \Rightarrow N)^{\tau}\colon = \bigwedge M \rightarrow \bigvee N$,
- $(\Gamma/G_1;...;G_n)^{\tau}\colon = \Gamma^{\tau} \vee \Box G_1^{\tau} \vee ... \vee \Box G_n^{\tau}$.

Given the above definition, the intended interpretation of the tree-hypersequent of the example above, is:

$$\Delta^{\tau} \vee \Box(\Gamma^{\tau} \vee \Box\Sigma^{\tau}) \vee \Box(\Gamma_1^{\tau} \vee \Box(\Sigma_1^{\tau} \vee \Box\Theta^{\tau}) \vee \Box\Sigma_3^{\tau})$$

In order to display the rules of the calculi, we will use the notation $G[\Gamma]$ (or $G[H]$) to refer to a tree-hypersequent together with a specific occurrence in it of a sequent Γ (or a tree-hypersequent H). You may think, if you like, of $G[\ ]$ as a "tree-hypersequent with one hole," an object which becomes a real tree-hypersequent whenever a sequent Γ (or a tree-hypersequent H) is appropriately put into the hole.

We can even use the notation $G[\Gamma][[\Box\alpha,\Sigma]]$, where $[[\Box\alpha,\Sigma]]$ represents all the sequents in G which are successive to Γ and contain(ed) the formula $\Box\alpha$ on the left side.

The calculus *CSK* is composed of:

Initial Tree-hypersequents.

$G\ [p,\ \Gamma,\ p]$

Propositional Rules.

$$\frac{G[\Gamma,\alpha]}{G[\neg\alpha,\Gamma]}\,\neg A \qquad \frac{G[\alpha,\Gamma]}{G[\Gamma,\neg\alpha]}\,\neg K$$

$$\frac{G[\alpha,\beta,\Gamma]}{G[\alpha\wedge\beta,\Gamma]}\,\wedge A \qquad \frac{G[\Gamma,\alpha] \qquad G[\Gamma,\beta]}{G[\Gamma,\alpha\wedge\beta]}\,\wedge K$$

Modal Rules.

$$\frac{G[\Box\alpha,\Gamma/(\alpha,\Sigma/\underline{X});\underline{X}']}{G[\Box\alpha,\Gamma/(\Sigma/\underline{X});\underline{X}']}\,\Box A \qquad \frac{G[\Gamma/\Rightarrow\alpha;\underline{X}]}{G[\Gamma,\Box\alpha/\underline{X}]}\,\Box K$$

We underline that the addition of the formula $\Box\alpha$ to the left side of the sequent of the premise of the rule $\Box A$ only serves to make the rule invertible.

This is analogous to the repetition of the formula $\forall x(\alpha)$ in the premise of the rule which introduces the universal quantification, in some versions of the sequent calculus of first-order logic.

In order to introduce the cut-rule, we firstly need two new notions:

Definition 2.3. Given two tree-hypersequents, $G[\Gamma]$ and $G^{'}[\Gamma^{'}]$ together with an occurrence of a sequent in each, the relation of *equivalent position* between two of their sequents, in this case Γ and $\Gamma^{'}$, $G[\Gamma] \sim G^{'}[\Gamma^{'}]$, is defined inductively in the following way:

- $\Gamma \sim \Gamma^{'}$
- $\Gamma/\underline{X} \sim \Gamma^{'}/\underline{X}^{'}$
- If $H[\Gamma] \sim H^{'}[\Gamma^{'}]$, then $\Delta/H[\Gamma];\underline{X} \sim \Delta^{'}/H^{'}[\Gamma^{'}];\underline{X}^{'}$

Intuitively, given two tree-hypersequents, $G[\Gamma]$ and $G^{'}[\Gamma^{'}]$ together with an occurrence of a sequent in each, the relation of equivalent position between two of their sequents holds when, by considering $G[\Gamma]$ and $G^{'}[\Gamma^{'}]$ as trees, and Γ and $\Gamma^{'}$ as nodes of the trees, the two nodes have the same height in their respective trees. Consider for example the two tree-hypersequents $G \equiv \Delta/(\Gamma/\Sigma);(\Gamma_1/(\Sigma_1/\Theta);\Sigma_2)$ and $G^{'} \equiv \Delta^{'}/(\Gamma^{'}/\Sigma^{'});(\Gamma^{'}_1);(\Gamma^{'}_2/(\Sigma^{'}_1/\Theta^{'}))$. Then Γ and $\Gamma^{'}$ are in equivalent position, as are Γ and $\Gamma^{'}_1$, or Θ and $\Theta^{'}$.

Definition 2.4. Given two tree-hypersequents $G[\Gamma]$ and $G^{'}[\Gamma^{'}]$ together with an occurrence of a sequent in each, such that $G[\Gamma] \sim G^{'}[\Gamma^{'}]$, the operation of *product*, $G[\Gamma] \otimes G^{'}[\Gamma^{'}]$, is defined inductively in the following way:

- $\Gamma \otimes \Gamma^{'} := \Gamma \centerdot \Gamma^{'}$
- $(\Gamma/\underline{X}) \otimes (\Gamma^{'}/\underline{X}^{'}) := \Gamma \centerdot \Gamma^{'}/\underline{X};\underline{X}^{'}$
- $(\Delta/H[\Gamma];\underline{X}) \otimes (\Delta^{'}/H^{'}[\Gamma^{'}];\underline{X}^{'}) := \Delta \centerdot \Delta^{'}/(H[\Gamma] \otimes H^{'}[\Gamma^{'}]);\underline{X};\underline{X}^{'}$

Cut rule.

Given two tree-hypersequents $G[\Gamma,\alpha]$ and $G^{'}[\alpha,\Pi]$ together with an occurrence of a sequent in each, such that $G[\Gamma,\alpha] \sim G^{'}[\alpha,\Pi]$, the cut rule is:

$$\frac{G[\Gamma,\alpha] \quad G^{'}[\alpha,\Pi]}{G \otimes G^{'}[\Gamma \centerdot \Pi]}\, cut_{\alpha}$$

As the reader can easily see from the above definition, the cut rule should respect two important criteria. The first one says that, given two tree-hypersequents, we can cut on any two sequents belonging to them provided that they are in equivalent position. The second one says that after the cut the two tree-hypersequents should not be randomly mixed but according to the inductive definition of product. We underline that these two criteria are fundamental because they serve to assure that the result of a cut between two tree-hypersequents is still a *tree*-hypersequent, which is to say the tree shape is kept.

The corresponding rules of axiom D and axiom 4 are, respectively, the following two ones:

$$\frac{G[\Gamma/\Rightarrow]}{G[\Gamma]}\,ser. \qquad \frac{G[\Box\alpha,\Gamma/(\Box\alpha,\Sigma/\underline{X});\underline{X}^{'}]}{G[\Box\alpha,\Gamma/(\Sigma/\underline{X});\underline{X}^{'}]}\,tran.$$

Sequent calculi $CSKD$, $CSK4$ and $CSKD4$ can be obtained by adding to the basic sequent calculus CSK one or both the above rules.
In next section we will use the notation CSK^* (or, if necessary, $CSKD^*$ and $CSK4^*$) to denote the calculus CSK ($CSKD$, $CSK4$) and its extensions.

3. Admissibility of the Structural Rules

In this section we will show which structural rules are admissible in calculi CSK^*. Moreover, in order to show that the two rules of contraction are height-preserving admissible we will show that all the logical and modal rules are height-preserving invertible. The proof of the admissibility of the cut-rule will be shown in the fifth section.

Definition 3.1. We associate to each proof d in CSK^* a natural number $h(d)$ (height). Intuitively, the height corresponds to the length of the longest branch in a tree-proof d, minus one. However we omit the standard inductive definition.

Definition 3.2. $d \vdash^n G$ means that d is a proof of G in CSK^*, with $h(d) \leq n$. We write $\vdash^{\langle n \rangle} G$ for: "there exists a proof d such that $d \vdash^n G$."

Definition 3.3. Let G be a tree-hypersequent and $G^{'}$ be the result of the application of a certain rule $\mathcal{R}$ on G. We say that this rule $\mathcal{R}$ is *height-preserving admissible* when:

$$d \vdash^n G \quad \Rightarrow \quad \exists d^{'}(d^{'} \vdash^n G^{'})$$

We call a rule, $\mathcal{R}$, which transforms a tree-hypersequent G into a tree-hypersequent $G^{'}$, *admissible* when:

$$d \vdash^n G \quad \Rightarrow \quad \exists d^{'}(d^{'} \vdash G^{'})$$

Observation 3.4. In the sequent calculus for classical logic, we usually say that a formula of a sequent is principal in a rule when the rule operates on that formula. In a similar way we will call a sequent(s) principal in a tree-hypersequent when a certain rule operates on that sequent(s). In the following proofs of the (height-preserving) admissibility of structural rules and height-preserving invertibility of logical and modal rules, we will consider only these cases where the sequent(s) is (are) principal. All the other cases are dealt with easily, as shown in the two lemmas 3.14 and 3.15 which are proved at the end of the current section.

LEMMA 3.5. *Tree-hypersequents of the form $G[\alpha, \Gamma, \alpha]$, with α an arbitrary modal formula, are derivable in CSK^*.*

PROOF. By straightforward induction on α. ■

LEMMA 3.6. *The rule:*

$$\frac{G}{\Rightarrow /G}\ {\scriptstyle RN}$$

is height-preserving admissible in CSK^*.

PROOF. By induction on the derivation of the premise.

If G is an initial tree-hypersequent, then $\Rightarrow /G$ is also an initial tree-hypersequent.

If G is inferred by a logical rule, then the inference is clearly preserved. We will give an example using the logical rule $\neg K$:

$$\frac{{}^{\langle n-1\rangle}G[\alpha, \Gamma]}{{}^{\langle n\rangle}G[\Gamma, \neg\alpha]}\ {\scriptstyle \neg K} \qquad \rightsquigarrow^{2} \qquad \frac{{}^{\langle n-1\rangle} \Rightarrow /G[\alpha, \Gamma]}{{}^{\langle n\rangle} \Rightarrow /G[\Gamma, \neg\alpha]}\ {\scriptstyle \neg K}$$

If G is inferred by the modal rules, these are clearly preserved. We will give an example using the modal rule $\Box K$:

$$\frac{{}^{\langle n-1\rangle}G[\Gamma/ \Rightarrow \alpha; \underline{X}]}{{}^{\langle n\rangle}G[\Gamma, \Box\alpha/\underline{X}]}\ {\scriptstyle \Box K} \qquad \rightsquigarrow$$

$$\frac{{}^{\langle n-1\rangle} \Rightarrow /G[\Gamma/ \Rightarrow \alpha; \underline{X}]}{{}^{\langle n\rangle} \Rightarrow /G[\Gamma, \Box\alpha/\underline{X}]}\ {\scriptstyle \Box K}$$

If, finally, G is inferred by rule *ser.* or rule *tran.*, these are clearly preserved. We will give an example using the rule *ser.*:

---

[2] The symbol $\rightsquigarrow$ means: the premise of the right side is concluded by induction hypothesis on the premise of the left side.

$$\frac{{}^{\langle n-1\rangle}G[\Gamma/\Rightarrow]}{{}^{\langle n\rangle}G[\Gamma]}\,ser. \quad\rightsquigarrow\quad \frac{{}^{\langle n-1\rangle}\Rightarrow/G[\Gamma/\Rightarrow]}{{}^{\langle n\rangle}\Rightarrow/G[\Gamma]}\,ser.$$

■

Lemma 3.7. *The rules of weakening:*

$$\frac{G[\Gamma]}{G[\alpha,\Gamma]}\,WA \qquad \frac{G[\Gamma]}{G[\Gamma,\alpha]}\,WK$$

are height-preserving admissible in CSK^*.

Proof. By straightforward induction on the derivation of the premise. ■

Lemma 3.8. *The rule of external weakening:*

$$\frac{G[\Gamma/\underline{X}]}{G[\Gamma/\underline{X};\Gamma]}\,EW$$

is height-preserving admissible in CSK^*.

Proof. By straightforward induction on the derivation of the premise. ■

Lemma 3.9. *The rule of merge:*

$$\frac{G[\Delta/(\Gamma/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}{G[\Delta/(\Gamma\centerdot\Pi/\underline{X_1};\underline{X_2});\underline{Y}]}\,merge$$

is height-preserving admissible in CSK^*.

Proof. By induction on the derivation of the premise.

If the premise is an initial tree-hypersequent, then so is the conclusion.
If the premise is inferred by a logical rule, this inference is preserved. As the rule of merge has two principal sequents, we should analyze the following two cases: one in which the logical rule has been applied to the sequent Γ, one in which the logical rule has been applied on the sequent Π. These two cases are similar; hence we will only sketch the proof for one of them, taking as example the logical rule $\neg K$:

$$\frac{{}^{\langle n-1\rangle}G[\Delta/(\alpha,\Gamma/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Gamma,\neg\alpha/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}\,\neg K \quad\rightsquigarrow$$

$$\frac{{}^{\langle n-1\rangle}G[\Delta/(\alpha,\Gamma\centerdot\Pi/\underline{X_1};\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Gamma\centerdot\Pi,\neg\alpha/\underline{X_1};\underline{X_2});\underline{Y}]}\,\neg K$$

If the premise is inferred by the modal rule $\Box K$ (for the rule *ser.* the treatment is analogous), then as in the case of logical rules, there are two symmetric cases to analyze. We will give an example of just one case:

$$\dfrac{{}^{\langle n-1\rangle}G[\Delta/(\Gamma/\Rightarrow\alpha;\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Gamma,\Box\alpha/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}\,{\scriptstyle\Box K}\qquad\rightsquigarrow$$

$$\dfrac{{}^{\langle n-1\rangle}G[\Delta/(\Gamma\centerdot\Pi/\Rightarrow\alpha;\underline{X_1};\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Gamma\centerdot\Pi,\Box\alpha/\underline{X_1};\underline{X_2});\underline{Y}]}\,{\scriptstyle\Box K}$$

Finally, in the case where the premise is inferred by the rule *tran.* (for the rule $\Box A$ the treatment is analogous), there are, for a simple combination of principal sequents, two pairs of analogous cases to analyze: on the one hand, *tran.* applied between Δ and Γ, and between Δ and Π; on the other hand, *tran.* applied between Γ and $\underline{X_1}$, and between Π and $\underline{X_2}$. We will examine one case from each pair:

(1)

$$\dfrac{{}^{\langle n-1\rangle}G[\Box\alpha,\Delta/(\Box\alpha,\Gamma/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Box\alpha,\Delta/(\Gamma/\underline{X_1});(\Pi/\underline{X_2});\underline{Y}]}\,{\scriptstyle tran.}\qquad\rightsquigarrow$$

$$\dfrac{{}^{\langle n-1\rangle}G[\Box\alpha,\Delta/(\Box\alpha,\Gamma\centerdot\Pi/\underline{X_1};\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Box\alpha,\Delta/(\Gamma\centerdot\Pi/\underline{X_1};\underline{X_2});\underline{Y}]}\,{\scriptstyle tran.}$$

(2)[3]

$$\dfrac{{}^{\langle n-1\rangle}G[\Delta/(\Box\alpha,\Gamma/(\Box\alpha,\Sigma/\underline{X_1'});\underline{X_1''});(\Pi/\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Box\alpha,\Gamma/(\Sigma/\underline{X_1'});\underline{X_1''});(\Pi/\underline{X_2});\underline{Y}]}\,{\scriptstyle tran.}\qquad\rightsquigarrow$$

$$\dfrac{{}^{\langle n-1\rangle}G[\Delta/(\Box\alpha,\Gamma\centerdot\Pi/(\Box\alpha,\Sigma/\underline{X_1'});\underline{X_1''};\underline{X_2});\underline{Y}]}{{}^{\langle n\rangle}G[\Delta/(\Box\alpha,\Gamma\centerdot\Pi/(\Sigma/\underline{X_1'});\underline{X_1''};\underline{X_2});\underline{Y}]}\,{\scriptstyle tran.}$$

■

LEMMA 3.10. *The following rule:*

$$\dfrac{G[\Gamma/(\Sigma/\underline{X});\underline{X'}]}{G[\Gamma/(\Rightarrow/\Sigma/\underline{X});\underline{X'}]}\,{\scriptstyle tran.2.}$$

is admissible in those calculi which contain the rule tran.

PROOF. By induction on the derivation of the premise. The cases where the premise is an initial tree-hypersequent or is preceded by a logical rule are trivial. We analyze the cases in which the last applied rule is one of the modal rules or is the rule *ser.* or is the rule *tran.*

$[\Box K]$ (for the rule *ser.* the treatment is analogous):

$$\dfrac{{}^{\langle n-1\rangle}G[\Gamma/\Rightarrow\alpha;(\Sigma/\underline{X});\underline{Y}]}{{}^{\langle n\rangle}G[\Gamma,\Box\alpha/(\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle\Box K}\qquad\rightsquigarrow$$

$$\dfrac{G[\Gamma/\Rightarrow\alpha;(\Rightarrow/\Sigma/\underline{X});\underline{Y}]}{G[\Gamma,\Box\alpha/(\Rightarrow/\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle\Box K}$$

---

[3] We take $\underline{X_1}\equiv(\Sigma/\underline{X_1'});\underline{X_1''}$

$$[\Box A{:}]\quad \dfrac{{}^{\langle n-1\rangle}G[\Box\alpha,\Gamma/(\alpha,\Sigma/\underline{X});\underline{Y}]}{{}^{\langle n\rangle}G[\Box\alpha,\Gamma/(\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle \Box A}\qquad\rightsquigarrow$$

$$\dfrac{\dfrac{\dfrac{G[\Box\alpha,\Gamma/(\Rightarrow/\alpha,\Sigma/\underline{X});\underline{Y}]}{G[\Box\alpha,\Gamma/(\Box\alpha\Rightarrow/\alpha,\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle AA}}{G[\Box\alpha,\Gamma/(\Box\alpha\Rightarrow/\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle \Box A}}{G[\Box\alpha,\Gamma/(\Rightarrow/\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle tran.}$$

$$[tran.{:}]\quad \dfrac{{}^{\langle n-1\rangle}G[\Box\alpha,\Gamma/(\Box\alpha,\Sigma/\underline{X});\underline{Y}]}{{}^{\langle n\rangle}G[\Box\alpha,\Gamma/(\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle tran.}\qquad\rightsquigarrow$$

$$\dfrac{\dfrac{\dfrac{G[\Box\alpha,\Gamma/(\Rightarrow/\Box\alpha,\Sigma/\underline{X});\underline{Y}]}{G[\Box\alpha,\Gamma/(\Box\alpha\Rightarrow/\Box\alpha,\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle AA}}{G[\Box\alpha,\Gamma/(\Box\alpha\Rightarrow/\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle tran.}}{G[\Box\alpha,\Gamma/(\Rightarrow/\Sigma/\underline{X});\underline{Y}]}\,{\scriptstyle tran.}$$

■

Lemma 3.11. *All the logical and modal rules of CSK^* are height-preserving invertible.*

Proof. The proof proceeds by induction on the derivation of the premise of the rule considered. The cases of logical rules are dealt with in the classical way. The only differences — the fact that we are dealing with tree-hypersequents, and the cases where the rule before the logical rule is $\Box A$ or $\Box K$ or *ser.* or *tran.* — are dealt with easily.

The rule ($\Box A$) is trivially height-preserving invertible since the premise is concluded by weakening from the conclusion, and weakening is height-preserving admissible.

We show in detail the invertibility of the rule ($\Box K$). If $G[\Gamma,\Box\alpha/\underline{X}]$ is an initial tree-hypersequent, then so is $G[\Gamma/\Rightarrow\alpha;\underline{X}]$. If $G[\Gamma,\Box\alpha/\underline{X}]$ is preceded by a logical rule $\mathcal{R}$, we apply the inductive hypothesis on the premise(s), $G[\Gamma^{'},\Box\alpha/\underline{X}]$ ($G[\Gamma^{''},\Box\alpha/\underline{X}]$) and we obtain derivation(s), of height $n-1$, of $G[\Gamma^{'}/\Rightarrow\alpha;\underline{X}]$ ($G[\Gamma''/\Rightarrow\alpha;\underline{X}]$). By applying the rule $\mathcal{R}$, we obtain a derivation of height n of $G[\Gamma/\Rightarrow\alpha;\underline{X}]$. If $G[\Gamma,\Box\alpha/\underline{X}]$ is of the form $G[\Box\beta,\Gamma,\Box\alpha/(\Sigma/\underline{X}^{'});\underline{X}^{''}]$ and is concluded by the modal rule $\Box A$, we apply the inductive hypothesis on $G[\Box\beta,\Gamma,\Box\alpha/(\beta,\Sigma/\underline{X}^{'});\underline{X}^{''}]$ and we obtain a derivation of height $n-1$ of $G[\Box\beta,\Gamma/\Rightarrow\alpha;(\beta,\Sigma/\underline{X}^{'});\underline{X}^{''}]$. By applying the rule $\Box A$, we obtain a derivation of height n of $G[\Box\beta,\Gamma/\Rightarrow\alpha;(\Sigma/\underline{X}^{'});\underline{X}^{''}]$. If $G[\Gamma,\Box\alpha/\underline{X}]$ is concluded by the rule *ser.* or *tran.* or by the modal rule $\Box K$ in which $\Box\alpha$ is not the principal formula, these cases are analogous to the one of $\Box A$. Finally, if $G[\Gamma,\Box\alpha/\underline{X}]$ is preceded by the modal rule $\Box K$ and $\Box\alpha$ is a principal formula, the premise of the last step gives the conclusion. ■

LEMMA 3.12. *The rules of contraction:*

$$\frac{G[\alpha,\alpha,\Gamma]}{G[\alpha,\Gamma]}\,CA \qquad \frac{G[\Gamma,\alpha,\alpha]}{G[\Gamma,\alpha]}\,CK$$

are height-preserving admissible in CSK^*.

PROOF. By induction on the derivation of the premise $G[\Gamma,\alpha,\alpha]$. We only analyze the case of the rule CK. The case of the rule CA is symmetric.

If $G[\Gamma,\alpha,\alpha]$ is an initial tree-hypersequent, so is $G[\Gamma,\alpha]$.

If $G[\Gamma,\alpha,\alpha]$ is preceded by a rule $\mathcal{R}$ which does not have any of the two occurrences of the formula α as principal, we apply the inductive hypothesis on the premise(s) $G^{'}[\Gamma^{'},\alpha,\alpha]$ ($G^{''}[\Gamma^{''},\alpha,\alpha]$), obtaining derivation(s) of height $n-1$ of $G^{'}[\Gamma^{'},\alpha]$ ($G^{''}[\Gamma^{''},\alpha]$). By applying the rule $\mathcal{R}$ we obtain a derivation of height n of $G[\Gamma,\alpha]$

$G[\Gamma,\alpha,\alpha]$ is preceded by a logical or modal rule and one of the two occurrences of the formula α is principal. Hence the rule which concludes $G[\Gamma,\alpha,\alpha]$ is a K-rule and we have to analyze the following three cases: $\neg K$, $\wedge K$, $\Box K$.

$$[\neg K]:\ \dfrac{^{\langle n-1\rangle}G[\beta,\Gamma,\neg\beta]}{^{\langle n\rangle}G[\Gamma,\neg\beta,\neg\beta]}\,\neg K \quad \dashrightarrow^{4} \quad \dfrac{\dfrac{^{\langle n-1\rangle}G[\beta,\beta,\Gamma]}{^{\langle n-1\rangle}G[\beta,\Gamma]}\,i.h.}{^{\langle n\rangle}G[\Gamma,\neg\beta]}\,\neg K$$

$$[\wedge K]:\ \dfrac{^{\langle n-1\rangle}G[\Gamma,\beta,\beta\wedge\gamma] \qquad ^{\langle n-1\rangle}G[\Gamma,\gamma,\beta\wedge\gamma]}{^{\langle n\rangle}G[\Gamma,\beta\wedge\gamma,\beta\wedge\gamma]}\,\wedge K \quad \dashrightarrow$$

$$\dfrac{\dfrac{^{\langle n-1\rangle}G[\Gamma,\beta,\beta]}{^{\langle n-1\rangle}G[\Gamma,\beta]}\,i.h \qquad \dfrac{^{\langle n-1\rangle}G[\Gamma,\gamma,\gamma]}{^{\langle n-1\rangle}G[\Gamma,\gamma]}\,i.h.}{^{\langle n\rangle}G[\Gamma,\beta\wedge\gamma]}\,\wedge K$$

$$[\Box K]:\ \dfrac{^{\langle n-1\rangle}G[\Gamma,\Box\beta/\Rightarrow\beta;\underline{X}]}{^{\langle n\rangle}G[\Gamma,\Box\beta,\Box\beta/\underline{X}]}\,\Box K \quad \dashrightarrow$$

$$\dfrac{\dfrac{\dfrac{^{\langle n-1\rangle}G[\Gamma/\Rightarrow\beta;\Rightarrow\beta;\underline{X}]}{^{\langle n-1\rangle}G[\Gamma/\Rightarrow\beta,\beta;\underline{X}]}\,merge}{^{\langle n-1\rangle}G[\Gamma/\Rightarrow\beta;\underline{X}]}\,i.h.}{^{\langle n\rangle}G[\Gamma,\Box\beta/\underline{X}]}\,\Box K$$

$\blacksquare$

LEMMA 3.13. *The rule of external contraction:*

$$\frac{G[\Gamma/(\Sigma/\underline{X_1});(\Sigma/\underline{X_2});\underline{Y}]}{G[\Gamma/(\Sigma/\underline{X_1};\underline{X_2});\underline{Y}]}\,EC$$

is height-preserving admissible

Proof.[5]

$$\cfrac{\cfrac{G[\Gamma/(\Sigma/\underline{X_1});(\Sigma/\underline{X_2});\underline{Y}]}{\Gamma/(\Sigma\,\centerdot\,\Sigma/\underline{X_1};\underline{X_2});\underline{Y}]}\;merge}{\Gamma/(\Sigma/\underline{X_1};\underline{X_2});\underline{Y}]}\;c^*$$

■

Lemma 3.14. *Let $G[H]$ be any tree-hypersequent and $G^*[H]$ the result of the application of one of the structural rules — classical and external weakening, merge, tran2. and classical contraction — on $G[H]$. If for a rule $\mathcal{R}$ we have:*

$$\frac{G[H^{'}]}{G[H]}\,\mathcal{R}$$

then it holds that:

$$\frac{G^*[H^{'}]}{G^*[H]}\,\mathcal{R}$$

Proof. By induction on the form of the tree-hypersequent $G[H]$. ■

Lemma 3.15. *Let $G[H]$ be any tree-hypersequent and $G[H^{'}]$ the result of the application of one of the logical rules or the rule $\Box K$ on $G[H]$. If for a rule $\mathcal{R}$ we have:*

$$\frac{G^*[H^{'}]}{G[H^{'}]}\,\mathcal{R}$$

then it holds that:

$$\frac{G^*[H]}{G[H]}\,\mathcal{R}$$

Proof. By induction on the form of the tree-hypersequent $G[H^{'}]$. ■

4. The adequateness of the calculi

In this section we briefly prove that our calculi CSK^* prove exactly the same formulas as their corresponding Hilbert-style systems, that from now on, we will indicate with the notation K^*.

Theorem 4.1. [i] *If $\vdash \alpha$ in K^*, then $\vdash \Rightarrow \alpha$ in CSK^*.*

[ii] *If $\vdash G$ in CSK^*, then $\vdash (G)^\tau$ in K^*.*

---

[5] In the last inference of the proof, if the proof is read bottom up, we use the rule of negation twice in a role. From now on we indicate the repeated running applications of a same rule on a tree-hypersequent, by writing the rule with the symbol * as index.

Proof. By induction on the height of proofs in K^* and CSK^*, respectively. As concerns [ii], we omit the proof which is easy but quite tedious. However the technique to develop such proof consists of the following two steps: first of all, the sequent(s) affected by the rule should be isolated and the corresponding implication proved, then the implication should be transported up all along the tree so that, by modus ponens, the desired result is immediately achieved. In order to further acquaint the reader with the calculi CSK^* we verify [i].The classical axioms and the modus ponens rule are proved as usual, we just present the proof of the distribution axiom, axiom D, axiom 4 and the necessity rule.

$CSK^* \vdash\Rightarrow \Box(\alpha \to \beta) \to \Box\alpha \to \Box\beta$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Box(\alpha\to\beta)\Rightarrow /\alpha\Rightarrow\alpha \qquad \Box\alpha\Rightarrow /\beta\Rightarrow\beta}{\Box(\alpha\to\beta),\Box\alpha\Rightarrow /\alpha\to\beta,\alpha\Rightarrow\beta}{\scriptstyle \to A}}{\Box(\alpha\to\beta),\Box\alpha\Rightarrow /\alpha\Rightarrow\beta}{\scriptstyle \Box A}}{\Box(\alpha\to\beta),\Box\alpha\Rightarrow /\;\Rightarrow\beta}{\scriptstyle \Box A}}{\Box(\alpha\to\beta),\Box\alpha\Rightarrow\Box\beta}{\scriptstyle \Box K}}{\Box(\alpha\to\beta)\Rightarrow\Box\alpha\to\Box\beta}{\scriptstyle \to K}}{\Rightarrow\Box(\alpha\to\beta)\to\Box\alpha\to\Box\beta}{\scriptstyle \to K}$$

$CSKD^* \vdash\Rightarrow \Box\alpha \to \neg\Box\neg\alpha$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Box\alpha,\Box\neg\alpha\Rightarrow /\alpha\Rightarrow\alpha}{\Box\alpha,\Box\neg\alpha\Rightarrow /\neg\alpha,\alpha\Rightarrow}{\scriptstyle \neg A}}{\Box\alpha,\Box\neg\alpha\Rightarrow /\alpha\Rightarrow}{\scriptstyle \Box A}}{\Box\alpha,\Box\neg\alpha\Rightarrow /\;\Rightarrow}{\scriptstyle \Box A}}{\Box\alpha,\Box\neg\alpha\Rightarrow}{\scriptstyle ser.}}{\Box\alpha\Rightarrow\neg\Box\neg\alpha}{\scriptstyle \neg K}}{\Rightarrow\Box\alpha\to\neg\Box\neg\alpha}{\scriptstyle \to K}$$

$CSK4^* \vdash\Rightarrow \Box\alpha \to \Box\Box\alpha$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Box\alpha\Rightarrow /\Box\alpha\Rightarrow /\alpha\Rightarrow\alpha}{\Box\alpha\Rightarrow /\Box\alpha\Rightarrow /\;\Rightarrow\alpha}{\scriptstyle \Box A}}{\Box\alpha\Rightarrow /\Box\alpha\Rightarrow\Box\alpha}{\scriptstyle \Box K}}{\Box\alpha\Rightarrow /\;\Rightarrow\Box\alpha}{\scriptstyle tran.}}{\Box\alpha\Rightarrow\Box\Box\alpha}{\scriptstyle \Box K}}{\Rightarrow\Box\alpha\to\Box\Box\alpha}{\scriptstyle \to K}$$

if $CSK^* \vdash\Rightarrow \alpha$, then $CSK^* \vdash\Rightarrow \Box\alpha$

$$\dfrac{\dfrac{\Rightarrow\alpha}{\Rightarrow /\;\Rightarrow\alpha}{\scriptstyle RN}}{\Rightarrow\Box\alpha}{\scriptstyle \Box K}$$

■

5. Cut-elimination Theorem for CSK*

In this section we prove that the cut-rule is admissible in calculi CSK^*. In order to prove such a theorem we firstly have to show the following lemma.

LEMMA 5.1. *Given three tree-hypersequents together with a displayed occurrence of a sequent* Γ, $I[\Gamma]$, $J[\Gamma]$ *and* $H[\Gamma]$ *such that* $I[\Gamma] \sim J[\Gamma] \sim H[\Gamma]$, *if there is a rule* $\mathcal{R}$ *such that:*

$$\frac{J[\Gamma]}{I[\Gamma]}\ \mathcal{R}$$

then, for every Δ *it holds:*

$$\frac{J \otimes H[\Delta]}{I \otimes H[\Delta]}\ \mathcal{R}$$

PROOF. By induction on the form of the tree-hypersequents $I[\Gamma]$, $J[\Gamma]$ and $H[\Gamma]$. ■

Now we can prove that the cut-rule is admissible in the calculi CSK^*.

THEOREM 5.2. *Let* $G[\Gamma, \alpha]$ *and* $G^{'}[\alpha, \Pi]$ *be such that* $G[\Gamma, \alpha] \sim G^{'}[\alpha, \Pi]$. *If:*

$$\frac{\overset{\vdots\, d_1}{G[\Gamma, \alpha]} \quad \overset{\vdots\, d_2}{G[\alpha, \Pi]}}{G \otimes G^{'}[\Gamma \centerdot \Pi]}\ cut_\alpha$$

and d_1 *and* d_2 *do not contain any other application of the cut rule, then we can construct a proof of* $G \otimes G^{'}[\Gamma \centerdot \Pi]$ *without any application of cut rule.*

PROOF. The proof is developed by induction on the complexity of the cut formula, which is the number (≥ 0) of the occurrences of logical symbols in cut formula α, with subinduction on the sum of the heights of the derivations of the premises of cut. We will distinguish cases by the last rule applied on the left premise.

`Case 1.` $G[\Gamma, \alpha]$ is an initial tree-hypersequent. Then either the conclusion is also a tree-hypersequent or the cut can be replaced by various applications of the classical and external weakening rules on $G^{'}[\alpha, \Pi]$.

`Case 2.` $G[\Gamma, \alpha]$ is inferred by a rule $\mathcal{R}$ in which α is not principal. This case can be standardly solved, by induction on the sum of the heights of the derivations of the premises of cut. Indeed there is no rule which is able to change the position of the sequent where we cut, and, on the other hand, the definition of product assures us that the structure of the tree-hypersequent stay unchanged, therefore no problem arises. However, for the sake of clarity, let us make some examples. More particularly we will analyze those significant cases where the rule $\mathcal{R}$ has been applied on the sequent Γ, α. The others can be dealt with analogously, thanks to the lemma 5.1. Let us then suppose that the rule before $G[\Gamma, \alpha]$ is the rule $\Box K$ (the case where $\mathcal{R}$ is the rule *ser.* is analogous) applied on the sequent Γ, α and without α as principal formula. We have:

$$\dfrac{\dfrac{G[\Gamma, \alpha/ \Rightarrow \beta; \underline{X}]}{G[\Gamma, \alpha, \Box\beta/\underline{X}]}\,\Box K \qquad \begin{matrix}\vdots\\ G'[\alpha, \Pi/\underline{Y}]\end{matrix}}{G \otimes G'[\Gamma \centerdot \Pi, \Box\beta/\underline{X}; \underline{Y}]}\,cut_{\alpha}$$

We reduce to:

$$\dfrac{\dfrac{G[\Gamma, \alpha/ \Rightarrow \beta; \underline{X}] \quad G'[\alpha, \Pi/\underline{Y}]}{G \otimes G'[\Gamma \centerdot \Pi/ \Rightarrow \beta; \underline{X}; \underline{Y}]}\,cut_{\alpha}}{G \otimes G'[\Gamma \centerdot \Pi, \Box\beta/\underline{X}; \underline{Y}]}\,\Box K$$

Let us suppose that the rule before $G[\Gamma, \alpha]$ is the rule $\Box A$ (the case where $\mathcal{R}$ is the rule *tran.* is analogous) applied between the sequent Γ, α and the sequent successive to it, and without α as principal formula. We have:

$$\dfrac{\dfrac{G[\Box\beta, \Gamma, \alpha/(\beta, \Sigma/\underline{X}); \underline{X'}]}{G[\Box\beta, \Gamma, \alpha/(\Sigma/\underline{X}); \underline{X'}]}\,\Box A \qquad \begin{matrix}\vdots\\ G'[\alpha, \Pi/\underline{Y}]\end{matrix}}{G \otimes G'[\Box\beta, \Gamma \centerdot \Pi/(\Sigma/\underline{X}); \underline{X'}; \underline{Y}]}\,cut_{\alpha}$$

We reduce to:

$$\dfrac{\dfrac{G[\Box\beta, \Gamma, \alpha/(\beta, \Sigma/\underline{X}); \underline{X'}] \quad G'[\alpha, \Pi/\underline{Y}]}{G \otimes G'[\Box\beta, \Gamma \centerdot \Pi/(\beta, \Sigma/\underline{X}); \underline{X'}; \underline{Y}]}\,cut_{\alpha}}{G \otimes G'[\Box\beta, \Gamma \centerdot \Pi/(\Sigma/\underline{X}); \underline{X'}; \underline{Y}]}\,\Box A$$

Let us finally suppose that the rule before $G[\Gamma, \alpha]$ is the rule $\Box A$ (the case where $\mathcal{R}$ is the rule *tran.* is analogous) applied between the sequent Γ, α and the sequent which precedes it, and without α as principal formula. We have:

$$\dfrac{\dfrac{G[\Box\beta, \Delta/(\beta, \Gamma, \alpha/\underline{X}); \underline{X'}]}{G[\Box\beta, \Delta/(\Gamma, \alpha/\underline{X}); \underline{X'}]}\,\Box A \qquad \begin{matrix}\vdots\\ G'[\Lambda/(\alpha, \Pi/\underline{Y}); \underline{Y'}]\end{matrix}}{G \otimes G'[\Box\beta, \Delta \centerdot \Lambda/(\Gamma \centerdot \Pi/\underline{X}; \underline{Y}); \underline{X'}; \underline{Y'}]}\,cut_{\alpha}$$

We reduce to:

$$\dfrac{\dfrac{G[\Box\beta, \Delta/(\beta, \Gamma, \alpha/\underline{X}); \underline{X'}] \quad G'[\Lambda/(\alpha, \Pi/\underline{Y}); \underline{Y'}]}{G \otimes G'[\Box\beta, \Delta \centerdot \Lambda/(\beta, \Gamma \centerdot \Pi/\underline{X}; \underline{Y}); \underline{X'}; \underline{Y'}]}\,cut_{\alpha}}{G \otimes G'[\Box\beta, \Delta \centerdot \Lambda/(\Gamma \centerdot \Pi/\underline{X}; \underline{Y}); \underline{X'}; \underline{Y'}]}\,\Box A$$

`Case 3.` $G[\Gamma,\alpha]$ is inferred by a rule $\mathcal{R}$ in which α is principal. We distinguish two subcases: in one subcase $\mathcal{R}$ is a logical rule, in the other $\mathcal{R}$ is a modal rule.

`Case 3.1.` We suppose, as example, that the rule before $G[\Gamma,\alpha]$ is $\neg K$, we have:

$$\dfrac{\dfrac{G[\beta,\Gamma]}{G[\Gamma,\neg\beta]}\,{\scriptstyle \neg K} \qquad \begin{matrix}\vdots\\ G'[\neg\beta,\Pi]\end{matrix}}{G\otimes G'[\Gamma\centerdot\Pi]}\,{\scriptstyle cut_{\neg\beta}}$$

By applying lemma 3.11 on $G^{'}[\neg\beta,\Pi]$, we obtain $G^{'}[\Pi,\beta]$. We replace the previous cut with the following one which is eliminable by induction on the complexity of the cut formula:

$$\dfrac{G^{'}[\Pi,\beta] \quad G[\beta,\Gamma]}{G\otimes G'[\Gamma\centerdot\Pi]}\,{\scriptstyle cut_{\beta}}$$

`Case 3.2.` $\mathcal{R}$ is $\Box K$ and $\alpha\equiv\Box\beta$. We have the following situation:

$$\dfrac{\dfrac{G[\Gamma/\Rightarrow\beta;\underline{X}]}{G[\Gamma,\Box\beta/\underline{X}]}\,{\scriptstyle \Box K} \qquad \begin{matrix}\vdots\\ G'[\Box\beta,\Pi]\end{matrix}}{G\otimes G'[\Gamma\centerdot\Pi/\underline{X}]}\,{\scriptstyle cut_{\Box\beta}}$$

We have to consider the last rule $\mathcal{R}^{'}$ of d_2. If there is no rule $\mathcal{R}^{'}$ which introduces $G^{'}[\Box\beta,\Pi]$ because $G^{'}[\Box\beta,\Pi]$ is an initial tree-hypersequent, then we can solve the case as in 1. If $\mathcal{R}^{'}$ is a rule in which $\Box\beta$ is not principal, we solve the case as in 2. The only problematic cases are those cases where $\mathcal{R}^{'}$ is $\Box A$ or *tran*. We analyze them both.

$\Box A$:

$$\dfrac{\dfrac{G[\Gamma/\Rightarrow\beta;\underline{X}]}{G[\Gamma,\Box\beta/\underline{X}]}\,{\scriptstyle \Box K} \qquad \dfrac{G^{'}[\Box\beta,\Pi/(\beta,\Phi/\underline{Y});\underline{Y^{'}}]}{G'[\Box\beta,\Pi/(\Phi/\underline{Y});\underline{Y^{'}}]}\,{\scriptstyle \Box A}}{G\otimes G'[\Gamma\centerdot\Pi/\underline{X};(\Phi/\underline{Y});\underline{Y^{'}}]}\,{\scriptstyle cut_{\Box\beta}}$$

We reduce to:

$$\dfrac{G[\Gamma,\Box\beta/\underline{X}] \quad G^{'}[\Box\beta,\Pi/(\beta,\Phi/\underline{Y});\underline{Y^{'}}]}{G\otimes G'[\Gamma\centerdot\Pi/\underline{X};(\beta,\Phi/\underline{Y});\underline{Y^{'}}]}\,{\scriptstyle cut_{\Box\beta}}$$

$$\dfrac{\dfrac{G[\Gamma/\Rightarrow\beta;\underline{X}] \quad G\otimes G^{'}[\Gamma\centerdot\Pi/\underline{X};(\beta,\Phi/\underline{Y});\underline{Y^{'}}]}{G\otimes G\otimes G^{'}[\Gamma\centerdot\Gamma\centerdot\Pi/\underline{X};\underline{X};(\Phi/\underline{Y});\underline{Y^{'}}]}\,{\scriptstyle cut_{\beta}}}{G\otimes G'[\Gamma\centerdot\Pi/\underline{X};(\Phi/\underline{Y});\underline{Y^{'}}]}\,{\scriptstyle C^{*}}$$

where the first cut is eliminable by induction on the sum of the heights of the derivations of the premises of cut and the second cut is eliminable by induction on the complexity of cut formula.
tran.:

$$\dfrac{\dfrac{G[\Gamma/\Rightarrow\beta;\underline{X}]}{G[\Gamma,\Box\beta/\underline{X}]}\,\Box K \qquad \dfrac{G^{'}[\Box\beta,\Pi/(\Box\beta,\Phi/\underline{Y});\underline{Y^{'}}]}{G^{'}[\Box\beta,\Pi/(\Phi/\underline{Y});\underline{Y^{'}}]}\,tran.}{G\otimes G^{'}[\Gamma\,{\scriptstyle\blacksquare}\,\Pi/\underline{X};(\Phi/\underline{Y});\underline{Y^{'}}]}\,Cut_{\Box\beta}$$

In order to solve this case, we must check what can have introduced the tree-hypersequent $G^{'}[\Box\beta,\Pi/(\Box\beta,\Phi/\underline{Y});\underline{Y^{'}}]$. More particularly we go up the derivation until either a rule applies to a formula different from the $\Box\beta$'s or a rule different from *tran.* applies to some of the $\Box\beta$'s. This way we have the following situation:

$$\Diamond: \quad G^{'}[\Box\beta,\Pi]\;[[\Box\beta,\Psi]]^{6}$$

We then analyze each of the rules which can have inferred the tree-hypersequent $\Diamond$:

- $\Diamond$ is an axiom. Then, as $\Box\beta$ cannot be principal, even the conclusion of the cut is an axiom and the case is solved.
- $\Diamond$ has been inferred by a rule $\mathcal{R}^{''}$ which does not have any $\Box\beta$ as a principal formula. In this case the technique consists of: firstly, permuting the rule $\mathcal{R}^{''}$ and the n applications of the rule *tran.*, and, secondly, operating as in case 2.
- $\Diamond$ has been inferred by a rule $\mathcal{R}^{''}$ which has $\Box\beta$ as principal formula. $\mathcal{R}^{''}$ can only be the rule $\Box A$. We still have to distinguish two others possibilities. (1) the rule $\Box A$ has been applied to one of the sequents which follow the sequent $[\Box\beta,\Pi]$. Hence we have the following situation:

$$\dfrac{\dfrac{G[\Gamma/\Rightarrow\beta;\underline{X}]}{G[\Gamma,\Box\beta/\underline{X}]} \qquad \dfrac{\dfrac{\dfrac{G^{'}[\Box\beta,\Pi]\;[[\Box\beta,\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]}{G^{'}[\Box\beta,\Pi]\;[[\Box\beta,\Psi/(\Xi/\underline{Z});\underline{Z^{'}}]]}}{\vdots}}{G^{'}[\Box\beta,\Pi]\;[[\Psi/(\Xi/\underline{Z});\underline{Z^{'}}]]}}{G\otimes G^{'}[\Gamma\,{\scriptstyle\blacksquare}\,\Pi/\underline{X}][[\Psi/(\Xi/\underline{Z});\underline{Z^{'}}]]}$$

We proceed within the following three steps:

---

[6] For the sake of brevity, we omit to write: $/(\Phi/\underline{Y^{'}});\underline{Y^{''}}$

(i) we apply the rule $\Box A$ and the n applications of the rule *tran.* in a reverse order and so we obtain the tree-hypersequent:

$$G^{'}[\Box\beta,\Pi]\ [[\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]$$

(ii) we apply the rule *tran*2. to the tree-hypersequent $G[\Gamma/\Rightarrow\beta;\underline{X}]$ a number of time sufficient to get $\Rightarrow\beta$ in an equivalent position with the sequence β,Ξ of the tree-hypersequent $G^{'}[\Box\beta,\Pi]\ [[\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]$. This way we obtain a tree-hypersequent where $\Rightarrow\beta$ is no longer after Γ, but n empty sequences after. Let us note this as: $G[\Gamma/\underline{X}]\ \ [\Rightarrow\beta]$.

(iii) We are now able to apply two cuts: the first eliminable by induction on the sum of the heights, the second by induction on the complexity of the cut formula.

$$\dfrac{G[\Gamma,\Box\beta/\underline{X}]\quad G^{'}[\Box\beta,\Pi]\ [[\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]}{G\otimes G^{'}[\Gamma\centerdot\Pi/\underline{X}][[\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]}\ cut_{\Box\beta}$$

$$\dfrac{\dfrac{G[\Gamma/\underline{X}]\ \ [\Rightarrow\beta]\qquad G\otimes G^{'}[\Gamma\centerdot\Pi/\underline{X}][[\Psi/(\beta,\Xi/\underline{Z});\underline{Z^{'}}]]}{G\otimes G\otimes G^{'}[\Gamma\centerdot\Gamma\centerdot\Pi/\underline{X};\underline{X}][[\Psi/(\Xi/\underline{Z});\underline{Z^{'}}]]}\ cut_{\beta}}{G\otimes G^{'}[\Gamma\centerdot\Pi/\underline{X}][[\Psi/(\Xi/\underline{Z});\underline{Z^{'}}]]}\ C^{*}$$

(2) The rule $\Box A$, with $\Box\beta$ principal formula, has been applied on the sequent $[\Box\beta,\Pi]$. In this case we apply, as before, the rule $\Box A$ and the n applications of the rule *tran.* in a reverse order, and we proceed as at the beginning of this case. ■

6. Conclusions and Further Work

In this paper we have presented the tree-hypersequent method applied to the systems K, KD, $K4$ and $KD4$. Through the several sections we were given a chance to observe the advantages it has: it satisfies the subformula property, its rules are invertible and they fit the criteria required for a good sequent calculus, all the structural rules can be shown to be admissible, the contraction rules included. Moreover all the proof, as the calculi, are purely syntactic. Therefore the tree-hypersequents calculi seem to enjoy the qualities for being defined as good. Given this situation, two interesting questions seem to arise naturally: is it possible to obtain other results within the tree-hypersequent method? Is it possible to apply the tree-hypersequent method in order to obtain calculi for other systems of modal logic? Let us answer both of them, following the order.

As concerns the first hunch, it seems possible to prove two flavored results. As already remarked in [2], the tree shape of the hypersequents should help proving the interpolation theorem. Moreover it does not seem a hard work to adapt the technique introduced by Negri in order to prove the decidability theorem in a purely syntactic way (decidability through semantics has already been established in [2]).

As concerns the second question, in the light of what we have already suggested in [10] and what has been analyzed in details in [2], we can claim that the method of tree-hypersequents can be successfully applied to axioms B and T too (as concerns axiom 5, see again [2]). Moreover, as we have shown in [9], the method can be quite naturally modified in order to get a very simple sequent calculus for modal logic $S5$. Finally it also seems reasonable to apply the method to obtain a sequent calculus for the modal logic of provability GL, in a way similar to that employed by Negri.

On the other hand, we have still to investigate the following two questions: the application of the tree-hypersequents method to temporal logics (which seems quite complicated because of the tree shape of our syntactic objects), and a comparison between the tree-hypersequents method and the tableaux systems one (which seems quite natural).

Acknowledgements. I wish to thank Pierluigi Minari for his constant interest in this work and the substantial help with the notation used in the calculi; Brian Hill for the valuable discussions, comments and suggestions that he has dedicated to this paper.

References

[1] AVRON, A., 'The method of hypersequents in the proof theory of propositional non-classical logic', in Steinhorn, C., Strauss, J., Hodges, W., Hyland, M. (eds.), *Logic: from Foundations to Applications*, Oxford University Press, Oxford, 1996.

[2] BRÜNNLER, T., 'Deep sequent systems for modal logic', *Advances in Modal Logic AiML*, 6: 107–119, 2006.

[3] GOBLE, L., 'Gentzen systems for modal logic', *Notre Dame Journal of Formal Logic*, 15: 455–461, 1974.

[4] INDRZEJCZAK, A., 'Generalised sequent calculus for propositional modal logics', *Logica Trianguli*, 1: 15–31, 1997.

[5] KASHIMA, R., 'Cut-free sequent calculi for some tense logics', *Studia Logica*, 53: 119–135, 1994.

[6] MATSUMOTO, M., OHNISHI, K., 'Gentzen method in modal calculi', *Osaka Mathematical Journal*, 9 and 11: 115–120, 1959.

[7] MINTS, G., 'Indexed systems of sequents and cut-elimination', *Journal of Philosophical Logic*, 26: 671–696, 1997.

[8] Negri, S., ‘Proof analysis in modal logic’, *Journal of Philosophical Logic*, 34, 2005.
[9] Poggiolesi, F., ‘A cut-free simple sequent calculus for modal logic s5’, Submitted to the Review of *Symbolic Logic*, June 2007.
[10] Poggiolesi, F., ‘Sequent calculus for modal logic’, *Logic Colloquium*, 2006.
[11] Sambin, G., Valentini, S., ‘The modal logic of provability. The sequential approach’, *Journal of Philosophical Logic*, 11: 311–342, 1982.
[12] Wansing, H., ‘Sequent calculi for normal modal propositional logics’, *Journal of Logic and Computation*, 4: 125–142, 1994.

Francesca Poggiolesi
Center for Logic and Philosophy of Science (CLWF),
Vrije Universiteit Brussel, 5B457,
Pleinlaan 2, 1050 Brussels, Belgium
`francesca.poggiolesi@unifi.it`

Tomasz Kowalski
Yutaka Miyazaki

All Splitting Logics in the Lattice NExt(KTB)

Abstract. It is proved that there are only two logics that split the lattice **NExt**(**KTB**). The proof is based on the general splitting theorem by Kracht and conducted by a graph theoretic argument.

Keywords: splitting, lattice of modal logics, **KTB**.

1. Introduction

The logic **KTB** is the logic of tolerance relations, that is, a modal logic whose Kripke frames are characterised as reflexive and symmetric. These can also be viewed as the usual undirected graphs, if we assume the edge relation to be reflexive. One could argue that **KTB** is a basic logic of spatial locations (of *being nearby* or suchlike) in the same sense as $\mathbf{K_t}$ is the basic logic of time. The algebraic counterpart of **KTB** is the variety of KTB-algebras. These are Boolean Algebras with Operators (BAOs), of course, but their mathematically most outstanding feature is that the unary operator f satisfies

$$fx \wedge y = 0 \text{ iff } x \wedge fy = 0$$

which makes it into a *self-conjugate* operator. KTB-algebras are in a sense generic among varieties of BAOs with self-conjugate operators, namely, every such BAO has a term-reduct isomorphic to a KTB-algebra. Thus analysing the structure of the variety of KTB-algebras is a reasonable first step in an investigation of varieties of BAOs with self-conjugate operators. We will look closer at the subvariety lattice of the variety of KTB-algebras (speaking in algebraic terms), or (speaking in logical terms) at the lattice **NExt**(**KTB**) of *normal extensions* of **KTB**.

One useful way of analysing structure of lattices in general is via *splittings*. Somewhat informally, a splitting is a pair (a, b) of elements of a lattice L such that a is the largest element not above b. In such a case L splits (quite literally: see Figure 1) into the part above b and the part below a. Thus, splittings, if exist, provide a divide-and-conquer method for

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

dealing with L. Splittings proved to be particularly well-suited to investigating structure of lattices of subvarieties of a given 'big' variety. Examples include McKenzie's work on equational bases for varieties of modular lattices (cf. [18]), and Blok's results on *degree of Kripke incompleteness* for the lattices $\mathbf{NExt}(\mathbf{K}), \mathbf{NExt}(\mathbf{KT})$, and $\mathbf{NExt}(\mathbf{KD})$ (cf. [1, 2]). Blok obtained his results by demonstrating and exploiting the existence of many splittings in the variety of modal algebras. Later, Blok and Pigozzi in their papers on varieties with *equationally definable principal congruences (EDPC)* (cf. [4, 3, 5, 6]) proved that all varieties with EDPC have many splittings. Translated to modal logic, this result implies that $\mathbf{NExt}(\mathbf{L})$ has many splittings if $\mathbf{L}$ possesses a *master modality* (this was known for modal algebras before EDPC papers, cf. [21]). If $\mathbf{L}$ is moreover finitely axiomatised, then every finite subdirectly irreducible $\mathbf{L}$-algebra defines a splitting of $\mathbf{NExt}(\mathbf{L})$ (see Proposition 3.3 below). An important modal logic with that property is $\mathbf{K4}$ and it is worth reminding that Zakharyaschev's *canonical formulae*, a powerful tool for dealing with transitive frames, are connected to splittings via *characteristic formulae* of Jankov and Fine, of which they are a generalisation.

Without master modality (or EDPC in the general case) the situation is less clear. While $\mathbf{NExt}(\mathbf{K})$ has many splittings (cf. [1]), the lattice $\mathbf{NExt}(\mathbf{K_t})$ of normal *tense* logics has only one (cf. [14]). Outside modal domain, by the classical Jankov's result (cf. [11]) all finite subdirectly irreducible Heyting algebras are splitting, yet generalising Heyting algebras to FL_{ew}*-algebras* reduces the number of splittings to one (cf. [13]). On the other hand, the variety of all lattices has many splittings: so many that splitting lattices generate it (cf. [9]).

In this paper we consider splittings in the lattice $\mathbf{NExt}(\mathbf{KTB})$. From Makinson's theorem (cf. [17]) it follows that the logic of one reflexive point, that is, full relation on one point ($\mathbf{L}(\circ)$) splits this lattice. The second author has shown in [19] that the logic of full relation on two points ($\mathbf{L}(\circ\!\!-\!\!\circ)$) also splits $\mathbf{NExt}(\mathbf{KTB})$. We will complete the picture by showing that there are no logics other than two mentioned above that split $\mathbf{NExt}(\mathbf{KTB})$. Our technique makes use of General Splitting Theorem due to Kracht [16].

2. Preliminaries

Our notation in general follows Chagrov and Zakharyaschev [8], which we also ask the reader to consult for any modal logic notions left undefined. We deal with propositional modal logics, so our language $\mathcal{L}$ consists of (1) a denumerable set of propositional variables $\{p_0, p_1, p_2, \ldots\}$, (2) boolean and modal connectives $\wedge, \vee, \neg, \rightarrow, \Box, \Diamond$, (3) a pair of parentheses $(,)$. The set Φ of all formulas in this language is defined as usual. A *normal modal logic* (we call it simply a *logic* for short) over $\mathcal{L}$ is a set of formulas that contains all classical tautologies, the formula $\Box(p \wedge q) \rightarrow (\Box p \wedge \Box q)$, and is closed under *uniform substitution*, *modus ponens*, and *necessitation*. The smallest normal modal logics is denoted by $\mathbf{K}$. The largest one is the whole Φ, the unique *inconsistent logic* over $\mathcal{L}$.

For a logic $\mathbf{L}$ and a set Γ of formulas, the smallest normal modal logic containing both $\mathbf{L}$ and Γ is denoted by $\mathbf{L} \oplus \Gamma$, and the class of all normal extensions of $\mathbf{L}$ is denoted by $\mathbf{NExt}(\mathbf{L})$. The class $\mathbf{NExt}(\mathbf{L})$ is a complete lattice for any $\mathbf{L}$; we use the same symbol for the class and the lattice.

We put $\mathbf{T} := \Box p \rightarrow p$, $\mathbf{B} := p \rightarrow \Box\Diamond p$, and the logic $\mathbf{K} \oplus \{\mathbf{T}, \mathbf{B}\}$ is denoted by $\mathbf{KTB}$. Sometimes a normal modal logic containing $\mathbf{KTB}$ is called a *KTB logic*. Thus, we deal here with the lattice of KTB logics.

As usual, we use frames as semantic objects. A (*general*) *frame* is a triple $\mathcal{F} := \langle W, R, P \rangle$, where W is a non-empty set, R is a binary relation on W, and P is a subset of $\mathcal{P}(W)$ that satisfies (1) $W \in P$, (2) W is closed under $\cap$ (intersection), $-$ (set-theoretical complement), and the operation f_R defined for any $X \in P$ by $f_R(X) := R^{-1}(X)$. If W is finite, we can always assume $P = \wp(W)$ and forget about P altogether, thereby obtaining a *Kripke frame*. As we only need finite frames in the paper, we will use *frame* to mean *Kripke frame* from now on.

A *model* on a frame $\mathcal{F}$ is $\mathfrak{M} := \langle \mathcal{F}, V \rangle$, where V is a function from the set of variables into P, called a *valuation*. The notion of *truth* of a formula φ at a point x in a model $\mathfrak{M}$ ($\mathfrak{M} \models_x \varphi$) is defined as usual. We will also use the usual shorthand $x \models \varphi$ if the model is clear from context or does not have to be specified precisely. For a frame $\mathcal{F}$, a formula φ is *valid* in $\mathcal{F}$ ($\mathcal{F} \models \varphi$) if for any model $\mathfrak{M}$ on $\mathcal{F}$ and for any point $x \in W$, we have $\mathfrak{M} \models_x \varphi$. We write $\mathcal{F} \models \mathbf{L}$ to mean that $\mathcal{F} \models \varphi$ for all $\varphi \in \mathbf{L}$. The logic determined by $\mathcal{F}$ is denoted by $\mathbf{L}(\mathcal{F})$, specifically, $\mathbf{L}(\mathcal{F}) := \{\varphi \in \Phi \colon \mathcal{F} \models \varphi\}$.

It is well-known that the axioms $\mathbf{B}$ and $\mathbf{T}$ correspond, respectively, to symmetry and reflexivity of frames in the following sense. For a frame $\mathcal{F}$, (1) $\mathcal{F} \models \mathbf{B}$ if and only if $\mathcal{F} \Vvdash \forall x, y(xRy \text{ implies } yRx)$. (2) $\mathcal{F} \models \mathbf{T}$ if and only if $\mathcal{F} \Vvdash \forall x(xRx)$. We use $\Vvdash$ to express the fact that the frame satisfies the condition on the right hand side in the standard model-theoretical sense. Thus, a *KTB-frame* is a reflexive and symmetric frame.

We will also make use of *modal algebras*. An algebraic structure $\mathfrak{A} = \langle \mathsf{A}, \cap, \cup, -, f, 0, 1 \rangle$ is a modal algebra if: (1) $\langle A, \cap, \cup, -, 0, 1 \rangle$ is a Boolean algebra, and (2) f is a unary *normal operator*, that is, the following identities hold:

(i) $f0 = 0$ and
(ii) $f(a \vee b) = fa \vee fb$ for $a, b \in A$.

Formulas are interpreted in a modal algebra by a valuation in the standard way and a formula φ is defined to be *valid* in a modal algebra $\mathfrak{A}$ if the identity $\varphi = 1$ holds in $\mathfrak{A}$, which we will abbreviate by $\mathfrak{A} \models \varphi$ (abusing the model-theoretical notation a little). For a logic $\mathbf{L}$, we denote $\mathfrak{A} \models \mathbf{L}$ to mean that $\mathfrak{A} \models \varphi$ for any $\varphi \in \mathbf{L}$. The logic determined by a frame $\mathfrak{A}$ is denoted by $\mathbf{L}(\mathfrak{A})$, specifically, $\mathbf{L}(\mathfrak{A}) := \{\varphi \in \Phi \colon \mathfrak{A} \models \varphi\}$. A modal algebra $\mathfrak{A}$ is a *KTB-algebra* if it satisfies two further identities:

(iii) $a \leq fa$ and
(iv) $a \leq -f-fa$.

The second of these is in fact equivalent to the statement

$$fa \wedge b = 0 \text{ iff } a \wedge fb = 0$$

mentioned already in the introduction. Thus, in BAO terminology, a KTB algebra is a BAO with a single unary self-conjugate normal operator.

Since we will work with finite algebras and frames, it suffices to recall the duality between finite *atom structures* and finite modal algebras. Let $\mathfrak{A}$ be a finite modal algebra. Define a frame $\mathfrak{A}_* = \langle W_\mathfrak{A}, R_\mathfrak{A}\rangle$ by putting $W_\mathfrak{A}$ to be the set of all atoms of A and $R_\mathfrak{A}$ a binary relation on $W_\mathfrak{A}$ defined for any $a, b \in W_\mathfrak{A}$ by $(a, b) \in R_\mathfrak{A}$ iff $fb \geq a$. The frame $\mathfrak{A}_*$ is known as the atom structure of $\mathfrak{A}$ and both $\mathfrak{A}_*$ and $\mathfrak{A}$ validate exactly the same formulas. Conversely, for any finite frame $\mathcal{F} = \langle W, R\rangle$, the corresponding modal algebra $\mathcal{F}^*$ is defined by putting $\mathcal{F}^* = \langle P, \cap, \cup, -, f_R, \emptyset, W\rangle$. Then, both $\mathcal{F}$ and $\mathcal{F}^*$ validate the same set of formulas. Moreover, for the transformations $(\cdot)^*$ and $(\cdot)_*$ we have that $\mathfrak{A} \cong (\mathfrak{A}_*)^*$ for any finite modal algebra $\mathfrak{A}$, and $\mathcal{F} \cong (\mathcal{F}^*)_*$ for any finite frame $\mathcal{F}$. To be sure, these niceties break down for infinite structures and the full force of Stone-Jónsson-Tarski duality is needed.

3. Splitting

3.1. Splittings of lattices of modal logics in general

Splittings of lattices were first investigated by Whitman in [23].

Definition 3.1 (Splitting). Let $\mathfrak{L} = \langle L, \wedge, \vee\rangle$ be a lattice and $a \in L$. Then *a splits* $\mathfrak{L}$ if there exists $b \in L$ such that for any $x \in L$, either $x \leq a$ or $b \leq x$, but not both. Such a pair (a, b) is called a *splitting pair* of the lattice $\mathfrak{L}$. In this case the element b is a *splitting* of $\mathfrak{L}$ by a and denoted by $\mathfrak{L}/a$. The elements of a splitting pair (a, b) are sometimes called *splitting partners*.

See Figure 1 for a picture of splitting (but notice that the definition does not preclude the situation with $a < b$; if this happens then b is a unique cover of a and b a unique subcover of a, such a splitting may be called trivial). When splittings of a lattice of logics are considered and a logic $\mathbf{L}(\mathcal{F})$ splits the lattice, then it is often said that the frame $\mathcal{F}$ splits the lattice. Since the terminology of splittings is quite messy (perhaps *splitting* and *co-splitting* choice adopted in [15] is the least so, but it is not the most common), we offer the following rule-of-thumb principle. For lattices of varieties, think of splitting as determined 'low down' (point b in Figure 1) by a variety generated by a single (finite) subdirectly irreducible algebra $\mathfrak{B}$; its splitting partner being the largest variety not containing $\mathfrak{B}$. For lattices of logics, think of splitting as determined 'high up' (point a in Figure 1) by a logic of a single (finite) rooted frame $\mathcal{A}$; its splitting partner being the smallest logic not verified by $\mathcal{A}$. The algebra $\mathfrak{B}$ is a *splitting algebra*, the frame $\mathcal{A}$ a *splitting frame* and the logic of the splitting partner of $\mathcal{A}$ is a *splitting logic*[1] Notice that by duality between lattices of logics and lattices of varieties and between finite algebras and their atom structures, $\mathfrak{B}$ and $\mathfrak{B}_*$ ($\mathcal{A}^*$ and $\mathcal{A}$) determine dually isomorphic splittings.

In order to formulate General Splitting Theorem that we will make use of, we need some more notation. For $n \in \omega$ we use $\Box^n$ for the n-th iteration of the box operator, and put $\Box^{(n)}\varphi := \bigwedge_{i=0}^{n} \Box^i\varphi$. For a finite frame $\mathcal{F}$, we define the *diagram* $\Delta_{\mathcal{F}}$ of $\mathcal{F}$ as the following set of formulas: first we fix a

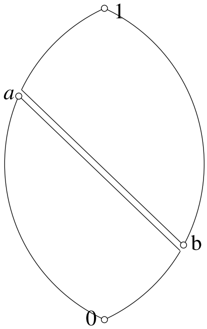


Figure 1. A splitting of a (bounded) lattice $\mathfrak{L}$

---

[1] On the purely linguistic level, think of anything that takes the adjective *splitting* as small. A single rooted frame is small, as opposed to a class of frames. The same goes for a single subdirectly irreducible algebra. But their logics are large and so a splitting logic, i.e., a splitting partner of a large logic, is small. Surely, it is not fully satisfactory, for the inconsistent logic turns out to be splitting in some classes, too. But nobody's perfect.

distinct propositional variable p_a for each element $a \in W$. Then,

$$\Delta_{\mathcal{F}} := \{p_a \to \Diamond p_b : aRb\} \cup \{p_a \to \neg\Diamond p_b : \neg(aRb)\} \cup \{p_a \to \neg p_b : a \neq b\} \cup \{\bigvee_{x \in W} p_x\}$$

The *characteristic formula* $\delta_{\mathcal{F}}$ for $\mathcal{F}$ is then defined by $\delta_{\mathcal{F}} := \bigwedge \Delta_{\mathcal{F}}$. In a frame $\mathcal{F} := \langle W, R\rangle$, a point $r \in W$ is called a *root* if for any $x \in W$, there exists a number n such that $rR^n x$. A frame with a root is called a *rooted* frame. The following theorem was shown by Kracht in [16] (see also Wolter [24]).

THEOREM 3.2 (General Splitting Theorem). *Let* $\Gamma_0 \in \mathbf{NExt}(\mathbf{K})$ *and* $\mathcal{F}$ *be a finite Kripke frame with root* r. *Then the following conditions are equivalent:*

1. $\mathcal{F}$ *splits* $\mathbf{NExt}(\Gamma_0)$.
2. *There is* $n \in \omega$ *such that for any frame* $\mathcal{G}$ *with* $\mathcal{G} \models \Gamma_0$, *if* $\Box^{(n)}\delta_{\mathcal{F}} \wedge p_r$ *is satisfiable in* $\mathcal{G}$, *then* $\Box^{(m)}\delta_{\mathcal{F}} \wedge p_r$ *is also satisfiable in* $\mathcal{G}$ *for any* $m > n$.

Thus, we have $\mathbf{NExt}(\Gamma_0)/\mathcal{F} = \Gamma_0 \oplus (\Box^{(n)}\delta_{\mathcal{F}} \to \neg p_r)$, where $\mathbf{NExt}(\Gamma_0)/\mathcal{F}$ is the splitting logic. In particular, if Γ_0 is finite, then the splitting logic is finitely axiomatisable. The next result holds in fact for any variety of algebras generated by their finite members (cf. [18]).

PROPOSITION 3.3. *Let* $\mathbf{L}_0$ *be a modal logic which has the finite model property, and* $\mathfrak{A}$ *an algebra for* $\mathbf{L}_0$. *If* $\mathbf{L}(\mathfrak{A})$ *splits* $\mathbf{NExt}(\mathfrak{A})$, *then there exists a finite subdirectly irreducible algebra* $\mathfrak{B}$ *such that* $\mathbf{L}(\mathfrak{A}) = \mathbf{L}(\mathfrak{B})$.

Since **KTB** has the finite model property, by the proposition above the only candidates for splitting frames in **NExt**(**KTB**) are the finite ones.

For contrast with many splittings in **NExt**(**K**), we also recall the following result.

THEOREM 3.4 (Blok [1]). *There exists only one splitting logic in* **NExt**(**KT**), *that is the inconsistent logic* Φ.

Notice that this is an example of a trivial splitting: the frame that splits **NExt**(**KT**) is the one-element reflexive frame and $\mathbf{L}(\circ)$ is the unique subcover of Φ.

3.2. Splittings of the lattice NExt(KTB)

On the positive side, we know two logics that split the lattice **NExt**(**KTB**). First, let us recall a classical theorem.

Theorem 3.5 (Makinson [17]). *Every consistent normal modal logic is contained in either* $\mathbf{L}(\bullet)$ *or* $\mathbf{L}(\circ)$*, where* $\bullet$ *is the frame of one irreflexive point, and* $\circ$ *is the frame of one reflexive point.*

Since $\mathbf{L}(\bullet)$ does not belong to $\mathbf{NExt}(\mathbf{KTB})$, we have that the logic $\mathbf{L}(\circ)$ splits the lattice $\mathbf{NExt}(\mathbf{KTB})$ (again, trivially). The second author discovered another logic which splits $\mathbf{NExt}(\mathbf{KTB})$.

Theorem 3.6 (Miyazaki [19]). $\mathbf{L}(\circ\!\!-\!\!\circ)$ *splits the lattice* $\mathbf{NExt}(\mathbf{KTB})$*, where* $\circ\!\!-\!\!\circ$ *is the full relation on two points.*

The proof of this theorem is based on Theorem 2.2, and the splitting partner of $\mathbf{L}(\circ\!\!-\!\!\circ)$ turns out to be $\mathbf{L}(\circ)$. This means that $\mathbf{L}(\circ\!\!-\!\!\circ)$ is the third greatest of all members in $\mathbf{NExt}(\mathbf{KTB})$. We will prove that no logic other than the two above splits $\mathbf{NExt}(\mathbf{KTB})$.

4. Connected KTB-frames

4.1. Subdirectly irreducible modal algebras and connected KTB-frames

We begin by recalling some notions from universal algebra. An algebra $\mathfrak{A}$ is a *subdirect product* of an indexed family $\{\mathfrak{A}_i \colon i \in I\}$ of the same type if there exists a one-to-one homomorphism $f : \mathfrak{A} \to \prod_{i\in I} \mathfrak{A}_i$ such that for any $i \in I$, $\pi_i \circ f : \mathfrak{A} \to \mathfrak{A}_i$ is onto, where π_i is a projection map to i-th coordinate. The map f is called a *subdirect representation* of $\mathfrak{A}$. A non-trivial algebra $\mathfrak{A}$ is *subdirectly irreducible* if for any subdirect representation $f : \mathfrak{A} \to \prod_{i\in I} \mathfrak{A}_i$ of $\mathfrak{A}$, there exists $j \in I$ such that $\pi_j \circ f : \mathfrak{A} \to \mathfrak{A}_j$ is one-to-one. By Birkhoff Theorem every algebra in a variety $\mathcal{V}$ is isomorphic to a subdirect product of subdirectly irreducible members in $\mathcal{V}$ (see e.g. [7]). Phrased in terms of logics this means that every modal logic is characterised by a class of subdirectly irreducible modal algebras.

Frame-theoretical counterparts of subdirectly irreducible modal algebras for KTB logics are *connected* frames.

Definition 4.1 (Connectedness). Let $\mathcal{F} = \langle W, R\rangle$ be a KTB-frame. Then, $\mathcal{F}$ is *connected* if for any $x, y \in W$, there is a number $n \in \omega$ such that $xR^n y$.

Connectedness for KTB-frames is thus exactly the same as connectedness for undirected graphs. More graph-theoretical notions will be recalled in the next section, connectedness appears here because finite connected frames are precisely the atom structures of finite subdirectly irreducible (in fact, simple) KTB-algebras.

PROPOSITION 4.2. *Let $\mathcal{F} = \langle W, R\rangle$ be a finite KTB-frame and $\mathfrak{A}$ be a finite KTB-algebra. Then, $\mathfrak{A}$ is subdirectly irreducible iff $\mathfrak{A}$ is simple. Moreover,*

1. *$\mathcal{F}^*$ is simple iff $\mathcal{F}$ is connected.*
2. *$\mathfrak{A}$ is simple iff $\mathfrak{A}_*$ is connected.*

The relation between subdirectly irreducible KTB algebras and connected KTB-frames is deeper and a version of it holds also in the infinite general case (cf. [19]), but this is of no concern to us here. All we need, is that Proposition 4.2 together with Proposition 3.3 imply that in search for frames that split **NExt**(**KTB**) we can restrict attention to the class of finite connected KTB-frames.

4.2. Graph-theoretical notions

This section is just cosmetics. We define some usual (and two not-so-usual) graph-theoretical notions for KTB-frames, simply by taking reflexivity into account. Let $\mathcal{F} = \langle W, R\rangle$ be a connected KTB-frame. Define a binary relation $\check{R}$ on W putting $x\check{R}y$ if and only if xRy and $x \neq y$.

When there is a finite sequence of points $\{a_i\}_{i=0}^{n} \subseteq W$ in $\mathcal{F}$ $(n \geq 1)$ such that $a_i\check{R}a_{i+1}$ for all $i \in [0, n-1]$, the list $t = [a_0, a_1, \ldots, a_n]$ is called a *path* in $\mathcal{F}$. In this case, the *length* $\ell(t)$ of the path t is n. In $\mathcal{F}$ we define a *distance* function $d : W \times W \to \omega$ by putting $d(x, y) := 0$ if $x = y$, and $d(x, y) := n + 1$ if not$(xR^n y)$ and $xR^{n+1}y$. The *diameter* $s(\mathcal{F})$ of the frame $\mathcal{F}$ is defined by $s(\mathcal{F}) := \max\{n \in \omega \colon \exists x, y \in W,\ d(x, y) = n\}$.

In a KTB-frame $\mathcal{F} = \langle W, R, P\rangle$, for a point $x \in W$, the *degree* of x is a number defined as $deg(x) := card\{y \in W \colon x\check{R}y\}$. In a connected frame of course $deg(x) \geq 1$ for all $x \in W$. We call an $a \in W$ a *tail* of $\mathcal{F}$ if there exists exactly one point $b \in W$ such that $b\check{R}a$. In this case, b is called the *base* of the tail a. In other words, a is a tail if and only if $deg(a) = 1$.

4.3. Pasting of KTB-frames

Our pasting construction is based on ideas from the first author's [12], developed independently, but resembling the *garland* construction of Kracht [15].

Let $\mathcal{F} = \langle W, R\rangle$ and $\mathcal{G} = \langle U, S\rangle$ be KTB Kripke frames, $a \in W$, and $b \in U$. The *pasting* $\mathcal{F} \biguplus_{(a,b)} \mathcal{G} = \langle X, T\rangle$ of $\mathcal{F}$ and $\mathcal{G}$ at a and b is a KTB Kripke frame defined as follows. First, we define an equivalence relation $\equiv_{(a,b)}$ on $W \cup U$, putting $\equiv_{(a,b)} := \{(x, x) \colon x \in W\} \cup \{(y, y) \colon y \in U\} \cup \{(a, b), (b, a)\}$. Then we define the set of worlds $X := \{[x] \colon x \in W \cup U\}$, where $[x] := \{y \in W \cup U \colon x \equiv_{(a,b)} y\}$, and the relation T is defined as: for $[x], [y] \in X$, ${}_{[x]}T_{[y]}$ if and only if for some $u \in [x]$ and for some $v \in [y]$, uRv or uSv holds. In other words, the pasting of $\mathcal{F}$ and $\mathcal{G}$ is a Kripke frame obtained by identifying a in $\mathcal{F}$ and b in $\mathcal{G}$ and leaving everything else untouched. Thus, since for all points $x \in W \cup U$ except a and b we have $[x] = \{x\}$, we will often omit square brackets in order to avoid an overloaded notation. The pasting construction will be used to build a sequence of KTB Kripke frames showing that a given frame cannot split the lattice $\mathbf{NExt}(\mathbf{KTB})$.

5. Few splittings theorem

Having prepared the tools, in this section we show that no KTB logics except $\mathbf{L}(\circ)$ and $\mathbf{L}(\circ\!\!-\!\!\circ)$ split the lattice $\mathbf{NExt}(\mathbf{KTB})$. Our proof strategy makes use of Theorem 2.2. This theorem implies that a finite connected KTB Kripke frame $\mathcal{F}$ does not split $\mathbf{NExt}(\mathbf{KTB})$ if there is a sequence $\{\mathcal{G}_n\}_{n\in\omega}$ of KTB-frames such that for every $n \in \omega$ the following two conditions are satisfied:

(a) $\Box^n \delta_{\mathcal{F}} \wedge p_0$ is satisfiable in $\mathcal{G}_n$,
(b) there is an $m > n$ such that $\Box^m \delta_{\mathcal{F}} \wedge p_0$ is not satisfiable in $\mathcal{G}_n$.

Let $\mathcal{F}$ be a given finite and connected KTB-frame, whose diameter is ℓ, and for a_0 and a_N in this frame, let $d(a_0, a_N) = s(\mathcal{F}) = \ell$. In order to construct a sequence $\{\mathcal{G}_n\}_{n\in\omega}$ with properties (a) and (b) for $\mathcal{F}$, we will build the following frame $\mathcal{G}_n$ for a given n. Suppose n satisfies $k\ell \geq n < (k+1)\ell$. First, we prepare $k+1$ copies $\mathcal{F}^{(0)}, \mathcal{F}^{(1)}, \mathcal{F}^{(2)}, \ldots, \mathcal{F}^{(k+1)}$ and paste $\mathcal{F}^{(0)}$ and $\mathcal{F}^{(1)}$ at $a_0^{(0)}$ and $a_0^{(1)}$, paste $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ at $a_N^{(1)}$ and $a_N^{(2)}$, etc., so that the copies are pasted alternately at a_0s and a_Ns. Next, we take a suitable finite connected KTB-frame $\mathcal{S}$, which we call a singularity, and paste it at the end of the construction obtained in the first step.

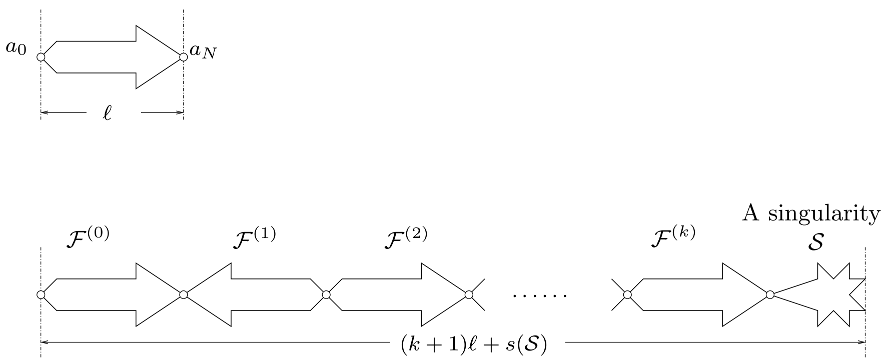


Figure 2. $\mathcal{G}_n$ for $k\ell \leq n < (k+1)\ell$

It is easy to see that the first ‘nonsingular’ part of the frame $\mathcal{G}_n$ above satisfies $\Box^n\delta_{\mathcal{F}} \wedge p_0$ at $a_0^{(0)}$ under a valuation $V(p_j) := \{a_j^{(i)} : 0 \leq i \leq k\}$ for all $j \in [0, N]$. Therefore, so does the entire frame, as the singularity (whatever it is) is too far away to interfere. But now have to choose the ‘singular’ part so that $\Box^m\delta_{\mathcal{F}} \wedge p_0$ were not satisfiable in $\mathcal{G}_n$ for some $m > n$. We also have to do it in some systematic way, because for $\mathcal{F}$ we can take any finite connected KTB-frame (the only two exceptions being the splitting frames).

To achieve that we divide all finite and connected KTB-frames into two groups: chains, and frames that are not chains. For frames what are not chains, we take a chain as its singularity, whereas for chains, singularities of some particular form are needed. A (*finite*) *chain* is a KTB-frame $\mathcal{C}_n := \langle W, R\rangle$, where $W := \{c_0, c_1, c_2, \cdots, c_{n-1}\}$, and $R := \{(c_i, c_j) \in W \times W : |i - j| \leq 1\}$ for $n \geq 1$.

Theorem 5.1. *For $\ell \geq 3$, $\mathcal{C}_\ell$ does not split* **NExt**(**KTB**).

Proof. For each c_i $(0 \leq i \leq \ell - 1)$ in $\mathcal{C}_\ell$, we prepare a variable p_i, and we construct $\Delta_{\mathcal{C}_\ell}$ and $\delta_{\mathcal{C}_\ell}$ as usual. For $\ell \geq 3$, define a frame $\mathcal{S}_\ell := \langle U, Q\rangle$ as follows. $U := \{t, s_0, s_1, s_2, \ldots, s_\ell\}$, and $Q := \{(s_i, s_j) \in (U - \{t\}) \times (U - \{t\}) : |i - j| \leq 1\} \cup \{(t, t), (t, s_1), (s_1, t)\}$. Of course, $s(\mathcal{C}_\ell) = \ell - 1$. Now we construct a sequence $\{\mathcal{G}_n\}_{n\in\omega}$. For any $n \in [k(\ell-1), (k+1)(\ell-1))$ for some $k \geq 0$, we prepare $(k+1)$ copies of $\mathcal{C}_\ell$ and define $\mathcal{G}_n$ as follows. For an even k, we put

$$\mathcal{G}_n := \mathcal{C}_\ell^{(0)} \biguplus_{(c_{\ell-1}^{(0)}, c_{\ell-1}^{(1)})} \mathcal{C}_\ell^{(1)} \biguplus_{(c_0^{(1)}, c_0^{(2)})} \mathcal{C}_\ell^{(2)} \biguplus_{(c_{\ell-1}^{(2)}, c_{\ell-1}^{(3)})} \cdots \biguplus_{(c_0^{(k-1)}, c_0^{(k)})} \mathcal{C}^{(k)} \biguplus_{(c_{\ell-1}^{(k)}, s_0)} \mathcal{S}_\ell.$$

Alternatively, for an odd k, we put

$$\mathcal{G}_n := \mathcal{C}_\ell^{(0)} \biguplus_{(c_{\ell-1}^{(0)}, c_{\ell-1}^{(1)})} \mathcal{C}_\ell^{(1)} \biguplus_{(c_0^{(1)}, c_0^{(2)})} \mathcal{C}_\ell^{(2)} \biguplus_{(c_{\ell-1}^{(2)}, c_{\ell-1}^{(3)})} \cdots \biguplus_{(c_{\ell-1}^{(k-1)}, c_{\ell-1}^{(k)})} \mathcal{C}^{(k)} \biguplus_{(c_0^{(k)}, s_0)} \mathcal{S}_\ell.$$

An example of $\mathcal{G}_n$ is shown below.

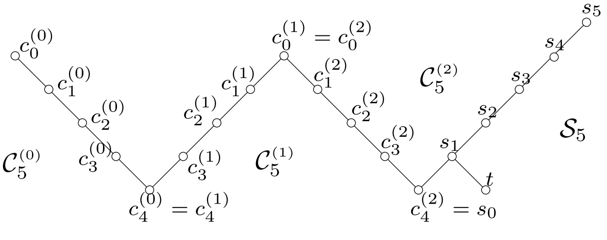


Figure 3. The shape of $\mathcal{G}_{11}$, where $k = 2$ and $\ell = 5$.

Then, note that $s(\mathcal{G}_n) = (k+1) \times s(\mathcal{C}_\ell) + s(\mathcal{S}) = (k+1)(\ell-1)+\ell$. The frame $\mathcal{G}_n$ satisfies $(\Box^n \delta_{\mathcal{C}_\ell}) \wedge p_0$ at the point $c_0^{(0)}$ under the valuation V given by $V(p_i) := \{c_i^{(j)} : 0 \leq j \leq k\}$ for $0 \leq i \leq \ell-1$, since $k(\ell-1) \leq n < (k+1)(\ell-1)$. It remains to show that $\Box^m \delta_{\mathcal{F}} \wedge p_0$ is not satisfiable in $\mathcal{G}_n$ for some $m > n$.

Put $m := (k+1)(\ell-1)+\ell$, and suppose that $\mathcal{G}_n$ satisfies $(\Box^m \delta_{\mathcal{C}_\ell}) \wedge p_0$ at some point a under some valuation V. Because $m = s(\mathcal{G}_n)$, wherever a is in $\mathcal{G}_n$, every point in $\mathcal{G}_n$ must make $\delta_{\mathcal{C}_\ell}$ true under V. In particular, for every point in $\mathcal{G}_n$ there is one and only one variable that is true at that point under V. Therefore, exactly one variable is true at t under V. Because t is a tail in $\mathcal{G}_n$, then $t \models p_0$ or else $t \models p_{\ell-1}$ must hold. Consider the case $t \models p_0$. Then because $c_0 R c_0$ and $c_0 R c_1$, we can have $s_1 \models p_0$ or $s_1 \models p_1$, but the former case never happens since in that case, $t \not\models p_0 \to \Diamond p_1$. Thus $s_1 \models p_1$. By the same reasoning, we can deduce that $s_i \models p_i$ for every $1 \leq i \leq \ell-1$. Now consider s_ℓ. Because of the form of $\delta_{\mathcal{C}_\ell}$, there must be precisely one variable that is satisfied at s_ℓ, moreover, since s_ℓ is a tail, that variable can only be p_0 or $p_{\ell-1}$. Since $p_{\ell-1}$ is true at a neighbouring $s_{\ell-1}$, the only choice is $s_\ell \models p_{\ell-1}$. But as $s_\ell \not\models \Diamond p_{\ell-2}$ that choice is also impossible. So, the assumption $t \models p_0$ leads to a contradiction. Similar reasoning shows that the assumption $t \models p_{\ell-1}$ leads to a contradiction, too. Therefore we conclude that $\mathcal{G}_n$ does not satisfy the formula $(\Box^m \delta_{\mathcal{C}_\ell}) \wedge p_0$, and thus $\mathcal{C}_\ell$ cannot split the lattice **NExt**(**KTB**). ■

The remaining candidates for splitting frames are those that are not chains. We will call a finite KTB-frame $\mathcal{F}$ a *non-chain*, if $\mathcal{F}$ is not isomorphic to $\mathcal{C}_n$ for any $n \geq 1$.

THEOREM 5.2. *Let $\mathcal{F}$ be a finite and connected KTB-frame that is a non-chain. Then $\mathcal{F}$ does not split the lattice* **NExt**(**KTB**).

PROOF. Suppose that $\mathcal{F} := \langle W, R \rangle$ and that $s(\mathcal{F}) = \ell$. Since $\mathcal{F}$ is a non-chain, we have that $\ell \neq 0$ and $|W| \geq 3$. The definition of diameter ensures that there exist $x, y \in W$ such that $d(x,y) = \ell$. Since W is finite, we may specify all points in W as $W := \{a_0, a_1, a_2, \ldots, a_N\}$. Also we may assume that $d(a_0, a_N) = \ell$. Taking a variable p_i for each a_i, we build $\Delta_{\mathcal{F}}$ and $\delta_{\mathcal{F}}$. Then we construct a sequence $\{\mathcal{G}_n\}_{n \in \omega}$ of KTB-frames as before, the only difference being that we take a chain $\mathcal{C}_{\ell+1}$ as the singularity. To be precise, let us fix an $n \in [k\ell, (k+1)\ell)$ $(k \in \omega)$. We first prepare $(k+1)$ copies of $\mathcal{F}$, that is, $\mathcal{F}^{(0)}, \mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(k)}$. Then we define the frame $\mathcal{G}_n$ putting, for an even k,

$$\mathcal{G}_n := \mathcal{F}^{(0)} \biguplus_{(a_N^{(0)},a_N^{(1)})} \mathcal{F}^{(1)} \biguplus_{(a_0^{(1)},a_0^{(2)})} \mathcal{F}^{(2)} \biguplus_{(a_N^{(2)},a_N^{(3)})} \cdots \biguplus_{(a_0^{(k-1)},a_0^{(k)})} \mathcal{F}^{(k)} \biguplus_{(a_N^{(k)},c_0)} \mathcal{C}_{\ell+1}.$$

Alternatively, for an odd k,

$$\mathcal{G}_n := \mathcal{F}^{(0)} \biguplus_{(a_N^{(0)},a_N^{(1)})} \mathcal{F}^{(1)} \biguplus_{(a_0^{(1)},a_0^{(2)})} \mathcal{F}^{(2)} \biguplus_{(a_N^{(2)},a_N^{(3)})} \cdots \biguplus_{(a_N^{(k-1)},a_N^{(k)})} \mathcal{F}^{(k)} \biguplus_{(a_0^{(k)},c_0)} \mathcal{C}_{\ell+1}.$$

Note that $s(\mathcal{G}_n) = (k+1)s(\mathcal{F}) + s(\mathcal{C}_{\ell+1}) = (k+1)\ell + \ell$. Then, it is easily shown that $\mathcal{G}_n$ satisfies $(\Box^n \delta_{\mathcal{F}}) \wedge p_0$ at the point $a_0^{(0)}$ under the valuation V given by $V(p_i) := \{a_i^{(j)} : 0 \leq j \leq k\}$ for $0 \leq i \leq N$.

Put $m := (k+1)\ell + \ell$, and suppose that $\mathcal{G}_n$ satisfies $(\Box^m \delta_{\mathcal{F}}) \wedge p_0$ at some point a under a valuation V. Arguing as before, we show that for every point in $\mathcal{G}_n$ there is one and only one variable that is true at this point under V. In particular, it is so for every point in the part $\mathcal{C}_{\ell+1}$ of $\mathcal{G}_n$. Thus we may assume that for every point $c_i \in \mathcal{C}_{\ell+1}$ $(0 \leq i \leq \ell)$, there is a variable $p_{f(i)}$ such that $c_i \models p_{f(i)}$. Put $A := \{a_{f(0)}, a_{f(1)}, a_{f(2)}, \ldots, a_{f(\ell)}\} \subseteq W$ in $\mathcal{F}$. Considering the degree of every c_i in $\mathcal{C}_{\ell+1}$ in $\mathcal{G}_n$, we get that $deg(a_{f(j)}) \leq 2$ for $1 \leq j < \ell$. It is possible that $deg(a_{f(0)})$ is larger than 2, but it is impossible that $a_{f(0)} R a_{f(l)}$, for $\Diamond p_{f(0)}$ is false at c_l. Thus A forms a subchain in $\mathcal{F}$, and since $\mathcal{F}$ is connected, there is a point $a_t \in W - A$ such that $a_{f(0)} R a_t$. Furthermore, it is only $a_{f(0)}$ in A that is connected to a point in $W - A$; this can be shown again by looking at degrees of points in $\mathcal{C}_{\ell+1}$.

Now we will prove that every point in A is distinct from each other. First, the point $a_{f(0)}$ is distinct from any point in $A - \{a_{f(0)}\}$. For suppose there is some j with $1 \leq j < \ell$, such that $a_{f(0)} = a_{f(j)}$. Then $a_{f(0)} R a_t$ in $\mathcal{F}$ implies that $c_j \models \Diamond p_{a_t}$ must hold, but this is impossible, since no point in $\{c_1, c_2, \ldots, c_\ell\}$ can see a point at which p_{a_t} is true under V. Second, the point $a_{f(1)}$ is distinct from any point in $A - \{a_{f(0)}, a_{f(1)}\}$. For suppose there is some j with $2 \leq j \leq \ell$, such that $a_{f(1)} = a_{f(j)}$. Then $a_{f(1)} R^2 a_t$ in $\mathcal{F}$ implies that $c_j \models \Diamond^2 p_{a_t}$ must hold, but this is impossible, since no point in $\{c_2, \ldots, c_\ell\}$ can in two steps reach a point at which p_{a_t} is true under V. Similarly we can show that for any $i \in [0, \ell - 1]$, the point $a_{f(i)}$ is different from any point in $\{a_{f(i+1)}, a_{f(i+2)}, \ldots, a_{f(\ell)}\}$. Therefore every point in A is distinct from each other, and so A forms a subchain of length ℓ in $\mathcal{F}$. But this leads to a contradiction since $s(\mathcal{F}) = \ell$ and $\mathcal{F}$ is not a chain. This completes our proof. ■

Putting Theorems 5.1 and 5.2 together, we obtain our main result.

THEOREM 5.3. *Let $\mathcal{F} := \langle W, R\rangle$ be a finite and connected KTB-frame, where $card(W) \geq 3$. Then $\mathcal{F}$ cannot split the lattice* **NExt(KTB)**.

6. Some questions and conjectures

For a normal modal logic $\mathbf{L}$ extending $\mathbf{L_0}$, following Fine [10], we define the *degree of* (Kripke) *incompleteness* $\xi_{\mathbf{NExt(K)}}(\mathbf{L})$ of $\mathbf{L}$ in $\mathbf{NExt}(\mathbf{L_0})$ by taking

$$\xi_{\mathbf{NExt(L_0)}}(\mathbf{L}) := card(\{\mathbf{L}' \in \mathbf{NExt}(\mathbf{L_0})\colon \forall\, \mathcal{F}, \mathcal{F} \models \mathbf{L} \text{ if and only if } \mathcal{F} \models \mathbf{L}'\}).$$

In [1] Blok obtained a complete characterisation of $\xi_{\mathbf{NExt(K)}}$ that we recall below.

Theorem 6.1 (Blok). *Let* $\mathbf{L}$ *be a normal modal logic.* $\xi_{\mathbf{NExt(K)}}(\mathbf{L}) = 2^{\aleph_0}$ *if and only if* $\mathbf{L}$ *is a join of splitting logics and* $\mathbf{L} \neq \Phi$*. Otherwise* $\xi_{\mathbf{NExt(K)}}(\mathbf{L})$ *is equal to* 1*.*

A modal logic $\mathbf{M}$ is a cocover of $\mathbf{L}$ if and only if (1) $\mathbf{M} \subseteq \mathbf{L}$, and (2) for any modal logic $\mathbf{M}'$, $\mathbf{M} \subseteq \mathbf{M}' \subseteq \mathbf{L}$ implies $\mathbf{M} = \mathbf{M}'$ or $\mathbf{M}' = \mathbf{L}$. For a normal modal logic $\mathbf{L}$ extending $\mathbf{L_0}$, we define $\chi_{\mathbf{NExt(L_0)}}(\mathbf{L})$ to be the number of cocovers of $\mathbf{L}$ in $\mathbf{NExt}(\mathbf{L_0})$. Blok [1] also contains the following characterisation of $\chi_{\mathbf{NExt(K)}}$.

Theorem 6.2 (Blok). *Let* $\mathbf{L}$ *be a normal modal logic.* $\chi_{\mathbf{NExt(K)}}(\mathbf{L}) = 2^{\aleph_0}$ *if and only if* $\mathbf{L}$ *is not a join of splitting logics and* $\mathbf{L} \neq \Phi$*. Otherwise* $\chi_{\mathbf{NExt(K)}}(\mathbf{L}) \leq \aleph_0$*.*

Both ξ and χ depend on what logic we take as base. It would be of considerable interest to establish a complete description of ξ and χ with respect to $\mathbf{KTB}$. However, our main theorem leaves little hope of obtaining a characterisation $\xi_{\mathbf{NExt(KTB)}}$ or $\chi_{\mathbf{NExt(KTB)}}$ similar to the ones for $\mathbf{NExt}(\mathbf{K})$. Comparing Theorem 5.3 with Theorem 3.4 reveals that $\mathbf{NExt}(\mathbf{KTB})$ may be similar to $\mathbf{NExt}(\mathbf{KT})$ instead. For $\xi_{\mathbf{NExt(KT)}}$ and $\chi_{\mathbf{NExt(KT)}}$, Blok also proved the following results ([1, 2]).

Theorem 6.3 (Blok). *Let* $\mathbf{L} \in \mathbf{NExt}(\mathbf{KT})$*, and* $\mathbf{L} \neq \Phi$*. Then,*

1. $\xi_{\mathbf{NExt(KT)}}(\mathbf{L}) = 2^{\aleph_0}$.
2. $\chi_{\mathbf{NExt(KT)}}(\mathbf{L}) = 2^{\aleph_0}$.

By analogy, the following conjecture seems to be plausible.

Conjecture 1. *Let* $\mathbf{L} \in \mathbf{NExt}(\mathbf{KTB})$ *and* $\mathbf{L} \subseteq \mathbf{L}(\circ\!\!-\!\!\circ)$*. Then,*

1. $\xi_{\mathbf{NExt(KTB)}}(\mathbf{L}) = 2^{\aleph_0}$.
2. $\chi_{\mathbf{NExt(KTB)}}(\mathbf{L}) = 2^{\aleph_0}$.

As far as the authors can tell, not much is known about these questions at present. As for (1) above, we know that there is a continuum of Kripke incomplete logics in $\mathbf{NExt}(\mathbf{KTB})$ (see [20]). However, all known examples of Kripke-incomplete logics $\mathbf{NExt}(\mathbf{KTB})$ are not finitely axiomatisable. We do not feel in a position to venture any conjectures here.

Question 1. *Is there a KTB-logic which is Kripke-incomplete and finitely axiomatisable?*

As for (2) in the above conjecture, Stevens and the first author have recently obtained the following partial result (cf. [22]).

Proposition 6.4 (Stevens and Kowalski). *The logic* $\mathbf{L}(\circ\!-\!\circ)$ *has at least* $\aleph_0$ *cocovers in* $\mathbf{NExt}(\mathbf{KTB})$*, that is,* $\chi_{\mathbf{NExt}(\mathbf{KTB})}(\mathbf{L}(\circ\!-\!\circ)) \geq \aleph_0$.

Question 2. *Does the logic* $\mathbf{L}(\circ\!-\!\circ)$ *have uncountably many cocovers in* $\mathbf{NExt}(\mathbf{KTB})$*?*

All we can say at present in regard to this question[2] is that 'empirical' evidence suggests strongly that if such an uncountable family exists, it should be n-transitive for some $n \in \omega$, that is, it should verify the formula $\Box^{n+1}p \leftrightarrow \Box^{n}p$.

Since the method used to gain a description of ξ might be also used to obtain a characterisation of χ with respect to our target modal logics, it is reasonable to expect that when one of the above questions is solved, the solutions to others may follow with less difficulty. Obtaining a clearer view of the lattice $\mathbf{NExt}(\mathbf{KTB})$ would certainly advance the knowledge about symmetric modal logics (i.e., logics extending **KB**), which remains quite a way behind the knowledge about transitive ones (extending **K4**).

Acknowledgements. We are grateful to Tadeusz Litak for many discussions and comments during the preparation of this paper. We are also grateful to an anonymous referee whose insightful comments helped to improve the presentation greatly. Being grateful to a and being grateful to b does not mean that $a \neq b$, of course. The second author's contribution to this research was supported by Grant-in-Aid for Young Scientists (B), No. 18700002, Japan Society for the Promotion of Science.

References

[1] Blok, W., *On the degree of incompleteness in modal logics and the covering relation in the lattice of modal logics*, Tech. Rep. 78-07, Department of Mathematics, University of Amsterdam, 1978.

[2] Blok, W., 'The lattice of modal logics: an algebraic investigation', *Journal of Symbolic Logic*, 45: 221–236, 1980.

[3] Blok, W., Köhler, P., Pigozzi, D., 'On the structure of varieties with equationally definable principal congruences II', *Algebra Universalis*, 18: 334–379.

---

[2] Added in proof: Stevens and the first author have recently obtained a positive answer.

[4] Blok, W., Pigozzi, D., 'On the structure of varieties with equationally definable principal congruences I', *Algebra Universalis*, 15: 195–227.

[5] Blok, W., Pigozzi, D., 'On the structure of varieties with equationally definable principal congruences III', *Algebra Universalis*, 32: 545–608.

[6] Blok, W., Pigozzi, D., 'On the structure of varieties with equationally definable principal congruences IV', *Algebra Universalis*, 31: 1–35.

[7] Burris, S.N., Sankappanavar, H.P., *A Course in Universal Algebra*, Springer Verlag, Berlin, 1981.

[8] Chagrov, A., Zakharyaschev, M., *Modal Logic*, Clarendon Press, Oxford.

[9] Day, A., 'Splitting lattices generate all lattices', *Algebra Universalis*, 7: 163–169.

[10] Fine, K., 'An incomplete logic containing S4', *Theoria*, 40: 23–29, 1974.

[11] Jankov, V.A., 'The relationship between deducibility in the intuitionistic propositional calculus and finite implicational structures', *Soviet Mathematics Doklady*, 4: 1203–1204.

[12] Kowalski, T., *An outline of a topography of tense logics*, Ph.D. thesis, Jagiellonian University, Kraków, 1997.

[13] Kowalski, T., Ono, H., 'Splittings in the variety of residuated lattices', *Algebra Universalis*, 44: 283–298.

[14] Kracht, M., 'Even more about the lattice of tense logics', *Archive of Mathematical Logic*, 31: 243–357.

[15] Kracht, M., *Tools and Techniques in Modal Logic*, Studies in Logics, vol. 42, Elsevier, Amsterdam.

[16] Kracht, M., 'An almost general splitting theorem for modal logic', *Studia Logica*, 49: 455–470, 1990.

[17] Makinson, D.C., 'Some embedding theorems for modal logic', *Notre Dame Journal of Formal Logic*, 12: 252–254, 1971.

[18] McKenzie, R., 'Equational bases and non-modular lattice varieties', *Transactions of the American Mathematical Society*, 156: 1–43.

[19] Miyazaki, Y., 'A splitting logic in NEXT(KTB)', *Studia Logica*, 85: 399–412, 2007.

[20] Miyazaki, Y., 'Kripke incomplete logics containing KTB', *Studia Logica*, 85: 311–326, 2007.

[21] Rautenberg, W., 'Splitting lattices of logics', *Archiv für Mathematische Logik*, 20: 155–159.

[22] Stevens, M., Kowalski T., *Minimal varieties of KTB-algebras*, manuscript.

[23] Whitman, Ph.M., 'Splittings of a lattice', *American Journal of Mathematics*, 65: 179–196.

[24] Wolter, F., *Lattices of modal logics*, Ph.D. thesis, Freie Universität, Berlin, 1993.

Tomasz Kowalski
Research School of Information Sciences
and Engineering
Australian National University
bldg 115, cnr North and Daley Rds
Canberra ACT, Australia
`tomasz.kowalski@anu.edu.au`

Yutaka Miyazaki
Meme Media Laboratory
Hokkaido University
Kita 13, Nishi 8, Kita-ku
Sapporo 060-8628, JAPAN
`y-miya@meme.hokudai.ac.jp`

Thomas Ågotnes
Wiebe van der Hoek
Juan A. Rodríguez-Aguilar
Carles Sierra
Michael Wooldridge

A Temporal Logic of Normative Systems

Abstract. We study *Normative Temporal Logic* (NTL), a formalism intended for reasoning about the temporal properties of normative systems. NTL is a generalisation of the well-known branching-time temporal logic CTL, in which the path quantifiers A ("on all paths…") and E ("on some path…") are replaced by the indexed deontic operators O_η ("it is obligatory in the context of the normative system η that …") and P_η ("it is permissible in the context of the normative system η that…"). After introducing the logic, we give a sound and complete axiomatisation. We then present a symbolic representation language for normative systems, and we identify four different model checking problems, corresponding to whether or not a model is represented symbolically or explicitly, and whether or not we are given a concrete interpretation for the normative systems named in formulae to be model checked. We show that the complexity of model checking varies from P-complete in the simplest case (explicit state model checking where we are given a specific interpretation for all normative systems in the formula) up to EXPTIME-hard in the worst case (symbolic model checking, no interpretation given). We present examples to illustrate the use of NTL, and conclude with discussions of related work (in particular, the relationship of NTL to other deontic logics), and some issues for future work.

Keywords: Normative Systems, Temporal Logic, Multi-Agent Systems.

1. Introduction

Normative systems, or social laws, have been widely promoted as an approach to coordinating multi-agent systems [26, 25, 20, 27, 28, 18]. Crudely, a normative system defines a set of constraints on the behaviour of agents, corresponding to obligations, which may or may not be observed by agents. The designer of a normative system typically has some objective in mind, such that if the constraints of the normative system are observed, then the objective is achieved [18].

A number of formalisms have been proposed for reasoning about normative behaviour in multi-agent systems, typically based on deontic logic [30, 12, 19]. However the computational properties of such formalisms — in particular, their use in the practical design and synthesis of normative systems

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*, Trends in Logic, 28

and the complexity of reasoning with them — has received relatively little attention. In this paper, we rectify this omission. We present a logic for reasoning about normative systems, which is closely related to the successful and widely-used temporal logic CTL [14]. The idea underpinning *Normative Temporal Logic* (NTL) is to replace the universal and existential path quantifiers of CTL with indexed deontic operators O_η and P_η, where $\mathsf{O}_\eta\varphi$ means that "φ is obligatory in the context of the normative system η", and $\mathsf{P}_\eta\varphi$ means "φ is permissible in the context of the normative system η". Here, φ is a temporal logic expression over the usual CTL temporal operators $\bigcirc$, $\Diamond$, $\Box$, and $\mathcal{U}$, and a syntactic construction rule similar to that in CTL applies: every temporal operator must be preceded by a deontic operator. A normative system η is understood to be a set of constraints on the behaviour of agents within the system. In NTL, obligations and permissions are thus, first, contextualised to a normative system η and, second, have a temporal dimension. NTL generalises CTL because by letting $\eta_\emptyset$ denote the empty normative system, which places no constraints on the behaviour of agents, the universal path quantifier A can be interpreted as $\mathsf{O}_{\eta_\emptyset}$. Because of its close relationship to CTL, much of the technical machinery developed for reasoning with CTL can be adapted for use in NTL [14, 11].

The remainder of the paper is structured as follows. After introducing the logic, we give a sound and complete axiomatisation. We then present a symbolic representation language for normative systems. We investigate the complexity of NTL model checking, and identify four different variations of the model checking problem, depending on whether a model is represented symbolically or explicitly, and whether we are given a concrete interpretation for the normative systems named in formulae to be model checked. We show that the complexity of model checking varies from P-complete in the simplest case (explicit state model checking where we are given a specific interpretation for all normative systems in the formula) up to EXPTIME-hard in the worst case (symbolic model checking, no interpretation given). We present two examples to illustrate the use of the logic. We conclude with a discussion of related work, (in particular, a discussion of the relation to other deontic and deontic temporal logics), and some issues for future research.

2. Normative Temporal Logic

2.1. Kripke Structures

Let $\Phi = \{p, q, \ldots\}$ be a finite set of atomic *propositional variables.* A *Kripke structure* (over Φ) is a quadruple

$$\mathcal{K} = \langle S, S^0, R, V \rangle,$$

where:

- S is a finite, non-empty set of *states*, with S^0 being the *initial states* ($\emptyset \subset S^0 \subseteq S$);
- $R \subseteq S \times S$ is a total binary relation on S, which we refer to as the *transition relation*[1]; and
- $V : S \to 2^\Phi$ labels each state with the set of propositional variables true in that state.

A *path* over R is an infinite sequence of states $\pi = s_0, s_1, \ldots$ which must satisfy the property that $\forall u \in \mathbb{N}$: $(s_u, s_{u+1}) \in R$. If $u \in \mathbb{N}$, then we denote by $\pi[u]$ the component indexed by u in π (thus $\pi[0]$ denotes the first element, $\pi[1]$ the second, and so on). A path π such that $\pi[0] = s$ is an *s-path.*

2.2. Normative Systems

Normative systems have come to play a major role in multi-agent systems research; for example, under the name of *social laws*, they have been shown to be a useful mechanism for coordination [27]. In this paper, a normative system should be understood simply as *a set of constraints on the behaviour of agents in a system.* More precisely, a normative system defines, for every possible system transition, whether or not that transition is considered to be legal or not, in the context of the normative system. Different normative systems may differ on whether or not a particular transition is considered legal. Formally, a normative system η (w.r.t. a Kripke structure $\mathcal{K} = \langle S, S^0, R, V \rangle$) is simply a subset of R, such that $R \setminus \eta$ is a total relation. We refer to the requirement that $R \setminus \eta$ is total as a *reasonableness* requirement: it prevents social laws which lead to states with no allowed successor. Let $N(R) = \{\eta \mid \eta \subseteq R \ \&\ R \setminus \eta \text{ is total}\}$ be the set of normative systems over R. The intended interpretation of a normative system η is that the presence of an arc (s, s') in η means that the transition (s, s') is forbidden in the context of η, hence, $R \setminus \eta$ denotes the allowed transitions. Since it is assumed that η is reasonable, we are guaranteed that such a transition always exists for every state. If π is a path over R and η is a normative system over R, then we say that π is *η-conformant* if it does not contain

---

[1] In the temporal logic literature, it is common to refer to a relation $R \subseteq S \times S$ as being total if $\forall s \in S, \exists s' \in S : (s, s') \in R$.

any transition that is forbidden by η, i.e., if $\forall u \in \mathbb{N}$, $(\pi[u], \pi[u+1]) \notin \eta$. We denote the set of η-conformant s-paths (w.r.t. some assumed R) by $\mathcal{C}_\eta(s)$.

Since normative systems in our view are just *sets* (of disallowed transitions), we can *compare* them, to determine, for example, whether one is *more liberal* (less restrictive) than another: if $\eta \subset \eta'$, then η places fewer constraints on a system than η', and hence η is more liberal. Notice that, assuming an *explicit* representation of normative systems, (i.e., representing a normative system η directly as a subset of R), checking such properties can be done in polynomial time. We can also operate on them with the standard set theoretic operations of union, intersection, etc. Taking the union of two normative systems η_1 and η_2 may yield (depending on whether $R \setminus (\eta_1 \cup \eta_2)$ is total) a normative system that is *more restrictive* (less liberal) than either of its parent systems, while taking the *intersection* of two normative systems will yield a normative system which is *less restrictive* (more liberal). The $\cup$ operation is intuitively the act of superposition, or composition of normative systems: imposing one law on top of another. Notice that, when operating on normative systems using such set theoretic operations, care must be taken to ensure the resulting normative system is reasonable.

EXAMPLE 2.1. *Consider two parallel circular train tracks. At one point both tracks go through the same tunnel. At the east and the west end of the tunnel there are traffic lights, which can be either green or red. A train controller controls the lights. The eastern light should be set to green if and only if there is a train waiting to enter the east end of the tunnel and there is no train waiting at the west end of the tunnel, and similarly for the western light. One train travels on each of the tracks, in opposite directions. We call the train that enters the tunnel at the eastern end the east train and the other train the west train. Obviously, the trains should not enter the tunnel if the light is red.*

We can model this situation by considering the physical properties and the normative properties separately*, as Kripke structures and normative systems respectively. We assume that each train can be in one of three states:* tunnel *(the train is in the tunnel);* waiting *(the train is waiting to enter the tunnel);* away *(the train is neither in the tunnel nor waiting). When away, the train can either be away or waiting in the next state; when waiting the train can either be waiting or in the tunnel in the next state; when the train is in the tunnel it leaves the tunnel and is away in the next state. Thus, we use propositional atoms* $eTunnel$, $eWaiting$, $eAway$, $wTunnel$, $wWaiting$, $wAway$ *to encode the position of the east and west train. We also use atoms* $eGreen$ *and* $wGreen$ *to represent the fact that the eastern/western lights*

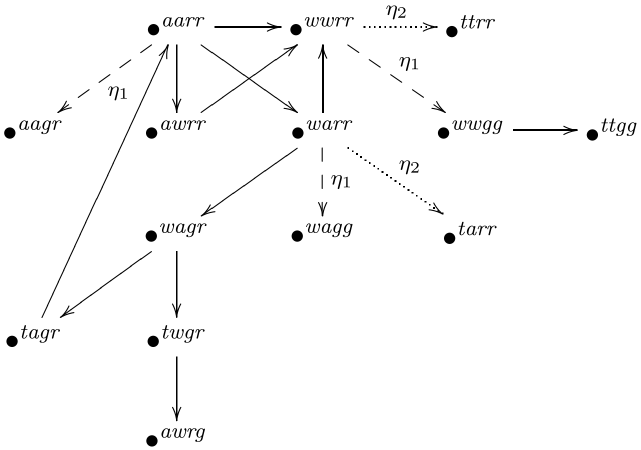


Figure 1. Kripke model of the trains example, including all physically possible transitions. Only a part of the model is shown. The transitions prohibited by the normative systems η_1 and η_2 are shown with dashed and dotted lines, respectively. The labelling of the states is abbreviated for readability: "twgr" stands for tunnel-waiting-green-red and means that $wTunnel$, $eWaiting$, $wGreen$ are true and that all other atoms (including $eGreen$) are false.

are green. Thus, $\neg eGreen$ means that the eastern light is red, and so on. Let $\mathcal{K}$ be the Kripke structure where the states correspond to all possible configurations of the atomic propositions, the (single) initial state is the state where both lights are red and both trains away, and the transitions are all physically possible transitions — illustrated in Figure 1. The transitions include entering on a red light, but exclude physically impossible transitions such as a train going directly from the tunnel *state to the* waiting *state.*

Let η_1 be the normative system corresponding to the normative requirement on the switching of the lights described above: η_1 contains all transitions between states s_1 and s_2 in which one of the lights are set to green (in s_2) without *the appropriate condition (as explained above) being true in s_1. The normative system η_1 is illustrated by labels on the transitions in Figure 1. The description above contains another normative requirement as well: trains should only enter the tunnel on a green light. Let η_2 be the normative system corresponding to that requirement: η_2 contains all transitions between states s_1 and s_2 such that a train is in the tunnel in s_2 only if the corresponding light is green in s_1. It is easy to see that $\eta_1, \eta_2 \in N(R)$, where R is the transition relation of $\mathcal{K}$.*

Finally, while the norms in this particular example are designed to avoid a crash, there are other problems, such as "deadlock" (both trains can wait forever for a green light), which they do not avoid. For simplicity, we will only consider the norms mentioned above.

2.3. Syntax of NTL

The language of NTL is a generalisation of CTL: the only issue that may cause confusion is that, within this language, we refer explicitly to normative systems, which are of course *semantic* objects. We will therefore assume a stock of syntactic elements Σ_η which will denote normative systems. An *interpretation* for symbols Σ_η with respect to a transition relation R is a function $I : \Sigma_\eta \to N(R)$. When R is a transition relation of Kripke structure $\mathcal{K}$ we say that I is an interpretation over $\mathcal{K}$. We will assume that the symbol $\eta_\emptyset$ always denotes the "emptyset" normative system, i.e., the normative system which forbids *no* transitions. Note that this normative system will be well-defined for *any* Kripke structure. Thus, we require that all interpretations I satisfy the property that $I(\eta_\emptyset) = \emptyset$. If the interpretation function I is clear from context or not relevant, we will sometimes identify the symbol η with the normative system it denotes.

The syntax of NTL is defined by the following grammar:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathsf{P}_\eta \bigcirc \varphi \mid \mathsf{P}_\eta(\varphi \mathcal{U} \varphi) \mid \mathsf{O}_\eta \bigcirc \varphi \mid \mathsf{O}_\eta(\varphi \mathcal{U} \varphi)$$

where $p \in \Phi$ is a propositional variable and $\eta \in \Sigma_\eta$ denotes a normative system. Sometimes we call α occurring in an expression $\mathsf{O}_\eta\alpha$ or $\mathsf{P}_\eta\alpha$ a *temporal formula* (although such an α is not a well-formed formula of NTL).

2.4. Semantic Rules

The semantics of NTL are given with respect to the satisfaction relation "$\models$". $\mathcal{K}, s \models_I \varphi$ holds when $\mathcal{K}$ is a Kripke structure, s is a state in $\mathcal{K}$, I an interpretation over $\mathcal{K}$, and φ a formula of the language, as follows:

$\mathcal{K}, s \models_I \top$;

$\mathcal{K}, s \models_I p$ iff $p \in V(s)$ (where $p \in \Phi$);

$\mathcal{K}, s \models_I \neg\varphi$ iff not $\mathcal{K}, s \models_I \varphi$;

$\mathcal{K}, s \models_I \varphi \vee \psi$ iff $\mathcal{K}, s \models_I \varphi$ or $\mathcal{K}, s \models_I \psi$;

$\mathcal{K}, s \models_I \mathsf{O}_\eta \bigcirc \varphi$ iff $\forall \pi \in \mathcal{C}_{I(\eta)}(s) : \mathcal{K}, \pi[1] \models_I \varphi$;

$\mathcal{K}, s \models_I \mathsf{P}_\eta \bigcirc \varphi$ iff $\exists \pi \in \mathcal{C}_{I(\eta)}(s) : \mathcal{K}, \pi[1] \models_I \varphi$;

$\mathcal{K}, s \models_I \mathsf{O}_\eta(\varphi \mathcal{U} \psi)$ iff $\forall \pi \in \mathcal{C}_{I(\eta)}(s), \exists u \in \mathbb{N}$, s.t. $\mathcal{K}, \pi[u] \models_I \psi$ and $\forall v, (0 \leq v < u) : \mathcal{K}, \pi[v] \models_I \varphi$

$\mathcal{K}, s \models_I \mathsf{P}_\eta(\varphi \mathcal{U} \psi)$ iff $\exists \pi \in \mathcal{C}_{I(\eta)}(s), \exists u \in \mathbb{N}$, s.t. $\mathcal{K}, \pi[u] \models_I \psi$ and $\forall v, (0 \leq v < u) : \mathcal{K}, \pi[v] \models_I \varphi$

The remaining classical logic connectives ("$\wedge$", "$\rightarrow$", "$\leftrightarrow$") are assumed to be defined as abbreviations in terms of $\neg$ and $\vee$, in the conventional manner. We define the remaining CTL-style operators for $\Diamond$ and $\Box$ as abbreviations:

$$\begin{array}{rcl} \mathsf{O}_\eta \Diamond \varphi & \equiv & \mathsf{O}_\eta(\top\, \mathcal{U}\, \varphi) \\ \mathsf{P}_\eta \Diamond \varphi & \equiv & \mathsf{P}_\eta(\top\, \mathcal{U}\, \varphi) \\ \mathsf{O}_\eta \Box \varphi & \equiv & \neg \mathsf{P}_\eta \Diamond \neg \varphi \\ \mathsf{P}_\eta \Box \varphi & \equiv & \neg \mathsf{O}_\eta \Diamond \neg \varphi \end{array}$$

Recalling that $\eta_\emptyset$ denotes the empty normative system, we obtain the conventional path quantifiers of CTL [14] as follows:

$$\begin{array}{rcl} \mathsf{A}\alpha & \equiv & \mathsf{O}_{\eta_\emptyset}\alpha \\ \mathsf{E}\alpha & \equiv & \mathsf{P}_{\eta_\emptyset}\alpha \end{array}$$

Thus the CTL universal path quantifier can be understood as obligation in the context of the empty normative system, which places no restrictions on which transitions the system takes, while the existential path quantifier can be understood as permission in the context of this normative system.

We write $\mathcal{K} \models_I \varphi$ if $\mathcal{K}, s_0 \models_I \varphi$ for all $s_0 \in S^0$, $\mathcal{K} \models \varphi$ if $\mathcal{K} \models_I \varphi$ for all I, and $\models \varphi$ if $\mathcal{K} \models \varphi$ for all $\mathcal{K}$.

EXAMPLE 2.2 (Example 2.1 continued). *Let $\mathcal{K}, \eta_1, \eta_2$ be as in example 2.1. Let I be such that $I(\eta_1) = \eta_1$, $I(\eta_2) = \eta_2$, $I(\eta_3) = \eta_1 \cup \eta_2$ (it is easy to see that also $\eta_1 \cup \eta_2 \in N(R)$). Let the formula*

$$crash = eTunnel \wedge wTunnel$$

denote a crash situation. We have that (recall that $\mathcal{K} \models_I \varphi$ means that φ is satisfied in all the initial states of $\mathcal{K}$ under I):

- $\mathcal{K} \models_I \mathsf{O}_{\eta_1} \bigcirc \neg wGreen$. *In the initial state, according to normative system η_1 it is obligatory that the western light stays red in the next state.*
- $\mathcal{K} \models_I \mathsf{P}_{\eta_1}(\neg eGreen\, \mathcal{U}\, eTunnel)$. *$\eta_1$ permits the eastern light to stay red until the east train is in the tunnel.*
- $\mathcal{K} \models_I \neg \mathsf{P}_{\eta_2}(\neg eGreen\, \mathcal{U}\, eTunnel)$. *$\eta_2$ does not permit the eastern light to stay red until the east train is in the tunnel.*
- $\mathcal{K} \models_I \mathsf{O}_{\eta_1} \Box (wGreen \rightarrow \neg eGreen)$. *It is obligatory in the context of η_1 that at least one of the lights are red.*
- $\mathcal{K} \models_I \mathsf{P}_{\eta_\emptyset} \Diamond crash$. *Without any constraining norms, the system permits a crash in the future.*

- $\mathcal{K} \models_I \mathsf{P}_{\eta_1} \Diamond crash$. *The normative system* η_1 *permits a crash.*
- $\mathcal{K} \models_I \mathsf{O}_{\eta_3} \Box \neg crash$. *It is obligatory, in the context of normative system* η_3*, that a crash never occurs;* η_3 *does not permit a crash at any point in the future.*

The following are examples of expressions involving nested operators. It is worth reflecting on the compositional meaning of nested operators. For example, $\mathsf{P}_{\eta_3} \Diamond \mathsf{P}_{\eta_1} \bigcirc crash$ *means that* η_3 *permits a computation along which in some future state* $\mathsf{P}_{\eta_1} \bigcirc crash$ *is true. However, in the evaluation of* $\mathsf{P}_{\eta_1} \bigcirc crash$ *in states along that computation, the system is* not *restricted by* η_3 *(but only by* η_1*).*

- $\mathcal{K} \models_I \mathsf{O}_{\eta_\emptyset} \Box ((wWaiting \wedge \neg wGreen) \rightarrow \neg \mathsf{P}_{\eta_2} \bigcirc wTunnel)$. *It is obligatory in the system that it is always the case that if the west train is waiting and the western light is red then the western train is not permitted by* η_2 *in the tunnel in the next state.*
- $\mathcal{K} \models_I \mathsf{P}_{\eta_2} \Diamond \mathsf{P}_{\eta_3} \bigcirc crash$. η_2 *permits a future state where a crash in the next state is permitted even by* η_3*.*
- $\mathcal{K} \models_I \mathsf{P}_{\eta_3} \Diamond \mathsf{P}_{\eta_1} \bigcirc crash$. η_3 *permits a future state where a crash in the next state is permitted by* η_1*.*
- $\mathcal{K} \models_I \mathsf{O}_{\eta_3} \Box \mathsf{O}_{\eta_2} \bigcirc \neg crash$. η_3 *does not permit a future state where a crash is permitted in the next state by* η_2*.*

2.5. Properties and Axiomatisation

The following Proposition makes precise the expected property that *a less liberal system has more obligations and less permissions than a more liberal system.*

PROPOSITION 2.3. *Let* $\mathcal{K}$ *be a Kripke structure,* I *an interpretation over* $\mathcal{K}$ *and* $\eta_1, \eta_2 \in \Sigma_\eta$.

$$\text{If } I(\eta_1) \subseteq I(\eta_2) \text{ then } \mathcal{K} \models_I \mathsf{O}_{\eta_1}\varphi \rightarrow \mathsf{O}_{\eta_2}\varphi \text{ and } \mathcal{K} \models_I \mathsf{P}_{\eta_2}\varphi \rightarrow \mathsf{P}_{\eta_1}\varphi$$

We now go on to exhaustively describe the universally valid properties, of NTL as well as some derived systems, by presenting sound and complete axiomatisations.

First, let NTL$^-$ be NTL without the empty normative system. Formally, NTL$^-$ is defined exactly as NTL, except for the requirement that Σ_η contains the $\eta_\emptyset$ symbol and the corresponding restriction on interpretations. An

(Ax1) All validities of propositional logic

(Ax2) $\mathsf{P}_\eta \Diamond \varphi \leftrightarrow \mathsf{P}_\eta(\top \mathcal{U} \varphi)$

(Ax2b) $\mathsf{O}_\eta \Box \varphi \leftrightarrow \neg \mathsf{P}_\eta \Diamond \neg \varphi$

(Ax3) $\mathsf{O}_\eta \Diamond \varphi \leftrightarrow \mathsf{O}_\eta(\top \mathcal{U} \varphi)$

(Ax3b) $\mathsf{P}_\eta \Box \varphi \leftrightarrow \neg \mathsf{O}_\eta \Diamond \neg \varphi$

(Ax4) $\mathsf{P}_\eta \bigcirc (\varphi \vee \psi) \leftrightarrow (\mathsf{P}_\eta \bigcirc \varphi \vee \mathsf{P}_\eta \bigcirc \psi)$

(Ax5) $\mathsf{O}_\eta \bigcirc \varphi \leftrightarrow \neg \mathsf{P}_\eta \bigcirc \neg \varphi$

(Ax6) $\mathsf{P}_\eta(\varphi \mathcal{U} \psi) \leftrightarrow (\psi \vee (\varphi \wedge \mathsf{P}_\eta \bigcirc \mathsf{P}_\eta(\varphi \mathcal{U} \psi)))$

(Ax7) $\mathsf{O}_\eta(\varphi \mathcal{U} \psi) \leftrightarrow (\psi \vee (\varphi \wedge \mathsf{O}_\eta \bigcirc \mathsf{O}_\eta(\varphi \mathcal{U} \psi)))$

(Ax8) $\mathsf{P}_\eta \bigcirc \top \wedge \mathsf{O}_\eta \bigcirc \top$

(Ax9) $\mathsf{O}_\eta \Box(\varphi \rightarrow (\neg \psi \wedge \mathsf{P}_\eta \bigcirc \varphi)) \rightarrow (\varphi \rightarrow \neg \mathsf{O}_\eta(\gamma \mathcal{U} \psi))$

(Ax9b) $\mathsf{O}_\eta \Box(\varphi \rightarrow (\neg \psi \wedge \mathsf{P}_\eta \bigcirc \varphi)) \rightarrow (\varphi \rightarrow \neg \mathsf{O}_\eta \Diamond \psi)$

(Ax10) $\mathsf{O}_\eta \Box(\varphi \rightarrow (\neg \psi \wedge (\gamma \rightarrow \mathsf{O}_\eta \bigcirc \varphi))) \rightarrow (\varphi \rightarrow \neg \mathsf{P}_\eta(\gamma \mathcal{U} \psi))$

(Ax10b) $\mathsf{O}_\eta \Box(\varphi \rightarrow (\neg \psi \wedge \mathsf{O}_\eta \bigcirc \varphi)) \rightarrow (\varphi \rightarrow \neg \mathsf{P}_\eta \Diamond \psi$

(Ax11) $\mathsf{O}_\eta \Box(\varphi \rightarrow \psi) \rightarrow (\mathsf{P}_\eta \bigcirc \varphi \rightarrow \mathsf{P}_\eta \bigcirc \psi)$

(R1) If $\vdash \varphi$ then $\vdash \mathsf{O}_\eta \Box \varphi$ (generalization)

(R2) If $\vdash \varphi$ and $\vdash \varphi \rightarrow \psi$ then $\vdash \psi$ (modus ponens)

(Obl) $\mathsf{O}_{\eta_\emptyset} \alpha \rightarrow \mathsf{O}_\eta \alpha$

(Perm) $\mathsf{P}_\eta \alpha \rightarrow \mathsf{P}_{\eta_\emptyset} \alpha$

Figure 2. The two systems NTL$^-$ ((Ax1)–(R2), derived from an axiomatisation of CTL [14]) and NTL ((Ax1)–(R2),(Obl),(Perm)). α stands for a temporal formula.

axiom system for NTL$^-$, denoted $\vdash^-$, is defined by axioms and rules (Ax1)–(R2) in Figure 2. NTL$^-$ can be seen as a *multi-dimensional* variant of CTL, where there are several indexed versions of each path quantifier[2]. Indeed, the axiomatisation has been obtained from an axiomatisation of CTL [14].

Going on to NTL, we add axioms (Obl) and (Perm) (Figure 2); the corresponding inference system is denoted $\vdash$. We then (by soundness, see below), have the following chain of implications in NTL (the second element in the

---

[2] Semantically, we can view the NTL structures as multi-dimensional CTL structures with one (total) transition relation $R \setminus I(\eta)$ for each normative system. This definition of multi-dimensional structures is different from *multiprocess temporal structures* as defined in [5, 14]. In the latter, only the *union* of the transition relations is required to be total.

chain is a variant of a deontic axiom discussed later). If something is naturally, or physically inevitable, then it is obligatory in any normative system; if something is an obligation within a given normative system η, then it is permissible in η; and if something is permissible in a given normative system, then it is naturally (physically) possible:

$$\models (\mathsf{A}\varphi \rightarrow \mathsf{O}_\eta\varphi) \qquad \models (\mathsf{O}_\eta\varphi \rightarrow \mathsf{P}_\eta\varphi) \qquad \models (\mathsf{P}_\eta\varphi \rightarrow \mathsf{E}\varphi)$$

THEOREM 2.4 (Soundness and Completeness).
For every φ in the language of NTL$^-$*, we have $\models \varphi$ iff $\vdash^- \varphi$. The same holds for $\vdash$ with respect to formulas from* NTL.

PROOF. (Sketch.) Soundness is straightforward.

For completeness, consider first NTL$^-$. Let φ_0 be a consistent formula. As noted earlier, we can view NTL$^-$ as a multi-dimensional extension of CTL. Rather than extending the tableau-based method for proving the completeness of CTL in [14], we describe[3] a construction which employs the CTL completeness result directly, viewing a formula as a CTL formula for one dimension δ at a time by reading O_δ and P_δ as CTL path quantifiers A and E, respectively, and treating formulae starting with a δ'-operator ($\delta' \neq \delta$) as atomic formulae. By completeness of CTL, we get a CTL model for the formula (if it is consistent), where the states are labelled with atoms such as $\mathsf{O}_{\delta'}\Gamma$ or $\mathsf{P}_{\delta'}\Gamma$ (for $\delta' \neq \delta$). Then, for each δ' and each state, we expand the state by taking the conjunction of η'-formulae the state is labelled with, construct a (single-dimension) CTL model of that formula, and "glue" the root of the model together with the state. Repeat for all dimensions and all states.

In order to keep the formulae each state is labelled with finite, we consider only subformulae of φ_0. A δ-atom is a subformula of φ_0 starting with either P_δ or O_δ. Let $At^{-\delta}$ denote the union all of δ'-atoms for all $\delta' \neq \delta$. Furthermore, we assume that φ_0 is such that every occurrence of $\mathsf{P}_\eta(\alpha_1 \,\mathcal{U}\, \alpha_2)$ ($\mathsf{O}_\eta(\alpha_1 \,\mathcal{U}\, \alpha_2)$) is immediately preceded by $\mathsf{P}_\eta \bigcirc$ ($\mathsf{O}_\eta \bigcirc$) — we call this *XU form.* Any formula can be rewritten to XU form by recursive use of the axioms (Ax6) and (Ax7). We start with a model with a single state labelled with the literals in a consistent disjunct of φ_0 written on disjunctive normal form. We continue by expanding states labelled with formulae, one dimension δ at a time. In general, let $at(\delta, s)$ be the union of the set of δ-atoms s is labelled with and the set of negated δ-atoms of XU form s is *not* labelled

---

[3] Due to lack of space we cannot give all the technical details here. For the interested reader, more details are available at `http://home.hib.no/ansatte/tag/misc/mctl.pdf`.

with. We can now view $\bigwedge at(\delta, s)$ as a CTL formula over a language with primitive propositions $\Phi \cup At^{-\delta}$. $\bigwedge at(\delta, s)$ is NTL$^-$ consistent. The following holds: any NTL$^-$ consistent formula is satisfied by a state s' in a CTL model M' viewing $\Phi \cup At^{-\delta}$ as primitive propositions, such that for any $\delta' \neq \delta$ and any state t of M', $\bigwedge(\delta', t)$ is NTL$^-$ consistent, and s' does not have any ingoing transitions (the proof is left for the reader). This ensures that we can "glue" the pointed model M', s' to the state s while labelling the transitions in the model with the dimension δ we expanded – M', s' satisfies the formulae needed to be true there. The fact that s' does not have any ingoing transitions ensures that we can append M', s' to s without changing the truth of δ-atoms at s'. The fact that φ_0 is of XU form ensures that all labelled formulae are of XU form, which again ensures that we don't add new labels to a state when we expand it (because all the formulae we expand start with a next-modality). The fact that $\bigwedge(\delta', t)$ is consistent for states t in the expanded model, ensures that we can repeat the process. Only a finite number of repetitions are needed, depending on the number of nested operators of different dimensions in the formula, after which we can remove the non-Φ labels without affecting the truth of φ_0 and obtain a proper model.

The same construction is used for NTL, treating $\eta_\emptyset$ as any other dimension, with the following difference. When expanding a node along dimension δ, when gluing the CTL model to the expanded node label the transitions with $\eta_\emptyset$ in addition to δ. Axioms (Obl) and (Perm) ensure that this is consistent with the $\eta_\emptyset$-atoms present at the node. ■

Going beyond NTL, we can impose further structure on Σ_η and its interpretations. For example, we can extend the logical language with basic statements like $\eta \equiv \eta'$ and $\eta \sqsubset \eta'$ ($\sqsubseteq$ can then be defined), with the obvious interpretation. Furthermore, we can add unions and intersections of normative systems by requiring Σ_η to include symbols $\eta \sqcup \eta'$, $\eta \sqcap \eta'$ whenever it includes η and η', and require interpretations to interpret $\sqcup$ as set union and $\sqcap$ as set intersection. As discussed earlier, we must then further restrict interpretations such that $R \setminus (I(\eta_1) \cup I(\eta_2))$ is always total. This would give us a kind of calculus of normative systems. Let $\mathcal{K}$ be a Kripke structure and I be an interpretation with the mentioned properties:

$$\mathcal{K} \models_I \mathsf{P}_{\eta \sqcup \eta'}\varphi \rightarrow \mathsf{P}_\eta \varphi \qquad \mathcal{K} \models_I \mathsf{P}_\eta \varphi \rightarrow \mathsf{P}_{\eta \sqcap \eta'}\varphi$$
$$\mathcal{K} \models_I \mathsf{O}_\eta \varphi \rightarrow \mathsf{O}_{\eta \sqcup \eta'}\varphi \qquad \mathcal{K} \models_I \mathsf{O}_{\eta \sqcap \eta}\varphi \rightarrow \mathsf{O}_\eta \varphi$$

(these follow from Proposition 2.3). Having such a calculus allows one to reason about the composition of normative systems, similar to the way one constructs complex programs from simpler ones in Dynamic Logic [16].

Of course we could drop the reasonableness constraint. This would make it possible that "too many" norms (i.e., too many constraints on agent behaviour) may prevent *any* transition from a given state.

3. Symbolic Representations

Our aim is for NTL to be used in the formal specification and analysis of normative systems. To this end, we envisage a computer program that will take as input a Kripke structure $\mathcal{K}$, representing some system of interest, together with an NTL formula φ representing a query about this system, and some normative systems I; the program will then determine whether or not the property expressed by φ holds of $\mathcal{K}, I$, i.e., whether or not $\mathcal{K} \models_I \varphi$. Such a program is called a *model checker* [11]. However, this raises the issue of exactly how the Kripke structure $\mathcal{K}$ and normative systems I are presented to the model checker. One possibility is to simply list all the states, the propositions true in these states, and the transitions in the transition relation. Such a representation is called an *explicit state* representation. In practice, explicit state representations of Kripke structures are almost never used. This is because of the *state explosion problem*: given a system with n Boolean variables, the system will typically have 2^n states, and so an explicit representation in the input is not practicable. Instead, practical reasoning tools provide *succinct*, *symbolic* representation languages for defining Kripke structures. In this section, we present such a language for defining models, and also introduce an associated symbolic language for defining normative systems[4].

3.1. A Symbolic Language for Models

The REACTIVE MODULES LANGUAGE (RML) was introduced by Alur and Henzinger as a simple but expressive formalism for specifying game-like distributed system models [2], and this language is used as the model specification language for several model checkers [4]. In this section, we consider a "stripped down" version of RML called SIMPLE REACTIVE MODULES LANGUAGE (SRML), introduced in [17]; this language represents the core of RML,

---

[4] Notice that when we refer to a "symbolic representation", we are referring to the use of a symbolic definition of the Kripke structure *in the input to the model checker*; however, the term "symbolic model checking" is also commonly used to refer to the internal representation used by a model checker, and in this paper, we are not concerned with this issue [11].

with some "syntactic sugar" removed to keep the presentation (and semantics) simple.

Here is an example of an agent in SRML (note that agents are referred to as "modules" in SRML):

$$
\begin{array}{l}
\texttt{module}\ \mathit{toggle}\ \texttt{controls}\ x \\
\quad \texttt{init} \\
\quad \ell_1 : \top \ \leadsto\ x' := \top \\
\quad \ell_2 : \top \ \leadsto\ x' := \bot \\
\quad \texttt{update} \\
\quad \ell_3 : x \leadsto\ x' := \bot \\
\quad \ell_4 : (\neg x) \leadsto x' := \top
\end{array}
$$

This module, named *toggle*, controls a single Boolean variable, x. Occurrences of the primed version x' refer to the fresh initial value of x (in `init`) or its value in the next state (`update`). The *choices* available to the agent at any given time are defined by those `init` and `update` rules[5]. The `init` rules define the choices available to the agent with respect to the initialisation of its variables, while the `update` rules define the agent's choices subsequently. In this example, there are two `init` rules and two `update` rules. The `init` rules define two choices for the initialisation of this variable: assign it the value $\top$ (i.e., "true") or the value $\bot$ (i.e., "false"). Both of these rules can fire initially, as their conditions ($\top$) are always satisfied; in fact, only one of the available rules will ever *actually* fire, corresponding to the "choice made" by the agent on that decision round. On the left hand side of the rules are *labels* (ℓ_i) which are used to identify the rules. Note that labels do not form part of the original RML language, and in fact play no part in the semantics of RML — their role will become clear below. We assume a distinguished label "▯"; the role of this label will also become clear below. With respect to `update` rules, the first rule says that if x has the value $\top$, then the corresponding choice is to assign it the value $\bot$, while the second rules says that if x has the value $\bot$, then it can subsequently be assigned the value $\top$. In other words, the module non-deterministically chooses a value for x initially, and then on subsequent rounds toggles this value. Notice that in this example, the `init` rules of this module are non-deterministic, while the `update` rules are deterministic: SRML (and RML) allow for non-determinism in both initialisation and update rules. An SRML *system* is a set of such modules.

---

[5] To be more precise, the rules are in fact *guarded commands*.

Formally, a rule γ over a set of propositional variables Φ and a set of labels $\mathcal{L}$ is an expression

$$\ell : \varphi \rightsquigarrow v_1' := \psi_1; \ldots ; v_k' := \psi_k$$

where $\ell \in \mathcal{L}$ is a label, φ (the guard) is a propositional logic formula over Φ, each v_i is a member of Φ and ψ_i is a propositional logic formula over Φ. We require that no variable v_i appears on the l.h.s. of two assignment statements in the same rule (hence no issue on the ordering of the updates arises). The intended interpretation is that if the formula φ evaluates to true against the interpretation corresponding to the current state of the system, then the rule is *enabled* for execution; executing the statement means evaluating each ψ_i against the current state of the system, and setting the corresponding variable v_i to the truth value obtained from evaluating ψ_i. We say that $v_1, \ldots, v_k$ are the *controlled variables* of γ, and denote this set by $ctr(\gamma)$. A set of rules is said to be *disjoint* if their controlled variables are mutually disjoint.

When dealing with the SRML representation of models, a state is simply equated with a propositional valuation (i.e., the set of states of an SRML system is exactly the set of possible valuations to variables within it: $S = 2^{\Phi}$)[6]. Given a state $s \subseteq \Phi$ and a rule $\gamma : \varphi \rightsquigarrow v_1' := \psi_1; \ldots ; v_k' := \psi_k$ such that s enables γ (i.e., $s \models \varphi$) we denote the result of *executing* γ on s by $s \oplus \gamma$. For example, if $s = \{p, r\}$, and $\gamma = p \rightsquigarrow q' := \top; r' := p \wedge \neg r$, then $s \oplus \gamma = \{p, q\}$. Note that if a variable does not have its value defined explicitly by a rule that is enabled in some state, then this variable is assumed to remain unchanged.

Given a state $s \subseteq \Phi$, and set Γ of disjoint rules over Φ such that every member of Γ is enabled by s, then the interpretation s' resulting from Γ on s is denoted by $s' = s \oplus \Gamma$ (since the members of Γ are disjoint, we can pick them in any order to execute on s).

As described above, there are two classes of rules that may be declared in a module: `init` and `update`. An `init` rule is only used to initialise the

---

[6] Thus the state space of an SRML system will be exponential in the number of variables in the system. One may then wonder how this squares with our requirement earlier that we want to avoid an representation for models that is overly large. The point is that while we cannot ultimately escape the fact that the number of possible states in a system will be exponential in the number of variables, if we want to *reason* about a system, then we still need some compact way of representing the system. This is the exactly the role played by S(RML). It provides a compact language for *defining* Kripke structures, and is suitable for use as the input to a model checker. A language which was not compact in this way would be useless in practice as the input language to a model checker, since the size of the input would be unfeasibly large.

values of variables, when the system begins execution. We will assume that the guards to `init` rule are "$\top$", i.e., every `init` rule is enabled for execution in the initialisation round of the system.

An SRML *module*, m, is a triple:

$$m = \langle \mathit{ctr}, \mathit{init}, \mathit{update} \rangle \text{ where:}$$

- $\mathit{ctr} \subseteq \Phi$ is the (finite) set of variables controlled by m;
- *init* is a (finite) set of *initialisation* rules, such that for all $\gamma \in \mathit{init}$, we have $\mathit{ctr}(\gamma) \subseteq \mathit{ctr}$; and
- *update* is a (finite) set of *update* rules, such that for all $\gamma \in \mathit{update}$, we have $\mathit{ctr}(\gamma) \subseteq \mathit{ctr}$.

Note that this definition permits variables to be unitialised by the `init` rules of the module. Such variables are by default assumed to be initialised to $\bot$.

Given a module m, we denote the controlled variables of m by $\mathit{ctr}(m)$, the initialisation rules of m by $\mathit{init}(m)$, and the update rules of m by $\mathit{update}(m)$. An SRML system ρ is then an $(n+2)$-tuple

$$\rho = \langle Ag, \Phi, m_1, \ldots, m_n \rangle$$

where $Ag = \{1, \ldots, n\}$ is a set of agents, Φ is a vocabulary of propositional variables, and for each $i \in Ag$, m_i is the corresponding module defining i's choices; we require that $\{\mathit{ctr}(m_1), \ldots, \mathit{ctr}(m_n)\}$ forms a partition of Φ (i.e., every variable in Φ is controlled by some agent, and no variable is controlled by more than one agent).

A *joint* rule (j.r.) is an indexed tuple $\langle \gamma_1, \ldots, \gamma_n \rangle$ of rules, with a rule $\gamma_i \in m_i$ for each $i \in Ag$. A j.r. $\langle \gamma_1, \ldots, \gamma_n \rangle$ is enabled by a propositional valuation s iff all its members are enabled by s.

It is straightforward to extract the Kripke structure $\mathcal{K}_\rho = \langle S_\rho, S^0_\rho, R_\rho, V_\rho \rangle$ corresponding to an SRML system ρ:

- the initial states S^0_ρ correspond to the valuations that could be generated by the `init` rules of ρ against the empty valuation;
- the remaining states in S_ρ are those that could be generated by some sequence of enabled update joint rules from some initial state;
- the transition relation R_ρ is defined by $(s, s') \in R_\rho$ iff there exists some update j.r. $\langle \gamma_1, \ldots, \gamma_n \rangle$ such that this j.r. is enabled in s and $s' = s \oplus \{\gamma_1, \ldots, \gamma_n\}$.

Notice that there is nothing in this definition which requires a Kripke structure $\mathcal{K}_\rho$ corresponding to a normative system ρ to be reasonable: it is the responsibility of the modeller, defining a normative system using SRML, to ensure this.

3.2. A Symbolic Language for Normative Systems

We now introduce the SRML *Norm Language* (SNL) for representing normative systems, which corresponds to the SRML language for models. The general form of a normative system definition in SNL is as follows:

$$\begin{array}{l} \texttt{normative-system} \quad id \\ \quad \chi_1 \ \texttt{disables} \ \ell_{1_1}, \ldots, \ell_{1_k} \\ \quad \cdots \\ \quad \chi_m \ \texttt{disables} \ \ell_{m_1}, \ldots, \ell_{m_k} \end{array}$$

Here, $id \in \Sigma_\eta$ is the name of the normative system; these names will be used to refer to normative systems in formulae of NTL. The body of the normative system is defined by a set of *constraint rules.* A constraint rule

$$\chi \ \texttt{disables} \ \ell_1, \ldots, \ell_k$$

consists of a condition part χ, which is a propositional logic formula over the propositional variables Φ of the system, and a set of rule labels $\{\ell_1, \ldots, \ell_k\} \subseteq \mathcal{L}$. The intuition is that if χ is satisfied in a particular state, then any SRML rule with a label that appears on the r.h.s. of the constraint rule will be disabled in that state, according to this normative system. Consider the following simple example.

$$\begin{array}{l} \texttt{normative-system} \quad forceTrue \\ \quad \top \ \texttt{disables} \ \ell_3 \end{array}$$

We here define a normative system $forceTrue$, which consists of a single rule. The condition part of the rule is $\top$, and hence always fires; in this case, the effect is to disable the rule with label ℓ_3. Since the condition part of this rule is always enabled, in the $forceTrue$ normative system, rule ℓ_3 can never fire.

Example 3.1 (Example 2.1 continued). *The following* SRML *modules describe the Kripke model $\mathcal{K}$ from Example 2.1.*

```
module $wtrain$ controls $wAway, wWaiting, wTunnel$
  init
  $[] : \top \rightsquigarrow wAway' := \top; wWaiting' := \bot; wTunnel := \bot$
  update
  $Wwait : wAway \vee wWaiting \rightsquigarrow wWaiting' := \top; wAway' := \bot$
  $Wstayaway : wAway \rightsquigarrow wAway' := \top$
  $Wenter : wWaiting \rightsquigarrow wWaiting' := \bot; wTunnel' := \top$
  $Wleave : wTunnel \rightsquigarrow wTunnel' := \bot; wAway := \top$
```

The module for the western train controls the variables describing its position. The four update rules correspond to the physical actions available. The module etrain for the eastern train is defined in the same way, with rules named Eenter and so on.

$\texttt{module}\ controller\ \texttt{controls}\ wGreen, eGreen$
$\quad\texttt{init}$
$\quad [] : \top \rightsquigarrow wGreen' := \bot; eGreen := \bot$
$\quad\texttt{update}$
$\quad RR : \top \rightsquigarrow wGreen' := \bot; eGreen' := \bot$
$\quad RG : \top \rightsquigarrow wGreen' := \bot; eGreen' := \top$
$\quad GR : \top \rightsquigarrow wGreen' := \top; eGreen' := \bot$
$\quad GG : \top \rightsquigarrow wGreen' := \top; eGreen' := \top$

The controller module controls the variables describing the lights. For every update the module chooses one of the rules corresponding to the four possible light settings.

The following SNL *specifications describe the normative systems* η_1 *and* η_2*.*

$\texttt{normative-system}\quad \eta_1$
$\quad (\neg wWaiting \vee eWaiting)\ \texttt{disables}\ GR, GG$
$\quad (\neg eWaiting \vee wWaiting)\ \texttt{disables}\ RG, GG$

$\texttt{normative-system}\quad \eta_2$
$\quad \neg wGreen\ \texttt{disables}\ Wenter$
$\quad \neg eGreen\ \texttt{disables}\ Eenter$

Formal Definition of SNL

Formally, an SNL constraint rule is a pair

$$c = \langle \varphi, L \rangle$$

where φ is a propositional formula over Φ, and $L \subseteq \mathcal{L}$ is a set of rule labels. An SNL normative system is then a pair

$$\eta = \langle id, C \rangle$$

where $id \in \Sigma_\eta$ is a unique identifier for the normative system and C is a set of SRML constraint rules. In any given state s, the set of SRML rules that are disabled by a normative system $\langle id, C \rangle$ will be the set of rules whose labels appear on the right hand side of constraint rules in C whose condition part is satisfied in s.

Given SNL normative systems η_1 and η_2, for some SRML system ρ, we say: η_1 is *at least as liberal* as η_2 in system ρ if for every state $s \in S_\rho$, every rule that is enabled according to η_2 is enabled according to η_1; and they are *equivalent* if for every state $s \in S_\rho$, the set of rules enabled according to η_1 and η_2 are the same.

THEOREM 3.2. *The problem of testing whether* SNL *normative system* η_1 *is at least as liberal as* SNL *normative system* η_2 *is* PSPACE*-complete, as is the problem of testing equivalence of such systems.*

PROOF. We do the proof for checking equivalence; the liberality case is similar. For membership of PSPACE, consider the complement problem: guess a state s, check that $s \in S_\rho$, (reachability of states in RML is in PSPACE [2]) and check that there is some rule enabled in s according to η_2 is not enabled in s according to η_1, or vice versa. Hence the complement problem is in NPSPACE, and so the problem is in PSPACE. For PSPACE-hardness, we reduce the problem of propositional invariant checking over (S)RML modules [2]. Given an SRML system ρ and propositional formula φ, define normative systems η_1 and η_2 as follows (where ℓ does not occur in ρ):

`normative-system` η_1	`normative-system` η_2
$\neg\varphi$ `disables` ℓ	$\bot$ `disables` ℓ

According to η_2, ℓ is always enabled; thus η_1 will be equivalent to η_2 iff φ holds across all reachable states of the system. ■

4. Model Checking

The model checking problem is an important computational problem for any logic, since model checking is perhaps the most successful approach to the automated verification of logical properties of systems [11]. We consider two versions of the model checking problem for NTL, depending on whether the model is presented explicitly or symbolically. For each of these cases, there are two further possibilities, depending on whether we are given an interpretation I for normative systems named in formulae or not. If we are given an interpretation for the normative systems named in the formula, then NTL model checking essentially amounts to a conventional model checking problem: showing that, under the given interpretation, the model and associated normative systems have certain properties. However, the *uninterpreted* model checking problem corresponds to the *synthesis* of normative systems: we ask whether *there exist* normative systems that would have the desired properties. Thus the uninterpreted model checking problems combine model checking with a satisfiability checking aspect.

4.1. Explicit State Model Checking

The *interpreted explicit state model checking problem* for NTL is as follows.

> Given a Kripke structure $\mathcal{K} = \langle S, S^0, R, V \rangle$, interpretation $I : \Sigma_\eta \rightarrow N(R)$ and formula φ of NTL, is it the case that $\mathcal{K} \models_I \varphi$?

It is known that the model checking problem for CTL may be solved in time $O(|\mathcal{K}| \cdot |\varphi|)$ [14], and is in fact P-complete [22]. The standard dynamic programming algorithm for CTL model checking may be trivially adapted for interpreted explicit state NTL model checking, and may be seen to have the same time complexity. More interesting perhaps is the case where we are *not* given an interpretation. The *uninterpreted explicit state model checking problem* for NTL is as follows.

> Given a Kripke structure $\mathcal{K} = \langle S, S^0, R, V \rangle$ and formula φ of NTL, does there exist an interpretation $I : \Sigma_\eta \rightarrow N(R)$ such that $\mathcal{K} \models_I \varphi$?

Notice that uninterpreted model checking has a very natural application, as follows. We have a Kripke structure $\mathcal{K}$ and want a normative system η that will ensure some property, so we write an NTL formula φ, which refers to η, describing this property (the property might, for example, be $\mathsf{O}_\eta \Box \neg fail$); the uninterpreted model checking problem then corresponds to the *feasibility* problem described in [18]: it asks whether there in fact exist a normative system that has the desired properties. We can show:

THEOREM 4.1. *The uninterpreted explicit state model checking problem for* NTL *is* NP*-complete.*

PROOF. For membership in NP, simply guess an interpretation I and verify that $\mathcal{K} \models_I \varphi$. Since interpretations are polynomial in the size of the Kripke structure and formula, guessing can be done in (nondeterministic) polynomial time, and checking is the interpreted explicit state model checking problem. Hence the problem is in NP. For NP-hardness, we reduce SAT. Given a SAT instance $\varphi(x_1, \ldots, x_k)$, we create an instance of the uninterpreted explicit state model checking problem as follows. For each propositional variable x_i in the SAT instance, we create two variables $t(x_i)$ and $f(x_i)$, and we define a Kripke structure with $3k + 1$ states, as illustrated in Figure 3; state s_0 is the initial state, and state s_{3k+1} is a final state, with the only transition possible from this state being back to itself. Now, given the input SAT instance $\varphi(x_1, \ldots, x_k)$, we denote by $\varphi^*(x_1, \ldots, x_k)$ the NTL formula obtained by systematically replacing every propositional variable x_i with $\mathsf{P}_\eta \Diamond t(x_i)$. Finally, we define the formula to be model checked as the

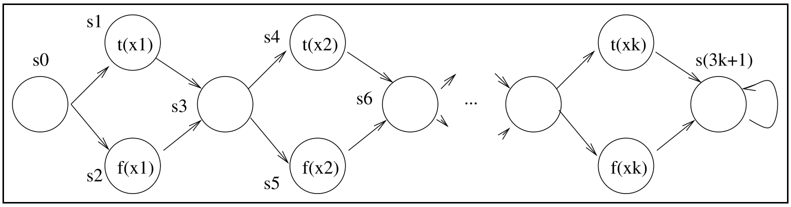


Figure 3. Reduction for Theorem 4.1.

conjunction of the following formulae.

$$\varphi^*(x_1, \dots, x_k) \tag{I}$$

$$\bigwedge_{1 \le i \le k} (\mathsf{P}_\eta \Diamond t(x_i) \rightarrow \neg \mathsf{P}_\eta \Diamond f(x_i)) \tag{II}$$

$$\bigwedge_{1 \le i \le k} (\mathsf{P}_\eta \Diamond f(x_i) \rightarrow \neg \mathsf{P}_\eta \Diamond t(x_i)) \tag{III}$$

$$\bigwedge_{1 \le i \le k} (\mathsf{P}_\eta \Diamond (t(x_i) \vee f(x_i))) \tag{IV}$$

If this formula is satisfied in the structure by some interpretation, then the interpretation for η must give a satisfying valuation for $\varphi(x_1, \dots, x_k)$; conversely, if $\varphi(x_1, \dots, x_k)$ is satisfiable, then any satisfying assignment defines an interpretation for η that makes the formula true in the structure. ■

4.2. Symbolic Model Checking

As we noted above, explicit state model checking problems are perhaps of limited interest, since such representations are exponentially large in the number of propositional variables. Thus we now consider the SRML *model checking problem for* NTL. Again, we have two versions, depending on whether we are given an interpretation or not. The interpreted version is as follows:

> Given an SRML system ρ, a set of SNL normative systems $I = \{\eta_1, \dots, \eta_k\}$ acting as an interpretation, and an NTL formula φ in which the only normative systems named are defined in I, is it the case that $\mathcal{K}_\rho \models_I \varphi$?

THEOREM 4.2. *The interpreted* SRML *model checking problem for* NTL *is* PSPACE-*complete.*

PROOF. PSPACE-hardness is by a reduction from the problem of propositional invariant verification for SRML, which is proved PSPACE-complete in [1][7]. Given a propositional formula φ and an (S)RML system ρ, let $I = \{\eta_\emptyset\}$, and simply check whether $\rho \models_I \mathsf{O}_{\eta_\emptyset} \Box \varphi$. Membership of PSPACE is by adapting the CTL symbolic model checking algorithm of Cheng [10]. ■

The *uninterpreted* SRML model checking problem for NTL is defined exactly as expected:

> Given an SRML system ρ and an NTL formula φ, does there exist a set of SNL normative systems $I = \{\eta_1, \ldots, \ldots, \eta_k\}$, one for each η named in φ, such that $\mathcal{K}_\rho \models_I \varphi$?

This problem is provably worse (under standard complexity theoretic assumptions) than the interpreted version.

THEOREM 4.3. *The uninterpreted* SRML *model checking problem for* NTL *is* EXPTIME-*hard.*

PROOF. We prove EXPTIME-hardness by reduction from the problem of determining whether a given player has a winning strategy in the two-player game PEEK-G_4 [29, p.158]. An instance of PEEK-G_4 is a quad:

$$\langle X_1, X_2, X_3, \varphi \rangle$$

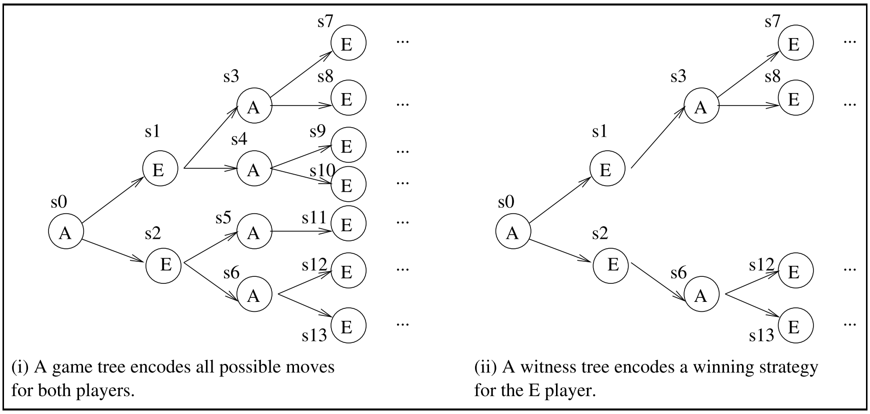
where:

- X_1 and X_2 are disjoint, finite sets of Boolean variables, with the intended interpretation that the variables in X_1 are under the control of agent 1, and X_2 are under the control of agent 2;
- $X_3 \subseteq (X_1 \cup X_2)$ are the variables deemed to be true in the initial state of the game; and
- φ is a propositional logic formula over the variables $X_1 \cup X_2$, representing the winning condition.

The game is played in a series of rounds, with the agents $i \in \{1, 2\}$ alternating (with agent 1 moving first) to select a value (true or false) for one of their variables in X_i, with the game starting from the initial assignment of truth values defined by X_3. Variables that were not changed retain the same truth value in the subsequent round. An agent wins in a given round if it makes a move such that the resulting truth assignment defined by that round makes the winning formula φ true. The decision problem associated with PEEK-G_4 involves determining whether agent 2 has a winning strategy in a given game instance $\langle X_1, X_2, X_3, \varphi \rangle$. Notice that PEEK-$G_4$ only requires "memoryless"

---

[7] In fact, the result of [2] is for RML in general, but the proof does not rely on any features of RML that are not present in SRML.



(i) A game tree encodes all possible moves for both players.

(ii) A witness tree encodes a winning strategy for the E player.

Figure 4. Game trees and witness trees.

(Markovian) strategies: whether or not an agent i can win depends only on the current truth assignment, the distribution of variables, the winning formula, and whose turn it is currently. As a corollary, if agent i can force a win, then it can force a win in $O(2^{|X_1 \cup X_2|})$ moves.

The idea of the proof is as follows. We can understand the possible plays of a finite two player game of perfect information as a tree (see Figure 4(i)), where nodes correspond to configurations of the game, and are choice points for the two players A (universal) and E (existential). Thus in an A node the universal player makes a choice, while in an E node the existential player makes a choice. We are interested in whether the E player has a winning strategy in such a game. If this is the case, then there will be a witness to this in the form of a sub-tree of the game tree, which characterises all plays of a winning strategy for E. This witness tree will be a sub-tree of the game tree with the following characteristics (see Figure 4):

- The starting position of the game is present in the witness tree.
- At every A node, all outgoing arcs of the game tree from this node must be present in the witness tree. (The E player strategy must win against all possible A moves.)
- At every E node, there can be only one outgoing node. (The E player's strategy can select only one move in any given state.)
- Every play in the witness tree must correspond to a win for the E player — that is, every possible infinite path through the witness tree from the starting position must contain a node in which the E player wins.

The idea of the reduction is to define an SRML system such that the computations of this system correspond to the plays of the PEEK-G_4 instance, and then define a NTL formula referring to a single normative system η, such that η will encode a witness tree for player 2.

Formally, given an instance $\langle X_1, X_2, X_3, \varphi\rangle$ of PEEK-G_4, we produce an instance of SRML model checking as follows. For each Boolean variable $x \in (X_1 \cup X_2)$, we create a variable with the same name in our SRML model, and we create an additional Boolean variable *turn*, with the intended interpretation that if $turn = \top$, then it is agent 1's turn to move, while if $turn = \bot$, then it is agent 2's turn to move. We have a module *move*, the purpose of which is to control *turn*, toggling its value in each successive round, starting from the initial case of it being agent 1's move.

```
module $move$ controls $turn$
  init
  [] $\top \rightsquigarrow turn' := \top$
  update
  [] $turn \rightsquigarrow turn' := \bot$
  [] $(\neg turn) \rightsquigarrow turn' := \top$
```

For each of the two PEEK-G_4 players $i \in \{1, 2\}$, we create an SRML module ag_i that `controls` the variables X_i. The module ag_i is as follows. It begins by deterministically initialising the values of all its variables to the values defined by X_3 (that is, if variable $x \in X_i$ appears in X_3 then this variable is initialised to $\top$, otherwise it is initialised to $\bot$). Subsequently, when it is this player's turn, it can non-deterministically choose at most one of the variables under its control and toggle the value of this variable; when it is not this player's turn, it has no choice but to do nothing, leaving the value of all its variables unchanged. The general structure of ag_1 is thus as follows, where $X_1 = \{x_1, \ldots, x_k\}$.

```
module $ag_1$ controls $x_1, \ldots, x_k$
  init
  // initialise to values from $X_3$
  [] $\top \rightsquigarrow x_1' := \ldots; x_k := \ldots$
  update
  $\ell_{1_1} : turn \rightsquigarrow x_1' := \bot$
  $\ell_{1_2} : turn \rightsquigarrow x_1' := \top$
  ...
  $\ell_{1_{2k}} : turn \rightsquigarrow x_k' := \bot$
  $\ell_{1_{2k+1}} : turn \rightsquigarrow x_k' := \top$
  $\ell_{1_{2k+2}} : \top \rightsquigarrow$ skip
```

Notice that an agent can always `skip`, electing to leave its variables unchanged; and, if it is not this agent's turn to move, this is the *only* choice it has. Agent ag_2 has a similar structure.

We now define the formula to model check. First, we define $chng(x_i)$ to mean that variable i changes value in some transition according to η:

$$chng(x_i) \equiv ((x_i \wedge \mathsf{P}_\eta \bigcirc \neg x_i) \vee (\neg x_i \wedge \mathsf{P}_\eta \bigcirc x_i))$$

Agent 2 is an existential player: if it is agent 2's turn, then at most one of its possible moves is allowed in the witness tree[8].

$$\mathsf{O}_\eta \square (\sum_{x_i \in X_2} chng(x_i) \leq 1)$$

If the E player changes the value of one of its variables, then this change is implemented in all its next states.

$$\bigwedge_{x_i \in X_2} \begin{array}{l} \mathsf{O}_\eta \square (\neg turn) \rightarrow \\ (chng(x_i) \rightarrow ((\mathsf{P}_\eta \bigcirc x_i \leftrightarrow \mathsf{O}_\eta \bigcirc x_i) \wedge (\mathsf{P}_\eta \bigcirc \neg x_i \leftrightarrow \mathsf{O}_\eta \bigcirc \neg x_i)))] \end{array}$$

If the E player leaves the value of one of its variables unchanged in one next state, then it is unchanged in all its next states.

$$\bigwedge_{x_i \in X_2} \begin{array}{l} \mathsf{O}_\eta \square (\neg turn) \rightarrow \\ (\neg chng(x_i) \rightarrow ((x_i \leftrightarrow \mathsf{O}_\eta \bigcirc x_i) \wedge (\neg x_i \leftrightarrow \mathsf{O}_\eta \bigcirc \neg x_i)))] \end{array}$$

Agent 1 is a universal player: all of its possible moves must be in the witness tree.

$$\mathsf{O}_\eta \square turn \rightarrow \left[\bigwedge_{x_i \in X_1} chng(x_i) \right]$$

Finally, the runs that remain must represent wins for agent 2:

$$\mathsf{O}_\eta (\neg \varphi) \mathcal{U} (\varphi \wedge turn)$$

Conjoining these formulae gives the formula to model check. We claim that this formula passes iff agent 2 has a winning strategy. For suppose that the formula passes. Then η defines a witness tree for agent 2. That it corresponds to a well-defined strategy follows from the other properties: for example, agent 2 is only allowed to make one choice in any given state when it is it's turn, which must win against all choices of agent 1. Similarly, if agent 2 has a winning strategy, then this strategy corresponds to a normative system η such that the formula passes under this interpretation. ■

---

[8] We use the $\sum$ notation here as an abbreviation for the obvious propositional equivalent.

5. Case Study: Traffic Control

We use a simple case study to illustrate some of the concepts we have introduced. The basis of the case study is as follows:

> Consider a circular road with two parallel lanes. Vehicles circulate on the two lanes clockwise. We consider three types of vehicles: cars, taxis, and ambulances. Each of the lanes is discretised into m positions, each position possibly occupied by a vehicle. In what follows, lane 1 stands for the outer lane, while lane 0 stands for the inner lane. We will refer to lane 0 as the right lane and to lane 1 as the left lane considering the direction of the vehicles. At each time step, cars and taxis can either stand still or change their position by one unit ahead, possibly changing lane at the same time. For instance a car could go from position 5 on the left lane to either position 6 on the right lane or position 6 on the left lane — or it could choose to stand still. Ambulances can stand still or change their position by one or two units, either straight or changing lanes at will.

To avoid crashes and make it possible for ambulances to get to hospitals faster, and to give taxis priority over private cars, we can imagine a number of norms that regulate the behaviour of the vehicles:

η_1 : Ambulances have priority over all other vehicles. By this we mean, in more detail, that other vehicles should stop whenever there is an ambulance behind them.

η_2 : Cars cannot use the rightmost (priority) lane.

η_3 : Vehicles have "right" priority. By this we mean that a vehicle should stop if there is another vehicle to its right. This is, of course, a very strict rule for prioritarisation which we adopt for simplicity. Otherwise, in order for a car to give way to another car with right priority, a signalling system should be used.

η_4 : Ambulances give "priority" to ambulances ahead. By an ambulance giving priority to ambulances ahead, we mean that the ambulance slows down (and therefore can only change its position by at most one unit) when there is another ambulance one unit right in front of it. Thus, norm η_4 is intended to avoid ambulances crashing when they are close and take 2-unit moves.

These norms act on the decisions that agents can make by constraining them. For instance, η_1 will force cars to stop in order to allow ambulances to overtake them.

Now, our goal in the remainder of this section is to show how the technical tools developed in this article can be used to analyse this scenario. A full scale "implementation" (involving *real* taxis and ambulances, etc) is beyond our current resources. However, what we can do instead is to take the key features of the scenario, as described above, and model them in SRML and SNL, abstracting away from lower level implementation details. We can then use model checking tools to investigate the properties of the scenario. Note that we cannot be certain that the results we obtain truly reflect reality; however, the models of systems and normative systems that we develop can, we believe, usefully inform subsequent development, and can help to identify potential issues with normative systems at an early stage of design.

Vehicle Modules

We model each vehicle as a module containing the rules that determine their physically legal movements. We define two types of modules, one for each of the two types of vehicles: those with 2-unit speed and those with 1-unit speed. Cars and taxis are vehicles of 1-unit speed, while ambulances are vehicles of 2-unit speed.

Assume that there are v vehicles named $1, \ldots, v$, each of which is either a car, a taxi or an ambulance. It is assumed that there are more positions than vehicles ($m > v$). It might be the case that none of the vehicles are cars, and the same for taxis and ambulances. Let $cars = \{c_0, \ldots, c_q\} \subseteq \{1, \ldots, v\}$ ($q \geq 0$) be the (names of) the cars. Similarly, $taxis = \{t_0, \ldots, t_r\}$ and $ambu = \{a_0, \ldots, a_s\}$. We first describe the Boolean variables. For each vehicle $1 \leq i \leq v$, position $1 \leq pos \leq m$ and $lane \in \{0, 1\}$, we have a Boolean variable $vpos_i(lane, pos)$. For example, $vpos_2(1, 7)$ means that vehicle number 2 is in position 7 on the left lane.

For each car $i \in cars$ we define a module *car-i* as in Figure 5 (for simplicity, the set $\{v'_1 := \psi_1, \ldots, v'_k := \psi_k\}$ is used as an abbreviation for the sequence of assignment operations $v'_1 := \psi_1; \ldots; v'_k := \psi_k$.)

Each car module controls the variables describing its own position. The initial position of car i is at position i along one of the tracks; the track is chosen non-deterministically (in what follows the initial positions of the cars are completely arbitrary, this particular choice was made as a simple way to ensure that two cars don't occupy the same position). In the update phase, a car module can perform one of four actions. First, the car can stand still, in which case there are no changes to the controlled variables (in this case the right hand side of the $still_i$ rule is just a dummy expression indicating no change). Second, the car can move straight ahead. In all

```
module $car\text{-}i$ controls $vpos_i(0,1), \ldots, vpos_i(0,m), vpos_i(1,1), \ldots, vpos_i(1,m)$
  init
  ▯ : $\top \rightsquigarrow vpos_i(0,i)' := \top; vpos_i(1,i)' := \bot; \{vpos_i(x,y)' := \bot \mid x \in \{0,1\}, y \neq i\}$
  ▯ : $\top \rightsquigarrow vpos_i(1,i)' := \top; vpos_i(0,i)' := \bot; \{vpos_i(x,y)' := \bot \mid x \in \{0,1\}, y \neq i\}$
  update
  $still_i : \top \rightsquigarrow vpos_i(0,1)' := vpos_i(0,1)$
  $straight_i : \bigvee_{x \in \{0,1\}, 1 \leq j \leq m} \Big( vpos_i(x,j) \wedge \bigwedge_{k \neq i} (\neg vpos_k(x, (j+1) mod\ m)) \Big) \rightsquigarrow$
    $vpos_i(0,1)' := vpos_i(0,m); vpos_i(1,1)' := vpos_i(1,m);$
    $vpos_i(0,2)' := vpos_i(0,1); vpos_i(1,2)' := vpos_i(1,1);$
      $\vdots$
    $vpos_i(0,m)' := vpos_i(0,m-1); vpos_i(1,m)' := vpos_i(1,m-1)$
  $right_i : \bigvee_{1 \leq j \leq m} \Big( vpos_i(1,j) \wedge \bigwedge_{k \neq i} (\neg vpos_k(0, (j+1) mod\ m)) \Big) \rightsquigarrow$
    $vpos_i(0,1)' := vpos_i(1,m); vpos_i(1,1)' := \bot;$
    $vpos_i(0,2)' := vpos_i(1,1); vpos_i(1,2)' := \bot;$
      $\vdots$
    $vpos_i(0,m)' := vpos_i(1,m-1); vpos_i(1,m)' := \bot$
  $left_i : \bigvee_{1 \leq j \leq m} \Big( vpos_i(0,j) \wedge \bigwedge_{k \neq i} (\neg vpos_k(1, (j+1) mod\ m)) \Big) \rightsquigarrow$
    $vpos_i(1,1)' := vpos_i(0,m); vpos_i(0,1)' := \bot;$
    $vpos_i(1,2)' := vpos_i(0,1); vpos_i(0,2)' := \bot;$
      $\vdots$
    $vpos_i(1,m)' := vpos_i(0,m-1); vpos_i(0,m)' := \bot$
```

Figure 5. Cars in the ambulance scenario.

of the three rules which move the car it is assumed that a car will only move to a position which is currently not occupied. This is a reasonable safety assumption about behaviour of cars, however it is neither sufficient (two cars might move simultaneously to the same position) nor necessary (the car currently occupying the position might move to another position at the same time) to avoid crashes. Alternatively, this assumption could have been implemented as a separate normative system. Thus, the guard of the $straight_i$ rule ensures that the position immediately in front of car i is currently available. The right hand side of this rule updates the position of car i by setting $vpos_i(x, y+1)$ to true if $vpos_i(x, y)$ were true before the rule was executed, and so on. The guard of the $right_i$ rule checks that the car is in the left lane and that one position ahead in the right lane is available, and the r.h.s. updates the position. Similarly for $left_i$. Note that the operations on vehicles' positions are modulo-m operations, where m is the number of positions in the road.

We similarly define a module *taxi-i* for each taxi $i \in taxis$ — see Figure 6.

Ambulances follow the same schema except that they have two-step rules which (also) can only be executed when the road is clear. We define a module *ambu-i* for each ambulance $i \in ambu$ — see Figure 7.

```
module $taxi\text{-}i$ controls $vpos_i(0,1),\ldots,vpos_i(0,m),vpos_i(1,1),\ldots,vpos_i(1,m)$
  init
    (as for car-i)
  update
    (as for car-i)
```

Figure 6. Taxis in the ambulance scenario.

```
module $ambu\text{-}i$ controls $vpos_i(0,1),\ldots,vpos_i(0,m),vpos_i(1,1),\ldots,vpos_i(1,m)$
  init
    (as for car-i)
  update
  $still_i$ : (as for car-i)
  $straight_i$ : (as for car-i)
  $right_i$ : (as for car-i)
  $left_i$ : (as for car-i)
  $straightstraight_i : \bigvee_{x\in\{0,1\},1\leq j\leq m}$
    $\Big(vpos_i(x,j)\wedge\bigwedge_{k\neq i}(\neg vpos_k(x,(j+1)mod\ m)\wedge\neg vpos_k(x,(j+2)mod\ m))\Big)\leadsto$
    $vpos_i(0,1)' := vpos_i(0,m-1); vpos_i(1,1)' := vpos_i(1,m-1);$
    $vpos_i(0,2)' := vpos_i(0,m); vpos_i(1,2)' := vpos_i(1,m);$
      $\vdots$
    $vpos_i(0,m)' := vpos_i(0,m-2); vpos_i(1,m)' := vpos_i(1,m-2)$
  $straightright_i : \bigvee_{1\leq j\leq m}$
    $\Big(vpos_i(1,j)\wedge\bigwedge_{k\neq i}(\neg vpos_k(1,(j+1)mod\ m)\wedge\neg vpos_k(0,(j+2)mod\ m))\Big)\leadsto$
    $vpos_i(0,1)' := vpos_i(1,m-1); vpos_i(1,1)' := \bot;$
    $vpos_i(0,2)' := vpos_i(1,m); vpos_i(1,2)' := \bot;$
      $\vdots$
    $vpos_i(0,m)' := vpos_i(1,m-2); vpos_i(1,m)' := \bot$
  $rightstraight_i : \bigvee_{1\leq j\leq m}$
    $\Big(vpos_i(1,j)\wedge\bigwedge_{k\neq i}(\neg vpos_k(0,(j+1)mod\ m)\wedge\neg vpos_k(0,(j+2)mod\ m))\Big)\leadsto$
    $vpos_i(0,1)' := vpos_i(1,m-1); vpos_i(1,1)' := \bot;$
    $vpos_i(0,2)' := vpos_i(1,m); vpos_i(1,2)' := \bot;$
      $\vdots$
    $vpos_i(0,m)' := vpos_i(1,m-2); vpos_i(1,m)' := \bot$
  $straightleft_i : \bigvee_{1\leq j\leq m}$
    $\Big(vpos_i(0,j)\wedge\bigwedge_{k\neq i}(\neg vpos_k(0,(j+1)mod\ m)\wedge\neg vpos_k(1,(j+2)mod\ m))\Big)\leadsto$
    $vpos_i(1,1)' := vpos_i(0,m-1); vpos_i(0,1)' := \bot;$
    $vpos_i(1,2)' := vpos_i(0,m); vpos_i(0,2)' := \bot;$
      $\vdots$
    $vpos_i(1,m)' := vpos_i(0,m-2); vpos_i(0,m)' := \bot$
  $leftstraight_i : \bigvee_{1\leq j\leq m}$
    $\Big(vpos_i(0,j)\wedge\bigwedge_{k\neq i}(\neg vpos_k(1,(j+1)mod\ m)\wedge\neg vpos_k(1,(j+2)mod\ m))\Big)\leadsto$
    $vpos_i(1,1)' := vpos_i(0,m-1); vpos_i(0,1)' := \bot;$
    $vpos_i(1,2)' := vpos_i(0,m); vpos_i(0,2)' := \bot;$
      $\vdots$
    $vpos_i(1,m)' := vpos_i(0,m-2); vpos_i(0,m)' := \bot$
```

Figure 7. Ambulances in the ambulance scenario.

In addition to the rules car and taxi modules have, ambulances have rules for moving two units at a time. $straightstraight_i$ moves two positions directly ahead, and the guard checks that both positions immediately ahead are unoccupied. $straightright_i$ means moving to the position which is two steps ahead in the right lane; given that the ambulance is currently in the left lane and the final position in the right lane is unoccupied. The difference between $straightright_i$ and $rightstraight_i$ is that the former also requires the position immediately ahead to be available, corresponding to driving straight and then turning right, and instead the latter also requires the position one step to the right to be available, corresponding to first turning right. Similarly for $straightleft_i$ and $leftstraight_i$.

Thus, the SRML description of the model consists of a collection of these modules. Note that the description of the modules here abstracts away from the number of vehicles in general as well as the number of the particular vehicle types. In reality, for a given number of vehicles, there are equally many modules. However, all the car (taxi, ambulance) modules are defined in the same way. For example, if there are four vehicles $\{1, 2, 3, 4\}$ and vehicles 1 and 2 are cars, vehicle 3 is a taxi, and vehicle 4 is an ambulance, then the SRML description consists of the modules named *car*-1, *car*-2, *taxi*-3 and *ambu*-4.

Normative Systems

We now go on to use SNL to describe the norms discussed above in the form of four separate normative systems (see Figure 8). Normative system η_1 has a constraint rule for each car $\{c_0, \ldots, c_q\}$ and each taxi $\{t_0, \ldots, t_r\}$, disabling any movement if they are immediately in front of an ambulance; η_2 has two constraint rules for each car, disabling switching to the right lane and moving ahead if already in the right lane; η_3 has one constraint rule for each vehicle, disabling any movement in the case that there is another vehicle immediately to the right; finally, η_4 has one constraint rule for each ambulance, disabling two-step moves in the case that there is another ambulance one step ahead in either lane.

Model Checking

For a fixed number of vehicles v, cars $cars$, taxis $taxis$ and ambulances $ambu$, let $\rho(v, cars, taxis, ambu)$ denote the SRML system defined above. As noted earlier, $\rho(v, cars, taxis, ambu)$ induces a Kripke structure $\mathcal{K}_{\rho(v,cars,taxis,ambu)}$. We can now model check formulae containing references to η_1, η_2, etc., by `normative-system` η_1

$\bigvee_{x\in\{0,1\},1\leq y\leq m,a\in ambu} vpos_{c_0}(x,(y+1)mod\ m) \wedge vpos_a(x,y)$
`disables` $straight_{c_0}, right_{c_0}$

$\vdots$

$\bigvee_{x\in\{0,1\},1\leq y\leq m,a\in ambu} vpos_{c_q}(x,(y+1)mod\ m) \wedge vpos_a(x,y)$
`disables` $straight_{c_q}, right_{c_q}$
$\bigvee_{x\in\{0,1\},1\leq y\leq m,a\in ambu} vpos_{t_0}(x,(y+1)mod\ m) \wedge vpos_a(x,y)$
`disables` $straight_{t_0}, right_{t_0}$

$\vdots$

$\bigvee_{x\in\{0,1\},1\leq y\leq m,a\in ambu} vpos_{t_r}(x,(y+1)mod\ m) \wedge vpos_a(x,y)$
`disables` $straight_{t_r}, right_{t_r}$

`normative-system` η_2

$\bigvee_{1\leq y\leq m} vpos_{c_0}(0,y)$ `disables` $straight_{c_0}$
$\top$ `disables` $right_{c_0}$

$\vdots$

$\bigvee_{1\leq y\leq m} vpos_{c_q}(0,y)$ `disables` $straight_{c_q}$
$\top$ `disables` $right_{c_q}$

`normative-system` η_3

$\bigvee_{1\leq j\leq v,1\leq y\leq m} (vpos_1(1,y) \wedge vpos_j(0,y))$ `disables` $straight_1, right_1$

$\vdots$

$\bigvee_{1\leq j\leq v,1\leq y\leq m} (vpos_v(1,y) \wedge vpos_j(0,y))$ `disables` $straight_v, right_v$

`normative-system` η_4

$\bigvee_{x,x'\in\{0,1\},1\leq y\leq m,a\in ambu}(vpos_{a_0}(x,y) \wedge vpos_a(x',(y+1)mod\ m))$
`disables` $straightstraight_{a_0}, straightright_{a_0}, rightstraight_{a_0}, straightleft_{a_0}, leftstraight_{a_0}$

$\vdots$

$\bigvee_{x,x'\in\{0,1\},1\leq y\leq m,a\in ambu}(vpos_{a_s}(x,y) \wedge vpos_a(x',(y+1)mod\ m))$
`disables` $straightstraight_{a_s}, straightright_{a_s}, rightstraight_{a_s}, straightleft_{a_s}, leftstraight_{a_s}$

Figure 8. Normative systems η_1 to η_4.

using the SNL normative systems above as interpretations. Furthermore, we let η_{12} denote the normative system obtained by taking the union of η_1 and η_2, and so on. Let I denote the interpretation of all these normative systems.

Of primary interest is, of course, whether or not there will be crashes in the system under various circumstances. We define:

$$crash = \bigvee_{i\neq j,x\in\{0,1\},k} vpos_i(x,k) \wedge vpos_j(x,k)$$

We have the following.

(P1) Without norms, there might be crashes:

$$\mathcal{K}_{\rho(v,cars,taxis,ambu)} \models_I \mathsf{P}_{\eta_\emptyset} \Diamond crash$$

In other words, the unrestricted system permits a run along which a crash happens.

(P2) The combination of the normative systems η_1, η_2, η_3 and η_4 always ensures that there are no crashes:

$$\mathcal{K}_{\rho(v,cars,taxis,ambu)} \models_I \mathsf{O}_{\eta_{1234}} \Box\neg crash$$

In other words, it is an obligatory property of a run that a crash never happens.

The two properties above hold no matter how many vehicles of each type there are. The following are examples of normative properties which hold for particular configurations of vehicles:

(P3) If there are no ambulances ($ambu = \emptyset$), then η_3 ensures that there are no crashes:

$$\mathcal{K}_{\rho(v,cars,taxis,ambu)} \models_I \mathsf{O}_{\eta_3} \Box\neg crash$$

(P4) If there is only one ambulance ($ambu = \{a_0\}$), then the combination of the normative systems η_1, η_2 and η_3 ensures that there are no crashes:

$$\mathcal{K}_{\rho(v,cars,taxis,ambu)} \models_I \mathsf{O}_{\eta_{123}} \Box\neg crash$$

(P5) If there is more than one ambulance, then the combination of the normative systems η_1, η_2 and η_3 is not enough to ensure that there are no crashes:

$$\mathcal{K}_{\rho(v,cars,taxis,ambu)} \models_I \mathsf{P}_{\eta_{123}} \Diamond crash$$

A Note on Analysis

The properties above can be checked by manual inspection: this is a technically straightforward, but rather tedious process. Instead, we have analysed this case study using the MOCHA model checker [4]. MOCHA implements model checking for CTL and ATL [3] against models specified using the Reactive Systems language, of which SRML is a subset. Of course, we cannot directly check NTL and SNL properties in this way. Instead, to realise the effect of normative systems, we manually "implement" them by modifying the conditions of relevant SRML rules; this allows us to represent a (strict) subset of NTL properties as CTL formulae.

Figure 9 summarises some test results with different scenarios. A '0' under a norm means that the norm is not applied, and a '1' that it is. A crash is possible when a '1' appears under the *Crash* column, corresponding to the formula $\mathsf{P}_\eta \Diamond crash$ being true, where η is the combination of normative systems under consideration. These test results are of course in accordance with the results presented above: property **(P1)** can be observed in rows 1, 3 and 10. Line 17 is an instance of property **(P2)**. Line 2 is an instance of property **(P3)**. Line 9 is an instance of property **(P4)**. Property **(P5)** can be observed on lines 10–16.

	#Ambulances	**#Taxis**	**#Cars**	η_1	η_2	η_3	η_4	**Crash**
1	0	0	2	0	0	0	0	1
2	0	0	2	0	0	1	0	0
3	1	0	1	0	0	0	0	1
4	1	0	1	0	0	1	0	1
5	1	0	1	0	1	1	0	1
6	1	0	1	1	0	0	0	1
7	1	0	1	1	0	1	0	1
8	1	0	1	1	1	0	0	1
9	1	0	1	1	1	1	0	0
10	2	1	1	0	0	0	0	1
11	2	1	1	0	0	1	0	1
12	2	1	1	0	1	1	0	1
13	2	1	1	1	0	0	0	1
14	2	1	1	1	0	1	0	1
15	2	1	1	1	1	0	0	1
16	2	1	1	1	1	1	0	1
17	2	1	1	1	1	1	1	0

Figure 9. Testing $\eta_1, \eta_2, \eta_3, \eta_4$ with different numbers of ambulances, taxis and cars.

6. Discussion

6.1. Related Work

The work presented in this paper has its roots in several different communities, the most significant being the tradition of using deontic logic in computer science to reason about normative behaviour of systems [30, 12, 19], and the use of model checking and temporal logics such as CTL to analyse the temporal properties of systems [14, 11].

The two main differences between the language of NTL and the language of conventional deontic logic (henceforth "deontic logic") are, first, *contextual* deontic operators allowing a formula to refer to several different normative systems, and, second, the use of *temporal* operators. All deontic expressions in NTL refer to time: $\mathsf{P}_\eta \bigcirc \varphi$ ("it is permissible in the context of η that φ is true at the next time point"); $\mathsf{O}_\eta \Box \varphi$ ("it is obligatory in the context of η that φ always will be true"); etc. Conventional deontic logic contains no notion of time. In order to compare our temporal deontic statements with those of deontic logic we must take the temporal dimension to be implicit in the latter. Two of the perhaps most natural ways of doing that is to take "obligatory" ($\mathsf{O}\varphi$) to mean "*always* obligatory" ($\mathsf{O}_\eta \Box \varphi$), or "obligatory at the *next point in time*" ($\mathsf{O}_\eta \bigcirc \varphi$), respectively, and similarly for permission. In either case, all the principles of *Standard Deontic Logic* (SDL) (see, e.g., [9]) hold also for NTL, viz., $\mathsf{O}(\varphi \to \psi) \to (\mathsf{O}\varphi \to \mathsf{O}\psi)$ (K); $\neg\mathsf{O}\bot$ (D):[9] and from φ infer $\mathsf{O}\varphi$ (N). The two mentioned temporal interpretations of the (crucial) deontic axiom D are (both NTL validities):

$$\neg\mathsf{O}_\eta \Box \bot \text{ and } \neg\mathsf{O}_\eta \bigcirc \bot$$

A more detailed understanding of the relationship between NTL and other deontic logics would also be useful. Observe that our language is in one sense rather restricted: every deontic attitude is towards the future, never about the present or past. Indeed, when reasoning about normative behaviour of a system, it is also not easy to see what an obligation towards an objective, purely propositional formula, actually *means*. Our framework focuses on ideal *transitions*, rather than ideal *states*. This choice of design makes it also not easy to compare our set-up with other temporal deontic logics. We cannot, for instance, express properties like $\mathsf{O}_\eta \bigcirc \varphi \to \bigcirc \mathsf{O}_\eta \varphi$ (for a discussion of such "perfect recall"-like properties for temporal deontic logic, see [8]). There is another interesting direction to relax our class of formulae, however: namely, to allow for arbitrary linear temporal logic formulas $T\varphi$ in the scope of an obligation O_η.[10] This would allow us for instance to express that in system η, property φ needs to be true within three steps: $\mathsf{O}_\eta(\bigcirc \varphi \vee \bigcirc \bigcirc \varphi \vee \bigcirc \bigcirc \bigcirc \varphi)$. Semantically, this would also pave the way to define a norm as a restriction on *runs*, rather than *transitions*. One can for instance think of a norm that forbids any run in which some "unwanted" transition occurs more than n times. As a special case this would facilitate to enforce *fairness* and *liveness* conditions by a norm [15].

---

[9] Actually, the scheme D is $\mathsf{O}\varphi \to \neg\mathsf{O}\neg\varphi$, but for the logics that we consider here, both representations are equivalent.

[10] The '$\eta_\emptyset$ fragment' of this language would still be a strict subset of CTL*, so the overall language might still be well-behaved.

Contrary-to-Duty obligations are structures involving two obligations, where the second obligation "takes over" when the first is violated [21]. Logics that deal with this kind of obligation typical add actions, time, a default component or a notion of context (signalling that the primary obligation has been violated, and we are now entering a sub-ideal context) to their semantic machinery to deal with them [21]. NTL is already equipped with a temporal component, and it would certainly also be possible to label the transitions in our semantics with actions. However, given that we incorporate a suite of norms within one system, it seems NTL can also represent "sub-ideal" contexts. Technically, η_2 represents the secondary obligations that come into force when η_1 is violated, if the domain $dom(\eta_2) = \{t \mid \exists u\ (t, u) \in \eta_2\}$ is exactly the range $ran(\eta_1) = \{t \mid \exists s\ (s, t) \in \eta_1\}$ of η_1. We leave a detailed comparison between existing temporal deontic logics and NTL for future works, as well as any investigation into the usefulness of NTL to model contrary-to-duty obligations.

It has been argued that "*deadlines* are important norms in most interactions between agents" [13, page 40], and this naturally suggests the need for a temporal component in reasoning about systems with norms. Indeed, the authors of [13] used CTL in their paper *Designing a Deontic Logic of Deadlines* [7], and one of their authors reduces Strategic Deontic Temporal Logic to ATL in [6][11].

One of our concerns in this paper was to give a *computationally grounded* semantics for deontic modalities, in that we aim to give the semantics a clear computational interpretation; in this respect, our work is similar in spirit to the *deontic interpreted systems* model of Lomuscio and Sergot [19]. Perhaps the most obvious difference is that while we consider "bad transitions", Lomuscio and Sergot are concerned with "bad states".

We should also mention work by Sergot and Craven on the use of variants of the C+ language for representing and reasoning about deontic systems [23, 24]. The nC+ language they develop can be understood as an alternative to SRML/SNL for defining (symbolic) representations of Kripke models and normative systems. The main difference is that the C+ language provides a richer, higher level, arguably more general, logical framework for specifying models than SRML/SNL. In fact their work emerges from a rather different community — reasoning about action and non-monotonic reasoning in artificial intelligence. It would be interesting to undertake a more formal investigation of the relationship between the two frameworks, with respect to both expressive power and computational complexity. It seems

---

[11] The latter paper includes an application to Chisholm's paradox.

plausible that analogues of our SRML/SNL model checking problems will be more complex in the richer framework of nC+, although we emphasise that currently we have no results here. Note that the first of the papers cited above also presents an outline of how normative properties expressed in CTL can be evaluated using standard model checking tools, and uses an example similar to that used in this paper.

Finally, NTL is closely related to a system called Normative ATL, which was introduced in [31]. In fact, NTL is related to CTL [14] in the same way as that Normative ATL is related to ATL [3]: however, NTL (and specifically its semantics) is much simpler (and we believe more intuitive) than Normative ATL [31], and we present many more technical results associated with our logic.

6.2. Future Work

A number of issues suggest themselves for future work:

- Regarding NTL, tight bounds for complexity of uninterpreted symbolic model checking, and the complexity of satisfiability, which has not been addressed within this paper.
- The calculus of normative systems, as mentioned in Section 2, could be developed further. In this paper, we have considered only set-theoretic operations on normative systems, (taking their union and intersection), but other possible operators might be considered as well, such as what happens when a normative system is "restricted" to some set of players, or when it is restricted to those constraints that some players regard as in their best interests. To capture these latter concerns, we would need some notion of preference or goals.
- Another issue of interest is that of "reasonableness", and in particular the extent to which this constraint is necessary.
- We might usefully consider the possibility and implications of *non- compliance*. It seems inevitable that in real systems, some agents will fail to comply with a normative system, even if it is in their best interests to do so. This raises two issues: First, what is the right way to go about dealing with this possibility with respect to the design of normative systems themselves, and second, how are we to deal with these concerns at the language level?
- We might also consider prioritised collections of normative systems ("if this normative system fails, then use this").

- Finally, of course, a full implementation of a model checker encompassing the variations discussed in Section 4 would be desirable, and as ever, more detailed case studies would be useful to further evaluate the logic.

6.3. Conclusions

The design and application of normative systems and social laws is a major area of research activity in the multi-agent systems community. If we are going to make use of such social laws, then it seems only appropriate that we develop formalisms that allow us to explicitly and directly reason about them. In this paper, we have developed a Normative Temporal Logic that is intended for this purpose, being careful to give a semantics to deontic modalities that has a clear computational interpretation. We see the key advantages of NTL as follows. First, the fact that the formalism is so closely related to CTL is likely to be an advantage from the point of view of comprehension and acceptance within the mainstream model checking/verification community. Second, the fact that the language has a clear computational interpretation means that it can be applied in a computational setting without any ambiguity of interpretation. Third, the clear identification of different normative systems within the language, and the ability to talk about these directly, represents a novel step forward. While NTL arguably lacks some of the nuances of more conventional deontic and deontic temporal logics, we believe these advantages imply that the language and the approach it embodies merit further research.

Acknowledgements. We would like to thank the reviewers for their helpful comments and suggestions. This work was partially funded by projects AT (CONSOLIDER CSD2007-0022), IEA (TIN2006-15662-C02-01) and 2006 5 OI 099.

References

[1] Alur, R., Henzinger, T., *Computer-aided verification*, 1999. Manuscript.

[2] Alur, R., Henzinger, T.A., 'Reactive modules', *Formal Methods in System Design*, 15(11): 7–48, 1999.

[3] Alur, R., Henzinger, T.A., Kupferman, O., 'Alternating-time temporal logic', *Journal of the ACM*, 49(5): 672–713, 2002.

[4] Alur, R., Henzinger, T.A., Mang, F.Y.C., Qadeer, S., Rajamani, S.K., Taşiran, S., 'Mocha: Modularity in model checking', in *CAV 1998: Tenth International Conference on Computer-aided Verification*, LNCS, vol. 1427, Springer-Verlag, Berlin, Germany, 1998, pp. 521–525.

[5] Attie, Paul C., Allen Emerson, E., 'Synthesis of concurrent systems with many similar processes', *ACM Transactions on Programming Languages and Systems*, 20(1): 51–115, 1998.

[6] Broersen, J., 'Strategic deontic temporal logic as a reduction to ATL, with an application to Chisholm's scenario', in *Proceedings Eighth International Workshop on Deontic Logic in Computer Science (DEON'06)*, LNAI, vol. 4048, Springer, Berlin, 2006, pp. 53–68.

[7] Broersen, J., Dignum, F., Dignum, V., Meyer, J.-J. Ch., 'Designing a deontic logic of deadlines', in Lomuscio, A., Nute, D. (eds.), *Proceedings Seventh International Workshop on Deontic Logic in Computer Science (DEON'04)*, LNAI, vol. 3065, Springer, Berlin, 2004, pp. 43–56.

[8] Brunel, J., *Combining temporal and deontic logics. with an application to computer security*, Ph.D. thesis, IRIT, Toulouse, France, 2007.

[9] Carmo, Jose, Jones, Andrew J.I., 'Deontic logic and contrary-to-duties', in Gabbay, D.M., Guenthner, F. (eds.), *Handbook of Philosophical Logic, 2nd edition*, vol. 8, Kluwer Academic Publishers, Dordrecht, 2002, pp. 265–343.

[10] Cheng, A., '*Complexity results for model checking*', Tech. Rep. RS-95-18, BRICS, Department of Computer Science, University of Aarhus, 1995.

[11] Clarke, E.M., Grumberg, O., Peled, D.A., *Model Checking*, The MIT Press, Cambridge, MA, 2000.

[12] Dignum, F., 'Autonomous agents with norms', *Artificial Intelligence and Law*, 7: 69–79, 1999.

[13] Dignum, F., Broersen, J., Dignum, V., Meyer, J.-J. Ch., 'Meeting the deadline: Whey, when and how', in Hinchey, M.G., Rash, J.L., Truszkowski, W.F., Rouff C.A. (eds.), *Formal Approaches to Agent-Based Systems*, LNAI, vol. 3228, Springer, Berlin, 2004, pp. 30–40.

[14] Emerson, E.A., 'Temporal and modal logic', in van Leeuwen J. (ed.), *Handbook of Theoretical Computer Science Volume B: Formal Models and Semantics*, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1990, pp. 996–1072.

[15] Francez, N., *Fairness*, Springer-Verlag, Berlin, Germany, 1986.

[16] Harel, D., Kozen, D., Tiuryn, J., *Dynamic Logic*, The MIT Press, Cambridge, MA, 2000.

[17] van der Hoek, W., Lomuscio, A., Wooldridge, M., 'On the complexity of practical ATL model checking', in *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, Hakodate, Japan, 2005, pp. 201–208.

[18] van der Hoek, W., Roberts, M., Wooldridge, M., 'Social laws in alternating time: Effectiveness, feasibility, and synthesis', *Synthese*, 156(1): 1–19, 2007.

[19] Lomuscio, A., Sergot, M., 'Deontic interpreted systems', *Studia Logica*, 75(1): 63–92, 2003.

[20] Moses, Y., Tennenholtz, M., 'Artificial social systems', *Computers and AI*, 14(6): 533–562, 1995.

[21] Prakken, H., Sergot, M., 'Contrary-to-duty obligations', *Studia Logica* 57(1): 91–115, 1996.

[22] Schnoebelen, P., 'The complexity of temporal logic model checking', in Balbiani, P., Suzuki, N.-Y., Wolter, F., Zakharyascev, M. (eds.), *Advances in Modal Logic, vol. 4*, King's College Publications, London, 2003, pp. 393–436.

[23] Sergot, M., 'Modelling unreliable and untrustworthy agent behaviour', in Dunin-Keplicz, B., Jankowski, A., Skowron, A., Szczuka, M. (eds.), *Monitoring, Security, and Rescue Techniques in Multiagent Systems*, *Advances in Soft Computing*, Springer, Berlin, 2005, pp. 161–178.

[24] Sergot, M., Craven, R., 'The deontic component of action language nC+', in Goble, L., Meyer, J.-J. Ch. (eds.), *Deontic Logic and Artificial Systems*, *Proc. 8th International Workshop on Deontic Logic in Computer Science*, LNAI, 4048, Springer, Berlin, 2006, pp. 222–237.

[25] Shoham, Y., Tennenholtz, M., 'Emergent conventions in multi-agent systems', in Rich, C., Swartout, W., Nebel, B. (eds.), *Proceedings of Knowledge Representation and Reasoning (KR&R-92)*, 1992, pp. 225–231.

[26] Shoham, Y., Tennenholtz, M., 'On the synthesis of useful social laws for artificial agent societies', in *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, San Diego, CA, 1992, pp. 276–281.

[27] Shoham, Y., Tennenholtz, M., 'On social laws for artificial agent societies: Off-line design', in Agre, P.E., Rosenschein, S.J. (eds.), *Computational Theories of Interaction and Agency*, The MIT Press, Cambridge, MA, 1996, pp. 597–618.

[28] Shoham, Y., Tennenholtz, M., 'On the emergence of social conventions: Modelling, analysis, and simulations', *Artificial Intelligence*, 94(1–2): 139–166, 1997.

[29] Stockmeyer, L.J., Chandra, A.K., 'Provably difficult combinatorial games', *SIAM Journal of Computing*, 8(2): 151–174, 1979.

[30] Wieringa, R.J., Meyer, J.-J.Ch., 'Deontic logic in computer science', in Meyer, J.-J.Ch., Wieringa, R.J. (eds.), *Deontic Logic in Computer Science—Normative System Specification*, John Wiley & Sons, New York, 1993, pp. 17–40.

[31] Wooldridge, M., van der Hoek, W., 'On obligations and normative ability: Towards a logical analysis of the social contract', *Journal of Applied Logic*, 4(3–4): 396–420, 2005.

Thomas Ågotnes
Bergen University College
PO Box 7030
N-5020 Bergen
Norway
`tag@hib.no`

Juan A. Rodríguez-Aguilar
& Carles Sierra
Artificial Intelligence Research Institute — IIIA
Spanish Council for Scientific Research — CSIC
Barcelona, Spain
`{jar,sierra}@iiia.csic.es`

Wiebe van der Hoek
& Michael Wooldridge
Department of Computer Science
University of Liverpool
Liverpool
L69 3BX United Kingdom
`{wiebe,mjw}@csc.liv.ac.uk`

Melvin Fitting

Reasoning with Justifications

Abstract. This is an expository paper in which the basic ideas of a family of *Justification Logics* are presented. Justification Logics evolved from a logic called LP, introduced by Sergei Artemov [1, 3], which formed the central part of a project to provide an arithmetic semantics for propositional intuitionistic logic. The project was successful, but there was a considerable bonus: LP came to be understood as a logic of knowledge with explicit justifications and, as such, was capable of addressing in a natural way long-standing problems of logical omniscience. Since then, LP has become one member of a family of related logics, all logics of knowledge with explicit knowledge terms. In this paper the original problem of intuitionistic foundations is discussed only briefly. We concentrate entirely on issues of reasoning about knowledge.

Keywords: logic of knowledge, justification logic, modal logic.

1. Introduction

This is an expository paper in which the basic ideas of a family of *Justification Logics* are presented. Justification Logics evolved from a logic called LP, introduced by Sergei Artemov [1, 3], which formed the central part of a project to provide an arithmetic semantics for propositional intuitionistic logic. The project was successful, but there was a considerable bonus: LP came to be understood as a logic of knowledge with explicit justifications and, as such, was capable of addressing in a natural way long-standing problems of logical omniscience, [7]. Since then, LP has become one member of a family of related logics, all logics of knowledge with explicit knowledge terms. In this paper the original problem of intuitionistic foundations is discussed only briefly. We concentrate entirely on issues of reasoning about knowledge.

2. Hintikka's Logics of Knowledge

In [21] Hintikka developed an approach to logics of knowledge that has become the basis for much that followed. While the central ideas are generally familiar, a sketch of them will be useful. A logic with multiple agents is the

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

natural one but for the time being we will confine things to a single agent, and discuss widening the setting towards the end of the paper.

A propositional modal logic is constructed. It is customary to denote the necessity operator by K, standing for *it is known that*. We take $\supset$ and $\bot$ as basic, with other connectives defined in the usual way. Then a minimal logic of knowledge can be formulated as follows.

Axiom Schemes

K1 All instances of classical tautologies

K2 $K(X \supset Y) \supset (KX \supset KY)$

K3 $KX \supset X$

Rules of Inference

Modus Ponens $\dfrac{X \quad X \supset Y}{Y}$

Necessitation $\dfrac{X}{KX}$

Axiom **K3** can be seen as capturing part of the classic characterization of knowledge as justified, true belief: it says that what is known must be true. Without such an axiom we are capturing belief, not knowledge. Axiom **K2** is familiar from normal modal logics, but is somewhat problematic here. It says knowledge is closed under modus ponens—briefly, we know the consequences of what we know. This will be discussed further below. The **Necessitation** rule is also familiar from normal modal logics, and is also problematic here. It says all logical truths are known, and it too will be discussed further below.

These minimal axioms are generally extended with one or both of the following.

Axiom Schemes

Positive Introspection $KX \supset KKX$

Negative Introspection $\neg KX \supset K\neg KX$

The first says that if we know something, we know we know it. The second says if we don't know something, we know we don't know it. While these are increasingly strong, and increasingly doubtful assumptions, adding them both seems to have been the most common approach in the literature.

Historically, the justified knowledge approach sketched in this paper began with an analog of Hintikka's axioms including the one for Positive Introspection but not Negative Introspection. It was straightforward to provide an analog for the system without either Introspection axiom, and more recently an analog of the system with both Introspection axioms has appeared. To keep things relatively simple, we will follow the historical development here, and assume only Positive Introspection.

The semantics Hintikka introduced is a possible world one. A model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \Vdash \rangle$ consists of a collection $\mathcal{G}$ of *states of knowledge*, an accessibility relation $\mathcal{R}$ on them that is reflexive and transitive (since we have positive introspection), and a notion of truth at a state, which we write as $\mathcal{M}, \Gamma \Vdash X$, where $\mathcal{M}$ is a model, Γ is a state, and X is a formula. On propositional connectives $\Vdash$ is truth-functional at each world, and the usual Kripke condition is met,

$$\mathcal{M}, \Gamma \Vdash KX \Longleftrightarrow \mathcal{M}, \Delta \Vdash X \text{ for all } \Delta \in \mathcal{G} \text{ with } \Gamma \mathcal{R} \Delta \tag{1}$$

where this is usually read informally as: the agent knows X at state Γ if X is the case at all states the agent cannot distinguish from Γ. For a single agent this logic of knowledge is simply the well-known modal logic S4—things become more complex when multiple agents are involved.

Hintikka's approach has been successfully applied to many well-known puzzles and problems, but it is not the end of the matter. What is it that an agent has knowledge of, sentences or propositions? Hintikka's logic is quite unproblematic when taken to be a logic of propositions—in it, if $X \equiv Y$ is provable, so is $KX \equiv KY$. But two sentences might be equivalent and so express the same proposition, while that equivalence is not at all easy to see—we may not be aware of the equivalence. What we communicate directly is sentences, and propositions only indirectly. Wittgenstein argued that all mathematics is, essentially, a single tautology. In this sense, if we know the proposition that $2 + 2 = 4$, we know all mathematics—it's just a single proposition, the truth. But here the distinction between sentences and propositions is fundamental—mathematicians work with sentences directly, and propositions quite indirectly. Thought of as a logic of sentences, Hintikka's approach suffers from a fundamental difficulty usually referred to as *logical omniscience*. An agent turns out to know too much. This problem really breaks into two separate pieces, which we now discuss.

The first omniscience problem arises from the **Necessitation Rule**. According to this, an agent must know all tautologies. But a tautology could have as many symbols as there are atoms in the universe, and it is unlikely an agent actually would know the truth of such a formula. The second comes from axiom **K2**. It follows from this scheme that an agent would know the consequences of what it knows. This, too, seems unlikely in practice.

The usual sentence-based solution is to say that we are not really dealing with a logic of knowledge, but a logic of *potential* knowledge. KX informally means that X is *knowable*, rather than actually known. This has its negative uses—if $\neg KX$ is established then X is not knowable, so it is certainly not known, whatever we might mean by that. But still, a true logic of knowledge for sentences, not just knowability, would be a nice thing to have.

3. Awareness Logic

One of the reasons we might not know something that is knowable is that we haven't thought about it. In [11], Fagin and Halpern give this simple idea a formal treatment, producing a family of *awareness logics*. In these there is an explicit representation of the things one has thought about, so to speak.

Semantically, an awareness model is $\langle \mathcal{G}, \mathcal{R}, \mathcal{A}, \Vdash \rangle$, where $\mathcal{G}$, $\mathcal{R}$, and $\Vdash$ are as before, and $\mathcal{A}$ is a mapping assigning to each member of $\mathcal{G}$ a set of formulas. The members of $\mathcal{A}(\Gamma)$ are the formulas we are *aware* of at Γ. No special conditions are placed on this function; in particular, $\mathcal{A}(\Gamma)$ need not be complete, or consistent, or closed under subformulas. Syntactically, an operator A is added to the language, and AX is taken to be true at a state Γ just in case $X \in \mathcal{A}(\Gamma)$.

With this machinery added, 'actual' knowledge can be represented as a conjunction $KX \wedge AX$. We explicitly know those formulas that are knowable, and that we have thought about. This now allows us to avoid logical omniscience problems—we easily have models in which formulas that might be problems are not because we are not aware of them.

The authors of [11] consider various natural conditions one might place on $\mathcal{A}$ such as closure under subformulas, or preservation on passing to accessible states (monotonicity). Likewise one might want to say one is only aware of formulas that are not too complicated, or formulas whose possible justifications are not too long. The framework is, in fact, very general. It is more of a toolbox, able to contain many things useful or not, than a tool in itself.

4. Explicit Justifications

Now we start on the main subject matter of this survey—logics with explicit justification terms. Instead of KX, that is, "X is known," consider $t{:}X$, that is, "X is known for the explicit reason t." Of course these explicit reasons, or justification terms, should have some internal structure. We introduce the basic machinery.

4.1. Syntax and Axiom System

A justification for $X \supset Y$ applied to a justification for X should produce a justification for Y. The symbol $\cdot$, called *application*, is used and the basic principle is this.

1. $s{:}(X \supset Y) \supset (t{:}X \supset (s \cdot t){:}Y)$

Adding extra (perhaps useless) material to a justification still gives a justification, though a weaker one. The symbol + is used for this.

2. $s{:}X \supset (s + t){:}X$
3. $t{:}X \supset (s + t){:}X$

A justification can be verified for correctness. It has it's own justification. The symbol ! is used here. This is referred to as *checking* or *verification*.

4. $t{:}X \supset !t{:}t{:}X$

A logical truth has a justification and no further analysis is needed. Constant symbols are used for this, that is, if X is a 'basic' truth we can conclude $c{:}X$, where c undergoes no further analysis. Of course, c can be assigned a weight of some kind, reflecting the complexity of X, but this will not be done here.

In addition to logical truths, there are also facts of the world. These are, in a sense, inputs from outside the structure. They are represented by variables, thus $x{:}X$.

Gathering this together, we have the following formal language. First we have *justifications*, or *terms*, sometimes called *proof polynomials* when used in a mathematical setting.

Variables: x, y, z, ..., are justifications.

Constants: c, d, e, ..., are justifications.

Application: If s and t are justifications, so is $(s \cdot t)$.

Weakening: If s and t are justifications, so is $(s + t)$.

Checking: If t is a justification, so is $!t$.

Next, the definition of formulas.

Propositional Letters: P, Q, ..., are formulas.

Falsehood: $\bot$ is a formula.

Implication: If X and Y are formulas, so is $(X \supset Y)$.

Justification Formulas: If X is a formula and t is a justification then $t{:}X$ is a formula.

Axioms were, mostly, given above. Here they are in full. All formulas of the following form are axioms.

J0 A sufficient set of classical tautologies

J1 $s{:}(X \supset Y) \supset (t{:}X \supset (s \cdot t){:}Y)$

J2 $s{:}X \supset (s + t){:}X$

J3 $t{:}X \supset (s + t){:}X$

J4 $t{:}X \supset !t{:}t{:}X$

J5 $t{:}X \supset X$

Finally the rules of inference.

Modus Ponens: From X and $X \supset Y$ infer Y.

Axiom Necessitation: If X is an axiom, infer $c{:}X$, where c is a constant symbol.

The logic just described is called LP, standing for *logic of proofs*, [1, 3]. The name comes from the fact that it was originally created to represent arithmetic proofs, which are certainly justifications of a very special kind. Variations on LP will be discussed in Section 9.

4.2. Semantics

The standard epistemic semantics for LP comes from [13], and amounts to a blending of an earlier semantics from [22] with the usual Hintikka style semantics for logics of knowledge. One can see it as being in the tradition of Awareness Logics, but with the awareness function supplied with an additional structure of justifications. A model is $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \Vdash \rangle$, where $\mathcal{G}$ and $\mathcal{R}$ are as usual, with $\mathcal{R}$ reflexive and transitive, and with $\Vdash$ behaving on propositional connectives in the usual way. The new item is $\mathcal{E}$, which is an *evidence* function. The idea is, $\mathcal{E}$ assigns to each possible world Γ and to each justification t a set of formulas—those formulas that t is relevant to, or that t can serve as possible evidence for, at Γ. Evidence functions must meet certain conditions.

Monotonicity $\Gamma \mathcal{R} \Delta$ implies $\mathcal{E}(\Gamma, t) \subseteq \mathcal{E}(\Delta, t)$

Application $X \supset Y \in \mathcal{E}(\Gamma, s)$ and $X \in \mathcal{E}(\Gamma, t)$ imply $Y \in \mathcal{E}(\Gamma, s \cdot t)$

Weakening $\mathcal{E}(\Gamma, s) \cup \mathcal{E}(\Gamma, t) \subseteq \mathcal{E}(\Gamma, s + t)$

Checking $X \in \mathcal{E}(\Gamma, t)$ implies $t{:}X \in \mathcal{E}(\Gamma, !t)$

The one new condition on $\Vdash$ concerns the behavior of justification terms. It is the counterpart of (1) for standard logics of knowledge.

$$\begin{array}{c}\mathcal{M},\Gamma \Vdash t{:}X \iff \mathcal{M},\Delta \Vdash X \text{ for all } \Delta \in \mathcal{G} \text{ with } \Gamma\mathcal{R}\Delta \\ \text{and } X \in \mathcal{E}(\Gamma, t)\end{array} \tag{2}$$

In short, we have $t{:}X$ at Γ if X is knowable at Γ in the Hintikka sense, and t is relevant evidence for X at Γ. If we think of Hintikka semantics as capturing the idea of *true belief*, then what the present machinery captures is *justified* true belief.

There is also a stronger version of the semantics. A model $\mathcal{M}$ is said to be *fully explanatory* provided, if $\mathcal{M},\Delta \Vdash X$ for all $\Delta \in \mathcal{G}$ with $\Gamma\mathcal{R}\Delta$ then there is some justification t such that $\mathcal{M},\Gamma \Vdash t{:}X$. More informally, $\mathcal{M}$ is fully explanatory provided knowability of X at Γ (in the Hintikka approach) implies there is a justification for X at Γ. Under simple, reasonable conditions, designed to ensure that constants behave in corresponding ways semantically and proof theoretically, provability agrees with truth at all worlds of all models, and this agrees with truth at all worlds of all fully explanatory models, [13].

4.3. Awareness Logics Again

LP can be seen as an extension of awareness logic with the awareness function made more explicit. One can extract a variety of awareness logics from LP directly, as well. For instance, we might define $\#(t)$ to be the number of operation symbols in term t; then $\mathcal{A}(\Gamma) = \{X \mid X \in \mathcal{E}(\Gamma, t) \text{ and } \#(t) < n\}$ is a natural awareness function, for each choice of n—we are aware of formulas with possible justifications that are not too complicated. Or again, we might set $\mathcal{A}(\Gamma) = \{X \mid X \in \mathcal{E}(\Gamma, t) \text{ and } t \in S\}$, where S is a fixed set of justifications. Both these are quite plausible awareness functions, and others are easy to come by as well.

5. Internalization

In logics of knowledge, and more generally in normal modal logics, one has a *necessitation* rule: if X is provable, so is $\Box X$. In LP this takes a much stronger form, and is a constructively provable theorem rather than a basic rule. It is called *internalization* and is due to Artemov. It says, given a proof of X, then $t{:}X$ is provable for some closed term t, where t embodies the given proof of X. Here is a very simple example. First we have a proof in LP, of $t{:}P \supset (c \cdot t){:}(P \vee Q)$, where c is a proof constant for $P \supset (P \vee Q)$, and t is some arbitrary proof term.

1. $c{:}(P \supset (P \vee Q))$
2. $c{:}(P \supset (P \vee Q)) \supset (t{:}P \supset (c \cdot t){:}(P \vee Q))$
3. $t{:}P \supset (c \cdot t){:}(P \vee Q)$

Line 1 is by axiom necessitation, 2 is an instance of axiom scheme 1, and 3 is from 1 and 2 by modus ponens. Internalizing this proof, the following is also LP provable, where d is a proof constant for $c{:}(P \supset (P \vee Q)) \supset (t{:}P \supset (c \cdot t){:}(P \vee Q))$.

$$(d \cdot !c){:}(t{:}P \supset (c \cdot t){:}(P \vee Q))$$

Notice that $!c$ justifies line 1 of the proof above, d justifies line 2, and $(d \cdot !c)$ justifies line 3 by representing the application of modus ponens. We omit the proof of this formula in LP.

6. Information Hiding and Recovery

In LP justifications are explicit, while in Hintikka-style logics of knowledge they are hiden. The knowledge operator of a Hintikka logic is a kind of existential quantifier asserting the existence of a justification without saying what it is. Explicit justifications can easily be hidden behind such quantifier-like operators. Remarkably, explicit justifications can also be recovered. This is the content of a fundamental theorem in the subject of justification logics.

First, the easy direction. As noted above, one can think of K as a kind of existential quantifier: KX is read as "there is a reason for X". Then if X is a theorem of LP, with explicit justifications, and we replace each justification with K, we get a theorem of the Hintikka knowledge logic S4. This is easy to see. Each LP axiom scheme instance turns into an axiom of S4, and applications of LP rules of inference turn into applications of S4 rules of inference. Consequently an entire LP axiomatic proof converts into a proof in S4, and hence theorems convert as well.

A translation in the opposite direction is also possible, but much more difficult, and goes under the name *Realization Theorem.* It is due to Artemov, as is its first proof. Loosely it says, if X is a theorem of S4, there is some way of replacing occurrences of K with explicit justifications to produce a theorem of LP. But one can do better yet. In making the replacement of K symbols with justifications, negative occurrences of K can always be replaced with distinct variables, and positive occurrences with justifications that may be computed from those variables. Thus S4 theorems have a hidden input-output structure that is exposed by a realization in LP.

Here is an example whose verification is left to the reader. $(KP \vee KQ) \supset K(KP \vee KQ)$ is a theorem of S4. And here is a realization of it, provable in LP, and with negative occurrences of K replaced with distinct variables.

$$(x{:}P \vee y{:}Q) \supset (c{\cdot}!x + d{\cdot}!y){:}(x{:}P \vee y{:}Q)$$

In this, c and d are constant symbols introduced using the Axiom Necessitation rule, with c introduced for the tautology $x{:}P \supset (x{:}P \vee y{:}Q)$ and d for the tautology $y{:}Q \supset (x{:}P \vee y{:}Q)$.

Artemov's original proof of the Realization Theorem was entirely constructive, [3]. It extracted a provable realized version of an S4 theorem from a *cut-free sequent calculus* proof of the S4 theorem. Since then variations on the construction have been developed by several people, and the algorithm has become more efficient. In [13] a non-constructive proof was given, using the semantics described in Section 4.2. While not algorithmic, it goes more deeply into the role of the + operator. More recently a constructive proof along somewhat different lines, but still using a cut-free sequent calculus, was given in [15]—more about this in Section 8.

The Realization Theorem, and its associated algorithms, is central to understanding the significance of a logic like LP. It says we can reason Hintikka style and then, on demand, produce a conclusion with full justifications present. This holds great potential which is being explored by a number of researchers.

7. Original Intent

The logic LP is the first in a family of epistemic logics with explicit justifications; others will be discussed in section 9. But the original reason for its creation was quite different, and of considerable significance. It was part of a project to produce a constructive foundation for intuitionistic propositional logic. This project was successful. In this section we sketch the basic ideas.

There is a well-known BHK interpretation of the intuitionistic connectives (Brouwer, Heyting, Kolmogorov). Loosely, one thinks of intuitionistic truth as being rather like provability. This is often used informally to motivate intuitionistic logic. There have been various attempts to make the idea into a proper mathematical construct, Kleene's notion of realizability, for instance.

Gödel gave an axiomatization of the intuitive notion of provability, in [19]. He wrote *Bew* for the modal operator—we will use $\Box$ in this section. In his short paper he observed three fundamental things: 1) his axiomatization was equivalent to the Lewis logic S4; 2) intuitionistic logic embedded into it by inserting his ‘provable’ operator before each subformula; and 3) his axiomatization of provability did not correspond to the formal notion of provability in Peano arithmetic, because under an arithmetic interpretation $\Box X \supset X$ amounted to a consistency assertion. Thus the attempt was only partially successful, though tremendously influential. It was eventually realized that the logic of provability in Peano arithmetic was not S4, but GL, in which $\Box X \supset X$ is replaced with the Löb axiom, $\Box(\Box X \supset X) \supset \Box X$, but GL is not a logic into which one can embed intuitionistic logic, Gödel style.

Gödel made another suggestion, in [20], which remained largely unknown until the publication of his collected works. One might interpret S4 arithmetically, not as the logic of *provability*, but as the logic of *explicit proofs*. Sergei Artemov independently conceived the same idea, and carried it through to a successful conclusion, thus completing Gödel’s project. The chain of construction goes as follows. First, as Gödel noted, propositional intuitionistic logic embeds in S4, with intuitionistic connectives being translated as ‘provable’ versions of their classical counterparts. Second, via the Realization theorem, S4 embeds in LP. And third, LP does embed into Peano arithmetic, with LP terms mapping to Gödel numbers of explicit arithmetic proofs—this is the Artemov Arithmetic Completeness theorem. All this combines to provide the desired arithmetic semantics for intuitionistic logic.

It became clear as early as 1998 that logics with explicit proof terms could also be seen as logics of explicit justifications in a more general sense, [2, 6] for instance. The introduction of a Kripke-style semantics for LP, [13], provided a significant technical and conceptual tool for epistemic applications. Today work on understanding and applying LP and its relatives to epistemic problems proceeds at an increasing pace. But from an *epistemic* point of view, and taking various generalizations of LP into account, arithmetic completeness is not central in the way it was for the intuitionistic logic project. The Realization theorem, however, remains fundamental. This accounts for the minimal mention the arithmetic result gets here—it is important, but for something other than the subject of our immediate concern.

8. Realizations As First-Class Objects

Originally a realization was simply a tool for extracting the explicit content of an S4 theorem. More recently realizations have become objects for investigation in their own right. In [17, 16, 15] they are functions mapping *occurrences* of necessity operators to justification terms (the use of $\Box$ instead of K will be continued). Occurrences themselves are formally distinguished by breaking $\Box$ up into infinitely many copies, $\Box_1$, $\Box_2$, $\ldots$, with each having at most one occurrence in a formula. Further, positive and negative occurrences are distinguished, with even indexed operators in negative positions and odd indexed ones in positive positions. A formula with such subscripted modal operators is called *properly annotated.* Every modal formula can be properly annotated, and all properly annotated versions of the same formula are effectively interchangeable for our purposes. Then a realization is simply a function from positive integers to justification terms that maps even integers to distinct variables. If r is a realization function and X is an annotated formula, $r(X)$ is the result of replacing each subformula $\Box_i Y$ with $r(i){:}r(Y)$.

We also need the standard notion of substitution—recall that justification terms can contain variables. A substitution is a mapping, σ, from variables to justification terms. The result of applying a substitution σ throughout a formula X is denoted $X\sigma$. It is not hard to show that if X is a theorem of LP, so is $X\sigma$ for any substitution (though the role of constants may shift).

The direct use of realization functions, and of substitutions, has made it possible to state some algorithmic results concerning LP in a relatively coherent way—we do this in the rest of the section.

8.1. The Replacement Theorem

In $\mathsf{S4}$, as in every normal modal logic, one has a Replacement Theorem. If $A \equiv B$ is provable, then so is $X \equiv Y$, where Y is like X except that an occurrence of A as a subformula has been replaced with B. (Multiple replacements can be handled sequentially.) To state this more easily, we use the following notation. Suppose $X(P)$ is a formula with at most one occurrence of the propositional letter P. Then we write $X(Z)$ for the result of substituting Z for P in $X(P)$. Now the usual Replacement Theorem can be stated as follows. If $\vdash_{\mathsf{S4}} A \equiv B$ then $\vdash_{\mathsf{S4}} X(A) \equiv X(B)$.

We saw in the previous section that, in LP, positive and negative occurrences of subformulas sometimes play different roles. One problem with developing a Replacement Theorem for LP is that when $A \equiv B$ is expanded using $\wedge$, $\vee$, and $\neg$, one sees that A has both a positive and a negative occurrence. Fortunately this difficulty can be addressed, because there is a 'polarity preserving' version of Replacement. Suppose P has at most one *positive* occurrence in $X(P)$. Then if $\vdash_{\mathsf{S4}} A \supset B$ then $\vdash_{\mathsf{S4}} X(A) \supset X(B)$. Here is this result again, for purposes of comparison with the corresponding LP result given below.

$$\frac{A \supset B}{X(A) \supset X(B)} \tag{3}$$

We might have a hope, then, that a polarity preserving version of Replacement will transfer to LP. But the obvious transfer does not work, and a moment's thought suggests why. Suppose we have $\vdash_{\mathsf{LP}} A \supset B$, and P has at most one positive occurrence in $X(P)$. In $X(A)$ the subformula A might occur within the scope of a justification term, and when we replace A with B to produce $X(B)$ we should expect that justification term will need updating to incorporate its original reasoning, plus a reason accounting for the passage from A to B. That justification term itself may occur within the scope of another one, which will need updating, and so on up. In short, wherever appropriate, reasons must be modified to reflect the fact that A implies B.

With realization and substitution machinery available a proper, algorithmic, version of Replacement for LP can be given. Suppose $X(P)$, A and B are properly annotated modal formulas (not LP formulas), where P has a single positive occurrence in $X(P)$. Suppose also that r_0 is a realization function such that $\vdash_{\mathsf{LP}} r_0(A) \supset r_0(B)$. Then there is a pair, $\langle r, \sigma\rangle$ where r is a realization function and σ is a substitution, such that $\vdash_{\mathsf{LP}} r_0(X(A))\sigma \supset r(X(B))$. Schematically in LP we have the following, instead of (3).

$$\frac{r_0(A) \supset r_0(B)}{r_0(X(A))\sigma \supset r(X(B))} \tag{4}$$

In this, the substitution σ and the new realization function r take care of the 'justification adjustment' discussed above.

There are some conditions on the result above that must be stated. First, neither A nor B should share an index with $X(P)$. This is rather minor since annotations can always be reassigned in $X(P)$. Second and more serious, r_0 must be what is called *non self-referential on variables* over $X(A)$, that is, if $\Box_{2n}Z$ is a subformula of $X(P)$, and $r_0(2n)$ is the variable x, then x does not occur in $r_0(Z)$. This is especially important since it was shown in [10] that self-referential constant symbols are essential for completeness.

There are also stronger versions of the Replacement Theorem than we stated. There are, for instance, restrictions on the behavior of the substitution σ. But most importantly, the pair $\langle r, \sigma\rangle$ carries out a replacement of A by B not only in $X(P)$, but in subformulas as well. (With negative subformulas the implication in the conclusion is reversed.) Thus it is a kind of *uniform* replacement.

The proof of the LP Replacement Theorem is entirely algorithmic, with the algorithm depending on the complexity of $X(P)$.

8.2. Realization Merging

Suppose we have two different realization functions; is there some way of merging them into a single one? As with the Replacement Theorem, there is an algorithmic solution to this problem too, and the result takes the following form. If r_1 and r_2 are realization functions, and X is a properly annotated modal formula. Then there is a pair $\langle r, \sigma \rangle$, where r is a realization function, σ is a substitution, so that we have both $\vdash_{\mathsf{LP}} r_1(X)\sigma \supset r(X)$ and $\vdash_{\mathsf{LP}} r_2(X)\sigma \supset r(X)$. Actually, the full result says $\langle r, \sigma \rangle$ will merge not only X, but subformulas as well, but we can avoid the complexities in this survey paper.

Here is a simple example to show the utility of this. Suppose $A \supset C$ and $B \supset C$ are properly annotated modal formulas, and we have separate realization functions r_1 and r_2 such that $\vdash_{\mathsf{LP}} r_1(A \supset C)$ and $\vdash_{\mathsf{LP}} r_2(B \supset C)$. If we apply the algorithm referred to above, using $(A \vee B) \supset C$ for the formula X, a pair $\langle r, \sigma \rangle$ is produced, and it is not hard to show that $\vdash_{\mathsf{LP}} r((A \vee B) \supset C)$.

8.3. The Realization Theorem, Again

Both the Replacement result of section 8.1 and the Merging result of section 8.2 are special cases of a more general result which will not be stated here. Replacement, Merging, and one more consequence of the general theorem, together yield yet another algorithm for producing an LP provable realization of an S4 theorem, [15]. Like the original Artemov proof, it makes use of a cut-free S4 proof to create the provable realization. The relationship of this algorithm to the original one has not yet been determined.

8.4. What's Missing

A central open problem in this area concerns the familiar rule *modus ponens*. Suppose X and $X \supset Y$ are modal formulas, and we have LP provable realizations for both. Then there will be a provable realization for Y as well, by the following indirect argument. Since each of these has a provable realization, then X and $X \supset Y$ themselves are provable in S4. But then so is Y, and by the Realization Theorem, it will have an LP provable realization. The problem is, the only algorithmic ways currently known for producing a provable realization of Y is to begin with a cut-free proof of Y. This is expensive, and we wind up discarding any information contained in the provable realizations for X and $X \supset Y$ that we started with.

Is there a direct way of calculating a provable realization for Y, given provable realizations for X and $X \supset Y$? Nobody knows how to handle this. The various merging and replacement algorithms discussed above do not apply. All of them pay careful attention to polarity of subformulas, but notice that the occurrence of X in $X \supset Y$ is in a negative position, while in X itself it is positive. Until this has been dealt with, we do not have a fully satisfactory calculus of realization functions.

9. Generalizations

LP is a version of S4 with explicit justifications. S4 is just one single agent logic of knowledge. What about others; what about multiple agents; what about communication of justifications?

9.1. Single Agent Logics

Logics weaker than S4 were investigated early on. T works fine. One simply removes the obvious axioms from LP. Actually, the axiom necessitation rule needs modification too—it now reads: if X is an axiom, so is $c:X$ for a constant c, and now the same rule can be reapplied. This change is to compensate for the removal of the ! operator. K4 and K can be thought of as logics of belief rather than of knowledge, but again justification versions of them are straightforward to develop. For all these logics a Realization Theorem holds, [9], and completeness relative to a possible world semantics, as in section 4.2, can be shown.

In the other direction, S5 has also been given its justification version, independently in [25] and [23]. This requires the introduction of an additional operator, ?, dual to !. Whereas ! is designed for an explicit version of positive introspection, ? is intended to deal with negative introspection. Once again, Realization and semantical completeness have been established.

Several other single-knower variations have been considered, but this should be enough to give the general picture.

9.2. Multiple Agent Logics

There has already been work on justification versions of multiple agent Hintikka logics of knowledge. The idea has been to treat shared justifications as a kind of explicit common knowledge [4, 8, 5]. In fact, explicit justifications can be shown to satisfy the fixpoint condition usually imposed on common knowledge.

More recent work, still in progress, concerns multiple agent logics in which each agent has its own set of reasons, [26]. This is a natural setting for considering the results of communication, and some work has been done on the justification version of public announcements [24]. Among other things, this involves putting a syntactic counterpart of the semantic evidence function into the language in a way reminiscent of the treatment of the awareness function, discussed in section 3.

10. The Goal

What we are after is logics in which we can reason, not just about facts, but about reasons for facts, and in which we can reason about these reasons conveniently and efficiently. Work is very much in progress. We conclude with a quick summary of the current state of things.

For single agent logics of knowledge, both implicit (Hintikka style) versions and explicit (justification style) versions exist, and there is effective machinery to translate between them. How to handle *modus ponens* is still an open problem, but some progress is being made on this. In the meantime, cut-free sequent formulations serve as tools, though expensively. There are versions in which both implicit and explicit knowledge can be expressed. These are natural, and have a well-understood proof theory and semantics. What is missing for these is a version of the Realization Theorem, which now would take the form of 'self-realization.' Here the treatment of *modus ponens* seems to be even more of a central issue.

Multiple agent logics of knowledge with explicit justifications are not as well developed yet, partly because of the additional richness available. So far the most successful versions have had implicit knowledge individually, while explicit justifications were shared machinery. The two central issues currently being explored are: allowing a separate family of justifications for each agent, and permitting communication of justifications. This is perhaps the most active current area of development.

One might consider a move to a first-order logic of knowledge, implicit and explicit. One might have quantifiers over things or even quantifiers over justifications. There has been some work on this, [27, 18, 14], but the situation is complex and not yet well-understood.

Formalizing the reasoning of knowledge with justifications that are explicitly present has turned out to be a rich source of results and techniques. Already what has been achieved is significant, and progress remains steady. The range of logical systems that has resulted presents an exciting mix of expressiveness and succinctness, with strong proof-theoretic and semantic aspects. It is to be hoped that this work will lead to a better understanding of reasoning, its explanation and communication.

References

[1] Artemov, S., *Operational modal logic*, Technical Report MSI 95-29, Cornell University, December 1995.

[2] Artemov, S., *Logic of proofs: a unified semantics for modality and lambda-terms*, Technical Report CFIS 98-17, Cornell University, 1998.

[3] Artemov, S., 'Explicit provability and constructive semantics', *The Bulletin for Symbolic Logic*, 7(1): 1–36, 2001.

[4] Artemov, S., *Evidence-based common knowledge*, Technical Report TR-2004018, CUNY Ph.D. Program in Computer Science, 2004.

[5] Artemov, S., 'Justified common knowledge', *Theoretical Computer Science*, 357(1–3): 4–22, 2006.

[6] Artemov, S., Kazakov, E., Shapiro, D., *On logic of knowledge with justifications*, Technical Report CFIS 99-12, Cornell University, 1999.

[7] Artemov, S., Kuznets, R., 'Logical omniscience via proof complexity', in *Computer Science Logic 2006*, Lecture Notes in Computer Science, vol. 4207, Springer, Berlin, 2006, pp. 135–149.

[8] Artemov, S., Nogina, E., 'Introducing justification into epistemic logic', *Journal of Logic and Computation*, 15(6): 1059–1073, 2005.

[9] Brezhnev, V., 'On the logic of proofs', in Striegnitz, K. (ed.), *Proceedings of the Sixth ESSLLI Student Session*, Helsinki, 2001, pp. 35–46.

[10] Brezhnev, V., Kuznets, R., 'Making knowledge explicit: How hard it is', *Theoretical Computer Science*, 357: 23–34, 2006.

[11] Fagin, R., Halpern, J.Y., 'Beliefs, awareness and limited reasoning', *Artificial Intelligence*, 34: 39–76, 1988.

[12] Feferman, S. (ed.), *Kurt Gödel Collected Works*, Oxford, 1986–2003. Five volumes.

[13] Fitting, M.C., 'The logic of proofs, semantically', *Annals of Pure and Applied Logic*, 132: 1–25, 2005.

[14] Fitting, M.C., 'A quantified logic of evidence (short version)', in de Queiroz, R., Macintyre, A. Bittencourt, G. (eds.), *WoLLIC 2005 Proceedings*, Electronic Notes in Theoretical Computer Science, Elsevier, Amsterdam, 2005, pp. 59–70.

[15] Fitting, M.C., *Realizations and LP.* Available at http://comet.lehman.cuny.edu/fitting/, 2006.

[16] Fitting, M.C., *Realizing substitution instances of modal theorems.* Available at http://comet.lehman.cuny.edu/fitting/, 2006.

[17] Fitting, M.C., *A replacement theorem for LP*, Technical report, CUNY Ph.D. Program in Computer Science, 2006. http://www.cs.gc.cuny.edu/tr/.

[18] Fitting, M.C., 'A quantified logic of evidence', *Annals of Pure and Applied Logic*, 2007. Forthcoming.

[19] Gödel, K., 'Eine Interpretation des intuistionistischen Aussagenkalkuls', *Ergebnisse eines mathematischen Kolloquiums*, 4: 39–40, 1933. Translated as 'An interpretation of the intuitionistic propositional calculus', in [12] 296–301.

[20] Gödel, K., 'Vortrag bei Zilsel'. Translated as 'Lecture at Zilsel's', in Feferman, S. (ed.), *Kurt Gödel Collected Works*, Oxford, 1986–2003. Five volumes. III, 62–113, 1938.

[21] Hintikka, J., *Knowledge and Belief*, Cornell University Press, Ithaca, NY, 1962.

[22] Mkrtychev, A., 'Models for the logic of proofs', in Adian, S.I., Nerode, A. (eds.), *Logical Foundations of Computer Science*, Lecture Notes in Computer Science, vol. 1234, Springer, Berlin, 1997, pp. 266–275.

[23] Pacuit, E., 'A note on some explicit modal logics', in Acopoulos, C. Dimitr (ed.), *Proceedings of the Fifth Panhellenic Logic Symposium*, 2005, pp. 117–125.

[24] Renne, B., 'Bisimulation and public announcements in logics of explicit knowledge', in Artemov, S., Parikh, R. (eds.), *Proceedings of the Workshop on Rationality and Knowledge, 18th European Summer School in Logic, Language, and Information (ESSLLI)*, Málaga, Spain, 2006, pp. 112–123.

[25] Rubtsova, N., 'Evidence reconstruction of epistemic modal logic S5', in Grigoriev, D., Harrison, J., Hirsch, E.A. (eds.), *Computer Science—Theory and Applications*, Lecture Notes in Computer Science, vol. 3967, Springer-Verlag, Berlin, 2006, pp. 313–321.

[26] Yavorskaya, T. (Sidon), 'Logic of proofs with two proof predicates', in Grivoriev, D., Harrison, J., Hirsch, E. (eds.), *Computer Science—Theory and Applications*, Lecture Notes in Computer Science, vol. 3967, Springer, Berlin, 2006.

[27] Yavorsky, R., 'Provability logics with quantifiers on proofs', *Annals of Pure and Applied Logic*, 113(1–3): 373–387, 2002.

Melvin Fitting
The Graduate School & University Center, CUNY,
Department of Computer Science,
365 Fifth Avenue New York City,
NY 10016-4309, USA
`melvin.fitting@lehman.cuny.edu`

Janusz Czelakowski

Monotone Relations, Fixed Points and Recursive Definitions*

Abstract. The paper is concerned with reflexive points of relations. The significance of reflexive points in the context of indeterminate recursion principles is shown.

Keywords: fixed-point, monotone relation, chain-σ-continuous relation, definability by arithmetic recursion.

The focus of this paper is on the method of defining mathematical objects by means of fixed-points. The theory of fixed-points splits into two, to a large extent autonomous and conceptually independent, areas of research. Each of these fields is determined by the specific choice of underlying mathematical models:

(1) the theory of fixed-points developed in the setting of complete metric spaces (see e.g. Goebel and Kirk [1990], Kirk and Sims (eds.) [2001]).

(2) the theory of fixed-points carried out in order-complete partially ordered sets (see e.g. Gunter and Scott [1990]).

In this paper we are mainly concerned with the second branch of the theory of fixed-points. The basic idea of (2) is simple. At the outset one isolates a partially ordered set $(P, \leq)$, the universe of discourse, which exhibits some form of order-completeness. In the strongest case $(P, \leq)$ is assumed to be a complete lattice. But weaker versions of order-completeness are also plausible as e.g. inductivity, i.e., which means that the underlying poset is closed under the formation of suprema of chains, or continuity (alias directed completeness), i.e., the closure under the formation of supremas of directed subsets. In the weakest case, $(P, \leq)$ is assumed to have a zero and be closed under the suprema of chains of type ω.

In the standard approach to the order-oriented fixed-point theory, the focus is on mappings $\pi : P \rightarrow P$ exhibiting certain natural properties linked with the order of the poset $(P, \leq)$. The weakest natural property of π in

---

*The author is indebted to the referees for many suggestions which have improved the original exposition to its present form and for careful reading of the manuscript which have uncovered certain inaccuracies.

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

this context is that of monotonicity: for every pair $a, b \in P$, $a \leq b$ implies $\pi(a) \leq \pi(b)$. Some other properties of mappings such as e.g. expansivity or various forms of order-continuity are also important.

Any pair formed by $(P, \leq)$ together with an appropriate mapping $\pi : P \to P$ is called an *intended model*. The method of fixed-points is aimed at showing that for each such intended model, the mapping π has a least fixed-point in the sense of $\leq$, say a^*. This means that $\pi(a^*) = a^*$ and a^* is the least element b of P such that $\pi(b) = b$. From the perspective of definition theory, one *defines* an object as the least fixed-point of π because the standard conditions imposed on a correct definition, viz. the existence and uniqueness conditions are implied by the properties of π. Thus the definition of such an object as the least fixed-point is fully legitimate. It is clear however that the defining procedure based on the fixed-point method is highly impredicative. The reason is that the defined object viz. the least fixed-point of π, is not defined directly but indirectly, through an intermediate universe $(P, \leq)$.

But in this paper we are also concerned with a wider problem than that of *definability* objects; the focus is rather on proving the *existence* of objects, based on various fixed-point procedures, and not their uniqueness. There is also another significant difference — the method we want to outline concerns fixed-points (alias *reflexive points*) of *relations*, and not only functions. The underlying models we study are founded on order-complete posets $(P, \leq)$. But these posets are additionally equipped with certain binary relations, which need not be functions. We give a bunch of fixed-points theorems for relations and give some applications of these results.

The difference between the method of fixed points for mappings and that for relations in the theory of definition lies in the fact that in the case of relations we do not define certain objects unambiguously as least fixed-points, but merely prove the existence of such objects as fixed-points (of relations). We thus allow for a range of defined objects in the space P. The logical status of this method is somewhat similar to the well-known category method in mathematics.

From the perspective of formal semantics, the structures we study are Kripke frames augmented with an order relation $\leq$, where $\leq$ is assumed to be at least chain σ-complete, which means that the order has a zero and the supremum of every well-ordered chain of type ω exists.

The order oriented fixed-point theory constitutes the conceptual framework of the theory of *semantic domains*, a part of theoretical computer science. Investigations of semantic domains were initiated by Dana Scott in the end of the sixties and then developed by him and his collaborators (see e.g. Gunter and Scott [1990].) A recent paper by Desharnais and Möller [2005] surveys older results on fixed-points for relations (see e.g. Cai and Paige [1992]) and presents new results, relevant to computer science.

This work is a modest part of a more ambitious research project entitled *Infinitistic methods in the theory of definitions*, sponsored by KBN grant No. 2 H01 A 007 25. The project is centered on comparing various infinitistic methods of defining mathematical objects as e.g. transfinite recursion, the fixed-points method, the diagonal method, and the methods provided by calculus. This paper is thought as a preliminary account of fixed-points (= reflexive points) for relations.

The paper has a two-level structure. Many of the results of the paper are formulated in the *object* language of the theory of fixed-points, viz. the language of ordered Kripke frames. (The object language need not be first-order.) The object language and the fixed-point theorems proved in it constitute the first level of the research. But the work also contains results which belong to the *metatheory* of fixed-points. The second level of the research is constituted by the metalanguage of the fixed-point-theory and theorems establishing the relationship between fixed-point theorems and results belonging to other disciplines, especially to set theory. The paper contains an array of fixed-point theorems which are compared with the Axiom of Choice and its weak counterparts as well as with definability principles by arithmetic recursion. The results belonging to the metalevel are marked by the prefix *meta* to distinguish them from the object-level results.

In Chapter 2 we are also interested in the *constructive* dimension of the research which means that the focus is on the methods that guarantee attainability of the least fixed-point in ω steps in the pertinent posets. The fixed point-theory from the constructive perspective is coextensive with theoretical arithmetic, because, as it is shown below, fixed-point theorems for mappings presented in this paper are basically equivalent to various forms of *arithmetic recursion* (in the deterministic version). A more advanced theory of *inherently* infinitistic fixed-points for relations and its relationship with various forms of definability based on the Noetherian induction or transfinite recursion as well as the fixed-point approach to the well-known back and forth method will be presented in another paper.

1. Partially Ordered Sets

Let P be a set. A binary relation $\leq$ on P is an *order* (or *partial order*) on P iff $\leq$ satisfies the following conditions:

(i) $\leq$ is *reflexive*, i.e., $a \leq a$, for all $a \in A$;

(ii) $\leq$ is *transitive*, i.e., $a \leq b$ and $b \leq c$ implies $a \leq c$, for all $a, b, c \in A$;

(iii) $\leq$ is *antisymmetric*, i.e., $a \leq b$ and $b \leq a$ implies $a = b$, for all $a, b \in A$.

A *partially ordered set*, a *poset*, for short, is a set with an order defined on it. Each order relation $\leq$ on P gives rise to a relation $<$ of *strict order*: $a < b$ in P if and only if $a \leq b$ and $a \neq b$.

Let $(P, \leq)$ be a poset and let X be a subset of P. Then X inherits an order relation from P: given $xy \in X$, $x \leq y$ in X iff $x \leq y$ in P. We then also say that the order on X is *induced* by the order from P.

(1) An element $M \in X$ is called *maximal* in X whenever $M \leq x$ implies $M = x$, for every $x \in X$.

(2) An element $m \in X$ is called *minimal* in X whenever $x \leq m$ implies $m = x$, for every $x \in X$.

(3) An element $u \in P$ is called *an upper bound* of the set X if $x \leq u$, for every $x \in X$.

(4) An element $l \in P$ is called a *lower bound* of the set X if $l \leq x$, for every $x \in X$.

(5) An element $a \in P$ is called the *least upper bound* of the set X if a is an upper bound of X and $a \leq u$ for every upper bound u of X. If X has a least upper bound, this is called the *supremum* of X and is written '$\sup(X)$'.

(6) An element $b \in P$ is called the *greatest lower bound* of the set X if b is a lower bound of X and $l \leq b$ for every lower bound l of X. If X has a greatest lower bound, this is called the *infimum* of X and is written '$\inf(X)$'.

A set X may have more than one maximal element or none at all. A similar situation holds for minimal elements.

Instead of '$\sup(X)$' and '$\inf(X)$' we shall often write '$\vee X$' and '$\wedge X$'; in particular we write '$a \vee b$' and '$a \wedge b$' instead of '$\sup(\{a, b\})$' and '$\inf(\{a, b\})$'.

If the poset P itself has an upper bound u, then it is the only upper bound. u is then called the *greatest element* of P. In an analogous way the notion of the *least element* of P is defined.

A set $X \subseteq P$ is:

(a) an *upward directed* subset of P if for every pair $a, b \in X$ there exists an element $c \in X$ such that $a \leq c$ and $b \leq c$ (or, equivalently, if every finite non-empty subset of X has an upper bound in X);

(b) a *chain* in P (or: a *linear subset*) if, for every pair $a, b \in X$, either $a \leq b$ or $b \leq a$ (that is, if any two elements of X are comparable);

(c) a *well-ordered subset* of P (or: a *well-ordered chain* in P) if X is a chain in which every non-empty subset $Y \subseteq X$ has a minimal element (in Y).

Equivalently, in the presence of the Axiom of Dependent Choices (see below), X is well-ordered if and only if it is a chain and there is no strictly decreasing sequence $c_0 > c_1 > \ldots > c_n > \ldots$ of elements of X. Every well-ordered chain is isomorphic to a unique ordinal, called the *type* of X. The empty set $\emptyset$ is well-ordered and $\emptyset$ is its order type.

A downward directed subset is defined similarly; when nothing to the contrary is said, ‘ directed’ will always mean ‘ directed upwards’.

If the poset $(P, \leq)$ itself is a chain or directed, then it is simply called a chain or a directed poset.

Let $(P, \leq)$ be a poset and let X and Y be subsets of P. The set X is *cofinal with* Y if for every $a \in Y$ there exists $b \in X$ such that $a \leq b$. If X is cofinal with Y, then $\sup(Y)$ exists if and only if $\sup(X)$ exists. Furthermore $\sup(Y) = \sup(X)$.

THEOREM 1.1. *Let $(P, \leq)$ be a poset. Every countable directed subset D of $(P, \leq)$ contains a well-ordered subset of type $\leq \omega$. In particular, every countably infinite chain contains a cofinal well-ordered subchain of type ω.*

PROOF. Let a_n, $n \in \omega$, be an enumeration of the countably infinite directed set D. By Arithmetic Recursion (see Section 3) we define an increasing sequence b_n, $n \in \omega$, of elements of D. We put $b_0 := a_0$. Assume $b_0, b_1, \ldots, b_n$ have been defined so that $b_0 < b_1 < \ldots < b_n$. Then $b_{n+1} :=$ the unique element a_m of D with the least index m such that $b_n < a_m$ and $a_n < a_m$. The chain b_n, $n \in \omega$, is cofinal with D. ■

The above theorem is not true for uncountable directed subsets.

THEOREM 1.2 (Kuratowski -Zorn’s Lemma). *If every non-empty chain in a poset P has an upper bound, then the set P contains a maximal element.*

Kuratowski-Zorn’s Lemma is an equivalent form of the Axiom of Choice (on the basis of the familiar axioms of Zermelo-Fraenkel’s set theory ZF without the Axiom of Regularity).

DEFINITION 1.3. Let $(P, \leq)$ be a poset.

(a) $(P, \leq)$ is *directed-complete* if for every directed subset $D \subseteq P$, the supremum $\sup(D)$ exists in $(P, \leq)$.

(b) $(P, \leq)$ is *chain-complete* (or *inductive*) if for every chain $C \subseteq P$, the supremum $\sup(C)$ exists in $(P, \leq)$.

(c) $(P, \leq)$ is *well-orderably-complete* if for every well-ordered chain $C \subseteq P$, the supremum $\sup(C)$ exists in $(P, \leq)$. □

It is clear that every directed-complete poset is chain-complete and every inductive poset is well-orderably-complete. The above properties are thus successively weaker and weaker. It turns out however that in the presence of the Axiom of Choice they are mutually equivalent.

The empty subset of a poset is well-ordered. Hence, if $(P, \leq)$ is well orderably-complete, then the supremum of the empty subset exists and it is the least element in $(P, \leq)$. The least element of $(P, \leq)$ is often denoted by $\mathbf{0}$ and called the *zero* of the poset P. Thus every well-orderably-complete poset has a zero. It follows that in every inductive poset and in every directed-complete poset the least element exists.

Let $(P, \leq)$ be a poset. A mapping $\pi : P \to P$ is *monotone* (or *isotone*) if $a \leq b$ implies $\pi(a) \leq \pi(b)$ for every pair $a, b \in A$.

It is easy to see that if π is monotone and C is a chain (a well-ordered chain, a directed subset) in $(P, \leq)$, then the image $\pi[C] := \{\pi(a) : a \in C\}$ is also a chain (a well-ordered chain, a directed subset, respectively).

If $R \subseteq P \times P$ is a binary relation, the system $(P, \leq, R)$ is called an *ordered frame*. In the modal logic jargon, the elements of P are called *worlds*. The fact that aRb holds for some $a, b \in P$, bears various readings:

The world b is possible with respect to a,
The world b is alternative to a,
The world a sees the world b etc.

A relation $R \subseteq P \times P$ is called *serial* (or *left total*, or simply *total*) if for every $a \in P$ there exists an element $b \in P$ (not necessarily unique) such that aRb holds, symbolically

(1) $(\forall a \in P \exists b \in P)\ aRb.$

Definition 1.4. Let $R \subseteq P \times P$ be a binary relation defined on a non-empty set P. An element $a^* \in P$ is called a *fixed-point* of R if $a^* R a^*$ holds. □

Fixed-points of relations are also called *reflexive points*.

In the paper we apply the notation adopted by Desharnais and Möller [2005]. Given a binary relation R on a set P and $a \in P$, we define:

$$aR := \{x \in P : aRx\}$$

and

$$Ra := \{x \in P : xRa\}.$$

aR is called the *R-image* of the element a and Ra is the *R-preimage* of a.

In particular, given a poset $(P, \leq)$ and $a \in P$, we have:

$$a\!\leq = \{x \in P : a\!\leq x\},$$

and

$$\leq a = \{x \in P : x \leq a\}.$$

The sets $a\!\leq$ and $\leq\! a$ are also marked in some papers by $\uparrow\! a$ and $\downarrow\! a$, respectively.

A relation $R \subseteq P \times P$ is called $\forall$-*expansive* if it is serial and included in $\leq$, i.e.,

(2) $\quad (\forall a \in P)\emptyset \neq aR \subseteq a\!\leq.$

$\forall$ — expansive relations are also called *inflationary* (see e.g. Desharnais and Möller [2005]).

We begin the presentation of fixed-point theorems for relations with the following two, rather trivial, observations.

THEOREM 1.5 (The Fixed-Point Theorem for $\forall$-Expansive Relations). *Let $(P, \leq)$ be a poset in which every non-empty chain has an upper bound. Let $R \subseteq P \times P$ be a $\forall$-expansive relation. Then R has a fixed-point a^* which additionally satisfies the following condition*:

(3) $\quad$ *for every $b \in P$, if a^*Rb, then $b = a^*$.*

(This means that $a^*R = \{a^*\}$. In other words, the world a^* sees only itself.)

PROOF. By Kuratowski-Zorn's Lemma, applied to $(P, \leq)$, there exists at least one maximal element a^* (in the sense of $\leq$). We show that a^* is a fixed-point for R. By seriality, there exists $b \in P$ such that a^*Rb. (2) then implies that $a^* \leq b$. Since a^* is maximal, we have that $a^* = b$. Hence a^*Ra^*, i.e., a^* is a reflexive point of R. By maximality and (2), a^* also satisfies (3). ■

It is obvious that if R is a function on P, that is, R satisfies the condition: for every $a \in P$ there exists a unique element $b \in P$ such that aRb holds, symbolically:

$$(\forall a \in P\, \exists!\, b \in P)aRb,$$

then every fixed-point of R satisfies (3).

Let $(P, \leq)$ be a poset. A relation $R \subseteq P \times P$ is called $\exists$-*expansive* if for every $a \in P$ there exists $b \in P$ such that aRb and $a \leq b$, symbolically:

(4) $(\forall a \in P) aR \cap a\leq \neq \emptyset$.

Evidently, every $\exists$-expansive relation is serial and every $\forall$-expansive relation is $\exists$-expansive. It is clear that if R is a total function from P to P, then the properties of being $\forall$-expansive and $\exists$-expansive are equivalent for R.

Theorem 1.6 (The Fixed-Point Theorem for $\exists$-Expansive Relations). *Let $(P, \leq)$ be a poset in which every non-empty chain has an upper bound. Let $R \subseteq P \times P$ be a $\exists$-expansive relation. Then R has a fixed-point a^* which additionally satisfies the following condition*:

(5) $(\forall b \in P) a^*Rb$ *and* $a^* \leq b$ *implies* $b = a^*$.

(This means that $a^*R \cap a^*\leq = \{a^*\}$.)

Proof. Let R_0 be the intersection of the relations R and $\leq$, i.e., $R_0 := R \cap \leq$. The relation R_0 is serial and $\forall$-expansive. By Theorem 1.5, R_0 has a fixed-point a^* for which (3) holds. Consequently, a^*Ra^* and (5) readily follows. ■

The proof of Theorem 1.6 employs the Axiom of Choice (in the form of Zorn's Lemma). But in fact, the set-theoretic status of Theorems 1.5 and 1.6 is the same — each of the above fixed-point theorems is equivalent to the Axiom of Choice.

Metatheorem 1.7. *On the basis of Zermelo-Fraenkel set theory* ZF^- *(=* ZF *without the Axiom of Regularity), the following conditions are equivalent*:

(a) *The Axiom of Choice* (AC).

(b) *Theorem* 1.5.

(c) *Theorem* 1.6.

Proof. The implication (a) $\Rightarrow$ (b) directly follows from the proof of Theorem 1.5 because Zorn's Lemma is used here. The implication (b) $\Rightarrow$ (c) is present in the proof of Theorem 1.6. To prove the implication (c) $\Rightarrow$ (a), assume that Theorem 1.6 holds. We show that then Zorn's Lemma holds. For let $(P, \leq)$ be an arbitrary poset in which every non-empty chain has an upper bound. Let R be equal to $\leq$, i.e., $R := \leq$. The relation R is evidently $\exists$-expansive. By Theorem 1.6, R has a fixed-point a^* which satisfies (5). We show a^* is a maximal element in $(P, \leq)$. Assume $b \in P$ and $a^* \leq b$. So a^*Rb and $a^* \leq b$, by the definition of R. Condition (5) then gives that $b = a^*$. This means that a^* is a maximal element in $(P, \leq)$. ■

The above observation shows that the proofs of fixed-point results based on Theorems 1.5 and 1.6 would require adopting as strong set-theoretic assumptions as the Axiom of Choice.

EXAMPLES. 1. Let $P = [0, 1]$ be the closed unit interval of real numbers. Evidently, the system $(P, \leq)$ with the usual ordering $\leq$ of real numbers satisfies the hypothesis of Theorems 1.5 and 1.6. If the relation R is taken to be equal to $\leq$ on P, then R is $\forall$-expansive. Hence it has a fixed-point in $(P, \leq)$ which additionally satisfies (3). It is clear that 1 is the only such a fixed-point of R. On the other hand, every element of P is a fixed-point of R.

2. Let $\boldsymbol{A}$ be a non-trivial Boolean algebra. If F is a filter of $\boldsymbol{A}$ and $a \in A$, then $Fi(F, a)$ denotes the filter of $\boldsymbol{A}$ generated by the set $F \cup \{a\}$.

Let F_0 be a proper filter of $\boldsymbol{A}$ and let P be the poset consisting of all proper filters of $\boldsymbol{A}$ that include F_0. P is non-empty and ordered by inclusion. We define the relation $R \subseteq P \times P$ as follows. For $F, G \in P$:

$$FRG \text{ iff } [(\exists a \in A - F) G = Fi(F, a)] \vee [(\forall a \in A - F) A = Fi(F, a) \wedge F = G].$$

FRG says that *either* G is an extension of F obtained by adjoining a new element to F *or* $Fi(F, a)$ is an improper filter for all $a \in A - F$ and then $F = G$.

CLAIM 1. *R is serial.*

Indeed, suppose $F \in P$. If the filter $Fi(F, a)$ is proper for some $a \in A - F$, we evidently have that FRG for $G := Fi(F, a)]$. If $Fi(F, a)$ is improper for all $a \in A - F$, then FRG for $G := F$. ■

CLAIM 2. *R is $\forall$- expansive.*

Suppose FRG. It directly follows from the definition of R that $F \subseteq G$. ■

As the poset $(P, \subseteq)$ satisfies the assumption of Theorem 1.5, the above claims imply the existence of a fixed-point F^* of R that satisfies (3). It follows that F^* is actually an ultrafilter of $\boldsymbol{A}$ extending F_0. Thus every proper filter of $\boldsymbol{A}$ can be extended to an ultrafilter. □

Let $(P, \leq)$ be a poset. A mapping $\pi : P \to P$ is *expansive* if $a \leq \pi(a)$ for every $a \in A$.

COROLLARY 1.8 (Zermelo). *Let $(P, \leq)$ be a poset in which every non-empty chain has an upper bound. Let $\pi : P \to P$ be an expansive mapping. Then π has a fixed- point, i.e., there exists $a^* \in P$ such that $\pi(a^*) = a^*$. Furthermore, a^* can be assumed to be a maximal element in $(P, \leq)$.*

Proof. In the standard formulations of set theory, say in the theory ZF of Zermelo-Fraenkel, each function is identified with its graph. Accordingly, let R_π be the graph of π. Thus $aR_\pi b$ holds iff $b = \pi(a)$, for all $a, b \in P$. The relation R_π satisfies the hypotheses of both Theorems 1.5 and 1.6. R_π is serial, because π is a function. R_π is $\forall$-expansive because it is serial and π is expansive. In virtue of Theorem 1.5, R_π has a fixed-point a^*. In particular, $a^* R_\pi a^*$ holds which means that $a^* = \pi(a^*)$. ■

The first statement of Corollary 1.8 is provable in ZF^- and AC is not needed here. Moschovakis [1994] gives a proof of it which is based on Hartog's Theorem.

It is clear that if a poset $(P, \leq)$ is finite, then every non-empty chain has a supremum and there are maximal elements in $(P, \leq)$. It follows that if a mapping $\pi : P \to P$ is expansive, it has fixed-points.

The above theorems prove the existence of rather big fixed-points — they are maximal elements in a given poset. The theorems we present farther are concerned with the problem of finding possibly small fixed-points.

2. Monotone relations

Definition 2.1. Let $(P, \leq)$ be a poset. A binary relation $R \subseteq P \times P$ is called *monotone* if it satisfies:

$(*)$ $(\forall a, b, c \in P)[a \leq b$ and aRc implies $(\exists d \in P) bRd$ and $c \leq d]$

(see the figure below). ☐

$$\begin{tikzcd}
b \arrow[r, "R"] & d \\
a \arrow[u, "\leq"] \arrow[r, "R"'] & c \arrow[u, "\leq"']
\end{tikzcd}$$

In diagrams like this, horizontal arrows are labeled by R and the vertical arrows are labeled by the order sign $\leq$.

In the above definition it suffices to assume that a is strictly less than b, i.e., $a < b$, because if $a = b$, the succedent of the monotonicity condition $(*)$ is satisfied by the element $d := c$.

The monotonicity condition can be then written as:

$$(**)\quad (\forall b, c \in P)[\leq b \cap Rc \neq \emptyset \text{ implies } bR \cap \leq c \neq \emptyset]$$

or equivalently, by contraposition, as

$$(\forall b, c \in P)[bR \cap \leq c = \emptyset \text{ implies } \leq b \cap Rc = \emptyset].$$

If $\pi : P \to P$ is a function, then π is monotone if and only if the graph R_π of π is a monotone relation in the above sense.

NOTES 2.2.

1. Let $\mathcal{P}_0(X)$ denotes the set of all non-empty subsets of a set X.

There is an obvious one-to-one correspondence between serial relations defined on a set P and set-valued mappings from P to $\mathcal{P}_0(P)$. Indeed, if R is serial, then define the mapping $Tr(R) : P \to \mathcal{P}_0(P)$ by $Tr(R)a := aR$ $(= \{x \in P : aRx\})$, for all $a \in P$. Conversely, given a mapping $T : P \to \mathcal{P}_0(P)$, we define the serial relation $Re(T) \subseteq P \times P$ by: $aRe(T)b$ iff $b \in Ta$, for $a, b \in P$. Then $Re(Tr(R)) = R$ and $Tr(Re(T)) = T$.

Let $(P, \leq)$ be a complete lattice. Fujimoto [1984] introduces the notion of an *isotone* set-valued mapping $T : P \to \mathcal{P}_0(P)$ as a mapping which satisfies the condition: for any $a, b \in P$, if $a \leq b$ then for any $c \in Ta$ there exists a $d \in Tb$ such that $c \leq d$. It is not difficult to show that a *serial* relation $R \subseteq P \times P$ is monotone in the sense of the above Definition 2.1 if and only if the derived set-valued mapping $Tr(R) : P \to \mathcal{P}_0(P)$ is isotone in the sense of Fujimoto [1984]. (In fact, Fujimoto's definition of an isotone mapping also makes sense in the wider context of arbitrary posets.) We may therefore say that Definition 2.1, when restricted to serial relations, is equivalent to Fujimoto's definition of isotonicity.

2. Desharnais and Möller [2005] are also concerned with monotone relations R defined on *complete* lattices. The meaning of the term *'monotone relation'* defined by them differs from the one provided by Definition 2.1. In fact, in the context of complete lattices they consider *four* natural conditions imposed on R and each of them is regarded as generalization of the notion of a monotone mapping. In other words, they define four independent meanings of the term 'monotone relation'. The defining conditions introduced by them are marked below by (a), (b), (c), and (d), respectively. ☐

DEFINITIONS. Let $(P, \leq)$ be a complete lattice and let R be a binary relation on P. Consider the following conditions:

(a) $(\forall a, b \in P)[a < b \text{ implies } \inf aR \leq \inf bR]$,

(b) $(\forall a, b \in P)[a < b \text{ implies } \inf aR \leq \sup bR]$,

(c) $(\forall a, b \in P)[a < b \text{ implies } \sup aR \leq \inf bR]$,

(d) $(\forall a, b \in P)[a < b \text{ implies } \sup aR \leq \sup bR]$.

According to Desharnais and Möller [2005], each of these sentences is treated as the *definiens* of the monotonicity definition of R. Consequently, (a)–(d) individually yield four different definitions of monotonicity. $\square$

Desharnais and Möller prove that in the general case the above conditions are independent. But they note that if R is *serial*, then the following implications hold:

$$\text{(c)} \Rightarrow \text{(a)} \Rightarrow \text{(b) and (c)} \Rightarrow \text{(d)} \Rightarrow \text{(b)}.$$

Moreover, if R is a total function, all the four conditions are equivalent and they state that the function R is monotone.

The key theorem proved by Desharnais and Möller (Theorem 5) states that if R is serial and satisfies (c) then R has a least reflexive point.

None of these four conditions is equivalent to Definition 2.1. But in case the poset $(P, \leq)$ is a complete lattice, condition $(*)$ of Definition 2.1 and conditions (a)–(d) are not independent altogether. We have the following simple observations that relate the notions of monotonicity defined by Desharnais and Möller to the notion provided by Definition 2.1.

The first observation shows that in the context of complete lattices, Definition 2.1 is stronger than the definition of monotonicity provided by condition (d):

PROPOSITION 1. *Let $(P, \leq)$ be a complete lattice and let R be a binary relation on P. If R is monotone in the sense of Definition* 2.1*, then R is monotone in the sense of* (d)*.*

PROOF. Let $a, b \in P$ and $a < b$. We must show that $\sup aR \leq \sup bR$. If the set aR is empty, the thesis trivially holds, because $\sup aR = \sup \emptyset = \mathbf{0} \leq \sup bR$. We consider the case aR is non-empty. As R is monotone in the sense of Def. 2.1, it follows that for every $c \in aR$ there exists $d \in bR$ such that $c \leq d$. Consequently, for every $c \in aR$ it is the case that $c \leq \sup bR$. Whence $\sup aR \leq \sup bR$. $\blacksquare$

The second observation says that in the context of complete lattices, the definition of monotonicity of *serial* relations provided by condition (c) is stronger than Definition 2.1:

PROPOSITION 2. *Let $(P, \leq)$ be a complete lattice and let R be a serial binary relation on P. If R is monotone in the sense of (c), then R is monotone in the sense of Definition* 2.1*.*

PROOF. Suppose that $a \leq b$ and aRc for some $a, b, c \in P$. We claim that there exists $d \in P$ such that bRd and $c \leq d$.

The claim evidently holds if $a = b$, because it suffices to put $d := c$. We assume that $a < b$. But then condition (c) implies that $\sup aR \leq \inf bR$. As $c \in aR$, it follows that $c \leq \sup aR \leq \inf bR$. Hence $c \leq \inf bR$.

As R is serial, the set bR is non-empty. Let d be an element of bR. As $\inf bR \leq d$, it follows that $c \leq \inf bR \leq d$. Hence bRd and $c \leq d$. ■

It follows from the above facts that in the class of complete lattices the scope of Definition 2.1 for *serial* relations is limited by conditions (c) and (d) in the sense that the implications (c) $\Rightarrow$ $(*)$ $\Rightarrow$ (d) hold.

In this paper we adopt Definition 2.1 and from now on this definition constitutes the meaning of the term 'monotone relation'.

If a *finite* poset $(P, \leq)$ is endowed with a non-trivial monotone relation R, then R has reflexive points. The non-triviality of R means that there are elements $a, b \in P$ such that $a \leq b$ and aRb. Indeed, let $P_0 := \{y \in P : (\exists x \in P)(x \leq y \wedge xRy)\}$. The subset P_0 is finite and non-empty. Let a^* be a maximal element in P_0 in the sense of $\leq$. As a^* belongs to P_0, there exists x in P such that $x \leq a^*$ and xRa^*. Hence, by monotonicity, there exists $b \in P$ such that a^*Rb and $a^* \leq b$. It follows that b also belongs to P_0. Due to maximality of a^* we get that $a^* = b$. So a^*Ra^*.

In this section we are mainly concerned with *infinite* ordered frames. Our goal is to prove the existence of reflexive points in such frames by applying methods weaker than the Axiom of Choice.

A poset $(P, \leq)$ is called *chain-σ-complete* if $(P, \leq)$ has a zero element $\mathbf{0}$ and every well-ordered chain in $(P, \leq)$ of type $\leq \omega$ has a supremum.

Definition 2.3. Let $(P, \leq)$ be a chain-σ-complete poset and let $R \subseteq P \times P$ be a binary relation.

1. R is called *chain-σ-continuous* if R is monotone and it additionally satisfies the following condition:

(cont)$_\sigma$ For every strictly increasing sequence of elements of P of type ω

$$a_0 < a_1 < \ldots < a_n < a_{n+1} < \ldots$$

and for every increasing sequence

$$b_0 \leq b_1 \leq \ldots \leq b_n \leq b_{n+1} \leq \ldots$$

of elements of P of type $\leq \omega$, if it is the case that a_nRb_n for all n, then $\sup\{a_n : n \in \omega\} R \sup\{b_n : n \in \omega\}$.

(cont)$_\sigma$ can be equivalently formulated as follows:

for every chain C in $(P, \leq)$ of type ω and for every monotone mapping $f : C \to P$, if $aRf(a)$ for all $a \in C$, then $\sup(C) R \sup(f[C])$.

2. R is *chain-σ-continuous in the stronger sense* if it is chain-σ-continuous and additionally satisfies the following condition:

$(*)_\sigma$ For every chain D in $(P, \leq)$ of type ω and for every element $a \in P$, if aRd for all $d \in D$, then $aR \sup(D)$. □

It is not difficult to prove that R is chain-σ-continuous in the stronger sense if and only if it is monotone and satisfies:

(cont)$^*_\sigma$ For any two monotone mappings $f : \omega \to P$ and $g : \omega \to P$, if $f(n)Rg(n)$ for all $n \in \omega$, then $\sup(f[\omega]) R \sup(g[\omega])$.

It is easy to see that in light of Theorem 1.1, the condition (cont)$_\sigma$ can be equivalently formulated as a sentence in which one quantifies over arbitrary countably infinite chains C and not only over chains of type ω. The same remark applies to the condition $(*)_\sigma$.

The proof of Theorem 2.4 below requires a certain weak and plausible form of the Axiom of Choice, viz. the Axiom of Dependent Choices:

Axiom of Dependent Choices (DC).

For each set A, for each binary serial relation R on A and for each $a \in A$, there exists a function $f : \omega \to A$ such that $f(0) = a$ and $f(n)Rf(n+1)$ for all $n \in \omega$.

DC is unprovable from the standard axioms of ZF (Fraenkel, Mostowski). This fact together with the well-known Cohen's proof of the independence of AC from the axioms of ZF implies that DC itself is independent from ZF. Moreover DC is known to be strictly weaker than AC.

The following observation shows that the antecedent of (count)$_\sigma$ is not vacuously satisfied for certain relations:

OBSERVATION. *Assume the Axiom of Dependent Choices. Let $(P, \leq)$ be a poset. If a relation $R \subseteq P \times P$ is a monotone and serial, then for every chain $a_0 < a_1 < \ldots a_n < a_{n+1} < \ldots$ in $(P, \leq)$ of type ω, there exists a countable chain $b_0 \leq b_1 \leq \ldots \leq b_n \leq b_{n+1} \leq \ldots$ in $(P, \leq)$ of type less or equal to ω such that $a_n R b_n$ for all n.*

Indeed, let $a_0 < a_1 < \ldots a_n < a_{n+1} < \ldots$ be a chain of type ω. As R is serial, there exists an element $b_0 \in P$ such that if $a_0 R b_0$. Since $a_0 \leq a_1$ and $a_0 R b_0$, the monotonicity of R implies the existence of an element $b_1 \in P$ such that $a_1 R b_1$ and $b_0 \leq b_1$. As $a_1 \leq a_2$ and $a_1 R b_1$, there exists an element $b_2 \in P$ such that $a_2 R b_2$ and $b_1 \leq b_2$, again by monotonicity. Going farther, as $a_2 \leq a_3$ and a_2Rb_2, there exists an element $b_3 \in P$ such that a_3Rb_3 and $b_2 \leq b_3$. Continuing this procedure, we define a countable chain $b_0 \leq b_1 \leq \ldots \leq b_n \leq b_{n+1} \leq \ldots$ in $(P, \leq)$ such that a_nRb_n for all n. □

The above proof is an application of the so called Principle of Indeterminate Definability by Arithmetical Recursion. (We shall later discuss this principle in detail.) This principle is frequently used in mathematical proofs, often without mentioning it. The idea is that in the induction base one freely picks out one of several options, say c_0. Then, in the induction step, having defined elements $c_0, c_1, \ldots, c_n$, one has a range for freedom of picking out the consecutive element. Let it be c_{n+1}. We shall later check that the Principle of Indeterminate Definability by Arithmetical Recursion is equivalent to the Axiom of Dependent Choices.

(We leave it to the reader to check that DC really is used in the above proof. Here is an outline of the formal proof. Given a chain $a_0 < a_1 < \ldots a_n < a_{n+1} < \ldots$ of type ω in $(P, \leq)$, we define the following set of triples:

$$A := \{(a_n, a_{n+1}, c) : n \in \omega, a_nRc, c \in P\}.$$

Then the binary relation S is defined on the set A according to the formula:

$$(x_1, x_2, x_3)S(y_1, y_2, y_3) \text{ iff } (\exists n \in \omega)(x_1 = a_n \wedge y_1 = a_{n+1} \wedge x_3 \leq y_3).$$

The fact that R is monotone implies that S is serial. Indeed, suppose $(x_1, x_2, x_3) \in A$. Then, for some $n \in \omega$, $x_1 = a_n$, $x_2 = a_{n+1}$, and a_nRx_3. As $a_n < a_{n+1}$ and a_nRx_3, the monotonicity of R implies the existence of $y_3 \in P$ such that $a_{n+1}Ry_3$ and $x_3 \leq y_3$. Putting $y_1 := a_{n+1}$ and $y_2 := a_{n+2}$, we see that $(y_1, y_2, y_3) \in A$ and $(x_1, x_2, x_3)S(y_1, y_2, y_3)$.

As R is serial, there exists $b_0 \in P$ such that $(a_0, a_1, b_0) \in A$. By the seriality of S and DC, there exists a mapping $g : \omega \to A$ such that $f(0) = (a_0, a_1, b_0)$ and $g(n)Sg(n+1)$ for all $n \in \omega$. We then define $f : \omega \to P$ as follows. Let $n \in \omega$. Suppose $g(n) = (a_n, a_{n+1}, b)$. Then $f(n) := b$. Evidently, $f[\omega]$ is a chain such that $a_nRf(n)$ for all $n \in \omega$.)

The basic observation concerning fixed-points of σ-continuous relations is provided by the following theorem:

THEOREM 2.4. *Assume the Axiom of Dependent Choices. Let* $(P, \leq)$ *be a* σ*-chain-complete poset. Every chain-*σ*-continuous relation* $R \subseteq P \times P$ *such that the set* $\mathbf{0}R$ *is non-empty has a fixed-point* a^*. *Moreover,* a^* *can be assumed to have the following property: for every* $y \in P$, *if* $yR \subseteq {\leq} y$, *then* $a^* {\leq} y$.

Proof. $\mathbf{0}$ is the least element in $(P, \leq)$. $\mathbf{0}$ is the supremum of the empty chain. As R is chain-σ-continuous, R is a monotone relation in $(P, \leq)$.

We define:

$$Q := \{x \in P : xR \cap x{\leq} \neq \emptyset \text{ and } (\forall y \in P)(yR \subseteq {\leq}y \text{ implies } x \leq y)\}.$$

We notice that $\mathbf{0} \in Q$, because the set $\mathbf{0}R$ is non-empty.

Lemma 1. *$R\lceil Q$, the restriction of R to Q, is $\exists$-expansive in the poset* $(Q, \leq)$.

Proof of the lemma. Let $a \in Q$. As $aR \cap a{\leq} \neq \emptyset$, there exists $b \in P$ such that

(1) aRb and $a {\leq} b$.

We prove that $b \in Q$. This will show that $R\lceil Q$ is $\exists$-expansive. By (1) and the monotonicity of R in $(P, \leq)$, there exists $c \in P$ such that bRc and $b \leq c$ (see the figure below). This means that the set $bR \cap b{\leq}$ is non-empty.

$$\begin{tikzcd}
b \arrow[r, "R"] & c \\
a \arrow[u, "\leq"] \arrow[r, "R"'] & b \arrow[u, "\leq"']
\end{tikzcd}$$

Now let y be an arbitrary element of P such that $yR \subseteq {\leq}y$. We show $b \leq y$. As $a \in Q$ and $yR \subseteq {\leq}y$, we have that $a {\leq} y$. As aRb, the monotonicity of R implies the existence of an element $z \in P$ such that yRz and $b \leq z$ (see the diagram below). Since $z \in yR$ and $yR \subseteq {\leq}y$, we get that $z \leq y$. We thus have that $b \leq z \leq y$, which gives that $b \leq y$. This proves that $b \in Q$. ■

$$\begin{tikzcd}
y \arrow[r, "R"] & z \\
a \arrow[u, "\leq"] \arrow[r, "R"'] & b \arrow[u, "\leq"']
\end{tikzcd}$$

If C is a well-ordered chain in $(P, \leq)$ of type ω and $f : C \to P$ is a monotone mapping, then the f-image $f[C]$ is a chain of type $\leq \omega$. Hence

$\sup(f[C])$ exists in $(P, \leq)$. Evidently, every chain in $(Q, \leq)$ is also a chain in $(P, \leq)$. Consequently, by the σ-continuity of R:

(2) *For every chain C in $(Q, \leq)$ of type ω and for every monotone mapping $f : C \to P$ such that $xRf(x)$ for all $x \in C$, it is the case that* $\sup(C) R \sup(f[C])$.

LEMMA 2. *Let C be a chain in $(Q, \leq)$ of type ω. Assume that there exists a monotone mapping $f : C \to Q$ such that $xRf(x)$ and $x \leq f(x)$ for all $x \in C$. Then the supremum* $\sup(C)$ *belongs to Q.*

PROOF OF THE LEMMA. Let $m := \sup(C)$. We show $m \in Q$. By (2) we have that $mR \sup(f[C])$. Furthermore, as $x \leq f(x)$ for all $x \in C$, it follows that $m = \sup(C) \leq \sup(f[C])$. This shows that $\sup(f[C]) \in mR \cap m{\leq}$. Hence

(3) $mR \cap m{\leq} \neq \emptyset$.

Now let $y \in P$ be an element such that $yR \subseteq {\leq}y$. We claim that $m \leq y$. As $C \subseteq Q$, we have that $x \leq y$ for all $x \in C$, by the second conjunct of the definition of Q. It follows that $m = \sup(C) \leq y$. This and (3) prove that $m \in Q$. ■

We now proceed to the proof of the theorem. We inductively define a strictly increasing sequence $a_0 < a_1 < \ldots < a_n < a_{n+1} < \ldots$ of elements of Q (see the diagram below). The type of the sequence is smaller or equal to ω. We assume the Axiom of Dependent Choices and define: $a_0 := \mathbf{0}$. Suppose the elements $a_0 < a_1 < \ldots < a_n$ have been defined. As $a_n \in Q$ and, by Lemma 1, the relation $R\lceil Q$ is $\exists$-expansive, there exists an element $b \in Q$ such that $a_n \leq b$ and $a_n R b$. If $b = a_n$, the defining procedure terminates. In this case $a^* := a_n$ is already a fixed-point of R. As a^* belongs to Q, the second statement of the thesis of the theorem evidently holds for a^*. If $b \neq a_n$, we put: $a_{n+1} := b$. Clearly $a_n < a_{n+1}$.

It remains to consider the case when the sequence $a_0 < a_1 < \ldots < a_n < a_{n+1} < \ldots$ has type ω. In this case we put $C := \{a_n : n \in \omega\}$ and define $f : C \to P$ by $f(a_n) := a_{n+1}$ for all $n \in \omega$. f is well-defined and monotone. As $C \subseteq Q$, $aRf(a)$ and $a {\leq} f(a)$ for all $a \in C$, the supremum $\sup(C)$ belongs to Q, by Lemma 2. Furthermore $\sup(C) R \sup(f[C])$, by the σ-continuity of R. But evidently $\sup(C) = \sup(f[C])$ because $a_0 = \mathbf{0}$ (see the figure below). Putting $a^* := \sup(C)$, we thus see that $a^* R a^*$. So a^* is a reflexive point of R. Since $a^* \in Q$, it follows that for every $y \in P$, if $yR \subseteq {\leq}y$, then $a^* \leq y$. ■

$$\begin{tikzcd}
{} & \cdots\cdots & {} \\
a_2 \arrow[u, dash] \arrow[rr, "R"] & & a_3 \arrow[u, dash] \\
a_1 \arrow[u, "\leq"] \arrow[rr, "R"] & & a_2 \arrow[u, "\leq"'] \\
a_0 \arrow[u, "\leq"] \arrow[rr, "R"] & & a_1 \arrow[u, "\leq"'] \\
\mathbf{0} \arrow[u, "\leq"] \arrow[rr, "R"'] & & a_0 \arrow[u, "\leq"']
\end{tikzcd}$$

Notes 2.5.

(**1**). The assumption of σ-continuity, and hence monotonicity of R, in Theorem 2.4 is essential and cannot be dropped altogether. For let $P = [0, 1]$ be the closed unit interval of real numbers. Evidently, the system $(P, \leq)$ with the usual ordering $\leq$ of real numbers is a chain-σ-complete poset. The relation $R \subseteq P \times P$ is defined as follows:

$$aRb \text{ if and only if } (\exists n \in \omega, n \geq 1)|a - b| = (1/2)^n.$$

It is easy to see that:

(a) R is symmetric and serial,

(b) R is not monotone,

(c) R does not posses a fixed-point.

As to (b), observe that for the numbers $1/2$ and 1, we have $1/2R1$ and $1/2 < 1$. But there does not exists a number $d \in [0, 1]$ such that $1Rd$ and $1 \leq d$.

(**2**). The hypothesis that $(P, \leq)$ is chain-σ-complete is essential in Theorem 2.4. For let $P = \{1, 2, a, b\}$, where $1 < 2$ and $a < b$. Let R be the relation on P defined as follows:

$$R := \{\langle 1, a\rangle, \langle a, 1\rangle, \langle 2, b\rangle, \langle b, 2\rangle\}.$$

R is symmetric and monotone. R is also chain σ-continuous. But R does not possess a fixed-point. The poset $(P, \leq)$ is not chain-σ-complete because it does not have the least element, the supremum of the empty chain. $\square$

Let $(P, \leq)$ be a chain-σ-complete poset. A mapping $\pi : P \to P$ is called *order σ-continuous* if it is monotone and

$$\pi(\sup(C)) = \sup(\pi[C])$$

for every chain C in $(P, \leq)$ of type ω.

Since for any monotone mapping $\pi : P \to P$, the image $\pi[C]$ of any chain C in $(P, \leq)$ is a chain as well, we see that, in view of the chain-σ-completeness of $(P, \leq)$, the above formula thus postulates the equality of the two supremums and not their existence.

It is clear that a mapping $\pi : P \to P$ is σ-continuous in the above sense if and only if the graph R_σ is σ-continuous as a binary relation.

The following well-known observation is a direct consequence of Theorem 2.4.

THEOREM 2.6. *Let $(P, \leq)$ be a chain-σ-complete poset. Every σ-continuous mapping $\pi : P \to P$ has a least fixed-point a^*, i.e., $\pi(a^*) = a^*$ and*

$$(\forall b \in P)(\pi(b) \leq b \text{ implies } a^* \leq b).$$

PROOF. The above result does not require the Axiom of Dependent Choices but it requires the Arithmetic Induction Principle (see Metatheorem 3.1 below). We work with the graph R_π of π and proceed as the proof of Theorem 2.4. Since π is σ-continuous, the graph R_π i is a chain-σ-continuous relation and $\mathbf{0}R_\pi = \{\pi(\mathbf{0})\}$. By Theorem 2.4, there is a fixed-point a^* of R_π such that for every $b \in P$, $bR_\pi \subseteq b{\leq}$ implies that $a^* \leq b$. But the last condition simply says that for every $b \in P$, $\pi(b) \leq b$ implies $a^* \leq b$.

Let us also note that the chain $a_0 \leq a_1 \leq \ldots \leq a_n \leq a_{n+1} \leq \ldots$, defined as in the proof of Theorem 2.4 for the relation R_π, has the following properties:

$$a_0 = \mathbf{0},$$

$$a_{n+1} = \pi(a_n), \text{ for all } n.$$

Furthermore, $a^* = \sup(\{a_n : n \in \omega\})$. ■

Notes 2.7.

1. The above chain $a_0 \leq a_1 \leq \ldots \leq a_n \leq a_{n+1} \leq \ldots$ is known as *Kleene's approximation sequence* (Kleene [1952]. It can also be found in Tarski's [1955] classical paper. Many properties of such objects are compiled in Chapter 4 of Davey and Priestley [2002].

2. Apart from the notion of chain-σ-completeness of a poset, one can define the relative properties formulated in terms of countable well-ordered chains and countable directed subsets.

We say that a poset $(P, \leq)$ is:

(A) *linearly-σ-complete* if every countable chain C in $(P, \leq)$ (not necessarily of type ω) has a supremum,

(B) *directed-σ-complete* if every countable directed subset D in $(P, \leq)$ has a supremum.

It is clear that the chain-σ-completeness implies property (A) and that (B) implies the chain-σ-completeness. However, it readily follows from Theorem 1.1 that in the presence of the Axiom of Dependent Choices the above three properties are equivalent. In the sequel we will not carefully distinguish between these three situations. We shall however uniformly formulate the results discussed in this chapter in terms of the chain-σ-completeness of posets.

3. Theorem 2.6 should be compared with Theorem 3.6 in the next section — see Note following Theorem 3.6.

4. Let $(P, \leq)$ be a complete lattice and let $\pi : P \to P$ be a monotone mapping. Tarski's Fixed-Point Theorem (Tarski [1955]) states that the set of fixed-points of π forms a non-empty complete lattice for the ordering of $(P, \leq)$. The weak version of the above theorem, stating that under the above assumptions π has a fixed-point, is called the Knaster-Tarski theorem, because a special case of this theorem (for the lattices of powersets) was proved by Knaster already in 1928. There is vast literature devoted to Tarski's Theorem (in both versions). This theorem was generalized in various directions - see e.g. Berman and Blok [1988]. Fujimoto [1984] extends the Knaster-Tarski theorem to the case of set-valued mappings. Fujimoto's paper [1984] is chronologically the first work in the literature which is (implicitly) concerned with fixed-points of relations. In the paraphrased but equivalent form, expressed in terms of monotone relations, the main result proved by Fujimoto says the following: $\square$

Theorem. *Let $(P, \leq)$ be an inductive poset. Let R be a serial and monotone relation defined on $(P, \leq)$. Suppose moreover that for every $a \in P$ the subset $K_a := \bigcup\{{\leq} u : u \in aR\}$ of P is inductive. Then R has a fixed-point.*

In particular, if $(P, \leq)$ is a complete lattice, $\pi : P \to P$ is a mapping, and R is the graph of π, then the set K_a coincides with ${\leq}\pi(a)$ $(= \{x \in P : x \leq \pi(a)\})$. The last set is inductive. It follows from the above theorem that if π is monotone then it has as a fixed-point. Thus the Knaster-Tarski theorem is a consequence of Fujimoto's result.

For the sake of completeness of the presentation we shall insert here a short proof of Fujimoto's Theorem and show how to relate this theorem to Theorem 1.6.

If R is serial, then the set aR is non-empty, for all $a \in P$. Consequently, for any $a \in P$ the set K_a is non-empty, because it contains $\mathbf{0}$.

The inductivity of K_a says that if C is an arbitrary chain in $(P, \leq)$ such that for every $c \in C$ there is an element $u \in P$ such that $c \leq u$ and aRu, then taking $\sup C$ we also find an element u such that $\sup C \leq u$ and aRu. In other words, if a sees a world above each element of C, then a sees a world above $\sup C$ either.

We then define

$$Q := \{x \in P : xR \cap x{\leq} \neq \emptyset\}.$$

The following observation is due to Fujimoto:

Lemma 1. *Q is a non-empty inductive subset of $(P, \leq)$.*

Proof of the Lemma. As $\mathbf{0} \in Q$ by seriality, it follows that the supremum of the empty chain belongs to Q.

Let A be a non-empty chain in Q and let $b := \sup A$ in $(P, \leq)$.

Claim. $A \subseteq K_b$.

Proof. Suppose $a \in A$. As $A \subseteq Q$ the definition of Q implies that there exists an element $c \in P$ such that aRc and $a \leq c$. Evidently, $a \leq b$. Then by monotonicity of R, there exists $d \in P$ such that bRd and $c \leq d$.

$$\begin{tikzcd}
b \arrow[r, "R"] & d \\
a \arrow[u, "\leq"] \arrow[r, "R"'] & c \arrow[u, "\leq"']
\end{tikzcd}$$

As $a \leq c$ and $c \leq d$, we have that $a \leq d$. Thus bRd and $a \leq d$. This implies that $a \in K_b$. The claim is proved. ■

Since $A \subseteq K_b$ the assumption of the theorem implies that $\sup A$ belongs to K_b, i.e., $b \in K_b$. This means that there exists $c \in P$ such that bRc and $b \leq c$. Hence $b \in Q$.

This proves the lemma. ■

We also have:

LEMMA 2. *R is $\exists$-expansive on Q.*

PROOF OF THE LEMMA. Assume $x \in Q$. Then, by the definition of Q there exists $y \in P$ such that $x \leq y$ and xRy. As R is monotone on P, it follows that there exists $z \in P$ such that yRz and $y \leq z$. This shows that $y \in Q$. As xRy and $x \leq y$, we see that R is $\exists$-expansive on Q. ■

Applying Theorem 1.6 to Lemmas 1-2, we conclude that there is an element $a^* \in Q$ such that a^*Ra^* and for every $b \in Q$, a^*Rb and $a^* \leq b$ implies $b = a^*$. The second conjunct says that a^* is a maximal element in Q.

This concludes the proof of Fujimoto's Theorem. ■

The fact that Fujimoto's Theorem can be derived from Theorem 1.6, once Lemma 1 is established, is not surprising, because Theorem 1.6 is equivalent to Kuratowski-Zorn's Lemma. The original proof of Fujimoto's Theorem makes use of Zorn's lemma.

3. Arithmetic Recursion and Fixed-Points

We shall now investigate more closely the logical status of Theorems 2.4 and 2.6. We first recall the Induction Principle for natural numbers and the Principle of Definability by Arithmetic Recursion. We present the last principle in two versions.

ω is the least non-empty limit ordinal. The elements of ω are called *finite ordinals* or *natural numbers*. Thus $0 = \emptyset$ is the least natural number and $n + 1 = \{0, \ldots, n\}$, for all $n \in \omega$.

The Arithmetic Induction Principle.

For each set $X \subseteq \omega$,
$0 \in X \wedge (\forall n \in \omega)(n \in X \rightarrow n + 1 \in X) \rightarrow X = \omega$.

The following fundamental results in axiomatic number theory follow from the Induction Principle.

The Principle of Definability by Arithmetic Recursion. Version I.

Let A be a non-empty set and let $g : A \to A$ be a function. Then for every $a \in A$ there exists exactly one function $f : \omega \to A$ such that $f(0) = a$ and $f(n+1) = g(f(n))$, for all $n \in \omega$.

If A is a set and $n \in \omega$, then A^n is the set of functions from n to A. Note that $A^0 = \{\emptyset\}$. We then define $A^{<\omega} := \bigcup_{n\in\omega} A^n$.

The Principle of Definability by Arithmetic Recursion. Version II.

Let A be a non-empty set and let $G : A^{<\omega} \to A$ be a function. Then there exists exactly one function $F : \omega \to A$ such that $F(n) = G(F\lceil n)$, for all $n \in \omega$.

($F\lceil n$ is the restriction of F to the set $n = \{0, 1, \dots, n-1\}$. Hence $F\lceil n \in A^n$).

Both principles are also jointly called Arithmetic Recursion Principles. The second version of the Recursion Principle is called Complete Recursion in the literature (see e.g. Moschovakis [1994], p. 72).

METATHEOREM 3.1. *In the presence of the Arithmetic Induction Principle the following conditions are equivalent*:

(a) *The Principle of Definability by Arithmetic Recursion. Version I.*

(b) *The Principle of Definability by Arithmetic Recursion. Version II.*

(c) *Theorem* 2.6.

PROOF. (b) $\Rightarrow$ (a). Assume (b). Let $g : A \to A$ be a function and let a be an arbitrary but fixed element of A. We define $G : A^{<\omega} \to A$ as follows. Assume $p \in A^n$ for some $n > 0$, i.e., $p = \{\langle 0, a_0\rangle, \dots, \langle n-1, a_{n-1}\rangle\}$. Then $G(p) := g(a_{n-1})$. If $n = 0$, $A^0 = \{\emptyset\}$. Then $G(\emptyset) := a$, where a is defined as above. G is thus well-defined. As (b) holds, there exists a unique function $F : \omega \to A$ such that $F(n) = G(F\lceil n)$, for all $n \in \omega$. For $n = 0$ we have that $F(0) = G(F\lceil 0) = G(F\lceil\emptyset) = G(\emptyset) = a$. For each $n \in \omega$, $F(n+1) = G(F\lceil n+1) = G(F\lceil\{0, \dots, n\}) = G(\{\langle 0, F(0)\rangle, \dots, \langle n, F(n)\rangle\} = g(F(n))$. We thus see that F has the required properties, viz. $F(0) = a$ and $F(n+1) = g(F(n))$, $n \in \omega$. So (a) holds.

(c) $\Rightarrow$ (b). Let A be a non-empty set and let $G : A^{<\omega} \to A$ be a function. We claim that there exists a unique function $F : \omega \to A$ such that $F(n) = G(F\lceil n)$, $n \in \omega$.

We define $P := A^{<\omega} \cup A^{\omega}$. The set P is ordered by inclusion. We let $\mathbf{0}$ denote the empty function. $\mathbf{0}$ is the smallest element in $(P, \subseteq)$. We notice that if $p \in P$, then $Dom(p) = \omega$ or $Dom(p) = n$ for some $n \in \omega$. It follows from this remark that if

$$p_0 \subset p_1 \subset \ldots \subset p_n \subset p_{n+1} \subset \ldots$$

is a strictly increasing ω-chain of members of P, then the union $\bigcup_{n\in\omega} p_n$ is a *total* function from ω to A and hence a member of P. Consequently, the poset $(P, \subseteq)$ is chain σ-complete.

For each function $p \in P$ we define the function $\pi(p) \in P$ as follows. We consider two cases.

Case 1. p is a total function, i.e., $Dom(p) = \omega$.

We then put: $Dom(\pi(p)) = \omega$ and $\pi(p)(n) := G(p\lceil n)$, $n \in \omega$.

Case 2. p is a partial function, i.e., $Dom(p) = n$.

Then $Dom(\pi(p)) = n + 1$ and $\pi(p)(i) := G(p\lceil i)$, $i < n + 1$.

Note that in both cases $\pi(p)(0) = G(p\lceil 0) = G(\mathbf{0})$. We also have the following obvious observation:

Claim 1. *π maps P into P.*

Claim 2. *The mapping $\pi : P \to P$ is monotone.*

Proof of the claim. We observe that π acts as a substitution: for each $p \in P$ and each $i \in Dom(p)$, π replaces the value $p(i)$ by $G(p\lceil i)$. Furthermore, if p is not total and $Dom(p) = n$, π adjoins the pair $\langle n, G(p\lceil n)\rangle$ to the graph $\{\langle i, G(p\lceil i)\rangle : i \in Dom(p)\}$. It is then clear that for $p, q \in P$, $p \subseteq q$ implies $\pi(p) \subseteq \pi(q)$. ■

Claim 3. *$\pi : P \to P$ is chain σ-continuous.*

Proof of the claim. As π is monotone, it suffices to prove that $\pi(\sup C) = \sup \pi[C]$ for every chain $C \subseteq P$ of type ω. Let C be a strictly increasing chain

$$p_0 \subset p_1 \subset \ldots \subset p_n \subset p_{n+1} \subset \ldots$$

of elements of P. For each $n \in \omega$ we have that $Dom(p_n)$ is a natural number and $n \subseteq Dom(p_n)$. (The Arithmetic Induction Principle is used here.) Let $p := \bigcup_{n\in\omega} p_n$. Then $Dom(p) = \omega$ and, consequently $Dom(\pi(p)) = \omega$ and $\pi(p)(n) = G(p\lceil n)$ for any $n \in \omega$. We then have, for each $i \in \omega$ and each $a \in A$, the following chain of equivalent conditions:

$$\begin{aligned}
&\langle i, a\rangle \in \pi(\sup C),\\
&a = \pi(p)(i),\\
&(\exists n \in \omega) i < n \wedge a = G(p\lceil n)(i),\\
&(\exists n \in \omega) i < n \wedge a = G(p_n\lceil n)(i),\\
&(\exists n \in \omega) i < n \wedge a = \pi(p_n)(i),\\
&\langle i, a\rangle \in \textstyle\bigcup_{n\in\omega} \pi(p_n),\\
&\langle i, a\rangle \in \sup \pi[C].
\end{aligned}$$

■

CLAIM 4. *If $p \in P$ is a fixed-point of π, then p is a total function.*

PROOF OF THE CLAIM. Suppose p is a fixed-point of π and $Dom(p) = n$. According to the definition of $\pi(p)$, the function $\pi(p)$ is defined at n. As $\pi(p) = p$, it follows that p itself is defined at n, which is impossible. ■

We can now prove (b). In view of (c), π has a least fixed-point, say F. It follows from Claim 4 that F is a unique fixed-point of π. Indeed, suppose p is a fixed-point of π. As F is the least fixed-point, we have that $F \subseteq p$. But $\omega = Dom(F) = Dom(p)$, by Claim 4. Hence $F = p$.

As $\pi(F) = F$ and F is total, we have that $F(n) = \pi(F)(n) = G(F\lceil n)$, $n \in \omega$, by the definition of π. So (b) holds.

(a) $\Rightarrow$ (c). Let $(P, \le)$ be a chain-σ-complete poset. Assume $\pi : P \to P$ is a σ-continuous mapping. By (a) there exists exactly one function $f : \omega \to P$ such that $f(0) = \mathbf{0}$ and $f(n+1) = \pi(f(n))$, for all $n \in \omega$. The Arithmetic Induction Principle then yields:

CLAIM 5. *$f[\omega]$ is a chain of type $\le \omega$.*

PROOF OF THE CLAIM. Define the property Q of natural numbers (i.e., a subset of ω) by:

$$Q(n): \quad f(n) \le f(n+1).$$

Since $f(0) = \mathbf{0} \le f(1)$, $Q(0)$ holds. Assume $Q(n)$ holds, i.e., $f(n) \le f(n+1)$. As π is monotone, we get that $f(n+1) = \pi(f(n)) \le \pi(f(n+1)) = f(n+2)$. So $Q(n+1)$ holds. Thus $Q(n)$ holds for all $n \in \omega$. The claim follows. ■

As $(P, \le)$ is chain-σ-complete, $a^* := \sup f[\omega]$ exists in P.

CLAIM 6. *a^* is the least fixed-point of π.*

Proof of the claim. As $f(n) \leq a^*$ for all $n \in \omega$, we have, by monotonicity, that $\pi(f(n)) \leq \pi(a^*)$ for all $n \in \omega$. Since $\mathbf{0}$ is the least element in $(P, \leq)$, $a^* = \sup\{f(n+1) : n \in \omega\}$. Consequently, by continuity

$$\pi(a^*) = \pi(\sup\{f(n) : n \in \omega\}) = \sup\{\pi(f(n)) : n \in \omega\} =$$
$$\sup\{f(n+1) : n \in \omega\} = a^*.$$

So a^* is a fixed-point of π. To prove that a^* is the least fixed-point of π, suppose $b \in P$ and $\pi(b) \leq b$. Define the property Q of natural numbers as follows:

$$Q(n) : \quad f(n) \leq b.$$

Applying the Arithmetic Induction Principle and the fact that $\pi(b) \leq b$, we easily obtain that $Q(n)$ holds for all natural numbers n. Consequently, $a^* := \sup f[\omega] \leq b$. So (c) holds.

This concludes the proof of Metatheorem 3.1. ■

It is a well-known fact that each of the above statement (a), (b), (c) of Metatheorem 3.1 is provable in the standard set theory ZF of Zermerlo and Fraenkel. It is also a trivial fact from classical logic that if φ and ψ are theorems of whatever theory T, their equivalence $\varphi \leftrightarrow \psi$ is a theorem of T as well. Thus, in a trivial way, the above conditions (a)–(c) are all equivalent on the basis of ZF. Of course, we can say more: *any* two mathematical facts provable in ZF, say '$2 + 2 = 4$' and 'There exists a limit ordinal greater than 0', are deductively equivalent. But this observation gives us no *direct* insight into the proof of the equivalence of these facts, nor in particular into the equivalence of conditions the above conditions (a)–(c). The significance of Metatheorem 3.1 consists in the fact it underlies *logically relevants* facts that are needed in proving each of the conditions (a)–(b) on the basis of any other condition from this list.

We now pass to the discussion on the logical (or rather set-theoretic) status of Theorem 2.4. We first recall the Principle of Countable Choice (AC_ω):

The Principle of Countable Choice ($\mathbf{AC}_\omega$).
For each non-empty set A, for each binary serial relation $R \subseteq \omega \times A$, there exists a function $f : \omega \to A$ such that $nRf(n)$ for all $n \in \omega$.

It is well-known that AC_ω is constructively equivalent to the proposition: every countable, infinite family X of non-empty and pairwise disjoint sets admits a choice function, i.e., there exists a function f defined on X such that $f(A) \in A$ for all$A \in X$. (Two propositions are constructively equivalent if their equivalence can be established on the basis of the axioms of ZF^- without appealing to any choice principle whatsoever — see Moschovakis [1994], p. 127. ZF^- stands for Zermelo-Fraenkel set theory without the Axiom of Regularity.)

AC_ω is known to be weaker than the Axiom of Dependent Choices but independent of the axioms of ZF.

We define the following indeterminate versions of the Principle of Definability by Arithmetic Recursion.

We recall that $\mathcal{P}_0(A)$ is the set of non-empty subsets of a set A.

> **The Principle of Indeterminate Definability by Arithmetic Recursion. Version I.**
>
> *Let A be a non-empty set and let $g : A \to \mathcal{P}_0(A)$ be a function. Then for every $a \in A$ there exists a function $f : \omega \to A$ such that $f(0) = a$ and $f(n+1) \in g(f(n))$, for all $n \in \omega$.*

> **The Principle of Indeterminate Definability by Arithmetic Recursion. Version II.**
>
> *Let A be a non-empty set and let $G : A^{<\omega} \to \mathcal{P}_0(A)$ be a function. Then there exists a function $F : \omega \to A$ such that $F(n) \in G(F\lceil n)$, for all $n \in \omega$.*

The above theorems are also called Indeterminate Arithmetic Recursion Principles. They are often applied in various set-theoretic contexts. The procedures based on these principles are referred to as inductive ones. We mention as examples the proof of the Baire Category Theorem, the proof of the instance $\mathrm{MA}(\omega)$ of Martin's Axiom (see Kunnen [1999], Lemma 2.6 (c), p. 54), or the proof of the equivalence of various versions of the Axiom of Regularity in set theory. The proofs of the above results apply Indeterminate Arithmetic Recursion Principles.

The following observation establishes the relationship between the fixed-point theorem provided by the statement of Theorem 2.4 and the above recursion principles.

METATHEOREM 3.2. *Assume the Arithmetic Induction Principle. In the presence of the Principle of Countable Choice the following conditions are equivalent*:

(1) *The Axiom of Dependent Choices.*

(2) *The Indeterminate Arithmetic Recursion Principle. Version I.*

(3) *The Indeterminate Arithmetic Recursion Principle. Version II.*

(4) *The statement of Theorem* 2.4.

Proof. (1) and (2) are trivially equivalent because every binary relation R on an arbitrary set A can be viewed as a set-valued function $g : A \to \mathcal{P}(A)$, where $g(a) := \{b \in A : aRb\}$ for each $a \in A$, and every serial relation $R \subseteq A \times A$ can be viewed as a function $g : A \to \mathcal{P}_0(A)$ (see Note 2.2).

(3) ⇒ (2). Assume (3). Let $g : A \to \mathcal{P}_0(A)$ be a function and let a be an arbitrary but fixed element of A. We define $G : A^{<\omega} \to \mathcal{P}_0(A)$ as follows. Assume $p \in A$ for some $n > 0$, i.e., $p = \{\langle 0, a_0\rangle, \dots, \langle n-1, a_{n-1}\rangle\}$ for some $a_0, \dots, a_{n-1} \in A$. We then put $G(p) := g(a_{n-1})$. If $n = 0$, $A^0 = \{\emptyset\}$. Then $G(\emptyset) := \{a\}$, where a is defined as above. G is thus well-defined. As (3) holds, there exists a function $F : \omega \to A$ such that $F(n) \in G(F\lceil n)$, for all $n \in \omega$. For $n = 0$ we have that $F(0) \in G(F\lceil 0) = G(F\lceil\emptyset) = G(\emptyset) = \{a\}$. Hence $F(0) = a$. For each $n \in \omega$, $F(n+1) \in G(F\lceil n+1) = G(F\lceil\{0, \dots, n\}) = G(\{\langle 0, F(0)\rangle, \dots, \langle n, F(n)\rangle\} = g(F(n))$. We thus see that F has the required properties: $F(0) = a$ and $F(n+1) \in g(F(n))$, $n \in \omega$. So (2) holds.

(4) ⇒ (3). Let A be a non-empty set and let $G : A^{<\omega} \to \mathcal{P}_0(A)$ be a function. We claim that there exists a function $F : \omega \to A$ such that $F(n) \in G(F\lceil n)$ for all $n \in \omega$.

We define $P := A^{<\omega} \cup A^{\omega}$. The set P is ordered by inclusion. $\mathbf{0}$ denotes the empty function. $\mathbf{0}$ is the smallest element in $(P, \subseteq)$. ($(P, \subseteq)$ is thus identical with the poset defined in the proof of the implication (c) ⇒ (b) of Metatheorem 3.1.) The poset $(P, \subseteq)$ is chain σ-complete.

We define the following binary relation $R \subseteq P \times P$. Let $p, q \in P$. We consider two cases.

Case 1. p is a total function, i.e., $Dom(p) = \omega$.

Then pRq if and only if $Dom(q) = \omega$ and $q(n) \in G(p\lceil n)$ for all $n \in \omega$.

Case 2. p is a partial function, i.e., $Dom(p) = n$ for some $n \in \omega$.

Then pRq if and only if $Dom(q) = n+1$ and $q(i) \in G(p\lceil i)$ for all $i < n+1$.

Note that in both cases $q(0) \in G(p\lceil 0) = G(\mathbf{0})$. If $p = 0$, we have that $\mathbf{0}Rq$ iff $Dom(q) = 1 = \{0\}$ and $q(0) \in G(\mathbf{0})$. Hence $\mathbf{0}R = G(\mathbf{0}) \neq \emptyset$.

Claim 1. *The relation R is monotone.*

Proof of the claim. Let p, q, and r be elements of P such that $p \subseteq q$ and pRr. We shall prove that there exists $s \in P$ such that qRs and $r \subseteq s$.

$$\begin{tikzcd}
q \arrow[r] & s \\
p \arrow[u, "\subseteq"] \arrow[r, "R"'] & r \arrow[u, "\subseteq"']
\end{tikzcd}$$

We consider two cases.

Case 1. p is a total function, i.e., $Dom(p) = \omega$.
As $p \subseteq q$, we have that $Dom(q) = \omega$. Hence $p = q$. In this case we simply put $s := r$.

Case 2. p is a partial finite function, i.e., $Dom(p) = n$ for some $n \in \omega$.
As pRr, we have that $Dom(r) = n+1$ and $r(i) \in G(p\lceil i)$ for all $i < n+1$. In particular $r(n) \in G(p\lceil n)$. Moreover, as $p \subseteq q$, we have that $n = Dom(p) \subseteq Dom(q)$. We then consider two subcases:

Subcase 2A. $Dom(q) = m$ for some $m \in \omega$.
Then $n \leq m$. We define $s \in P$ as follows. $Dom(s) = m + 1$ and

$$s(i) := \begin{cases} r(i) & \text{if } i \leq n \\ \in G(q\lceil i) & \text{if } n+1 \leq i \leq m. \end{cases}$$

It is clear that $r \subseteq s$. Furthermore, as pRr, $s\lceil n+1 = r$ and $p \subseteq q$, it is also clear that $s(i) \in G(q\lceil i)$ for $i = 0, 1, \ldots, m$. This means that qRs.

Subcase 2B. $Dom(q) = \omega$.
Then $s \in P$ is defined as follows. $Dom(s) = \omega$ and

$$s(i) := \begin{cases} r(i) & \text{if } i \leq n \\ \in G(q\lceil i) & \text{if } n+1 \leq i. \end{cases}$$

We note that the Principle of Countable Choice is applied here, because one simultaneously picks out an element from each set from the countably infinite list of non-empty sets $G(q\lceil i)$, where $i \in \omega$ and $n + 1 \leq i$.

It is then clear that $r \subseteq s$ and $s(i) \in G(q\lceil i)$ for all $i \in \omega$. Hence qRs. ■

Claim 2. *R is chain σ-continuous.*

Proof of the claim. Let C be a strictly increasing chain

$$p_0 \subset p_1 \subset \ldots \subset p_n \subset p_{n+1} \subset \ldots$$

of elements of P of type ω, and let $f : C \to P$ be a monotone mapping such that $p_n R f(p_{n+1})$ for all $n \in \omega$. The sets $Dom(p_n)$, $n \in \omega$, form a strictly increasing sequence of natural numbers. Applying the Arithmetic Induction Principle we get that $n \subseteq Dom(p_n)$, for all $n \in \omega$. Furthermore $Dom(f(p_n)) = Dom(p_n) + 1$, and

$(*)$ $\quad f(p_n)(i) \in G(p_n \lceil i)$ for all $i \leq Dom(p_n)$, $n \in \omega$.

As $f : C \to P$ is monotone, we also have that the functions $f(p_n)$, $n \in \omega$, form an increasing chain

$$f(p_0) \subseteq f(p_1) \subseteq \ldots \subseteq f(p_n) \subseteq f(p_{n+1}) \subseteq \ldots$$

of type $\leq \omega$. It follows that $p := \bigcup_{n\in\omega} p_n$ and $q := \bigcup_{n\in\omega} f(p_n)$ are well-defined functions and $Dom(p) = Dom(q) = \omega$. Consequently, $p, q \in P$ and, by $(*)$,

$$q(i) \in G(p\lceil i) \text{ for all } i \in \omega.$$

Hence pRq. This proves the claim. ■

Claim 3. *If $p \in P$ is a fixed-point of R, then $Dom(p) = \omega$.*

Proof of the claim. Suppose $Dom(p) = n$ for some $n \in \omega$. According to the definition of $\pi(p)$, the function $\pi(p)$ is defined at n. As pRp, the definition of R yields that p is defined at n, which is impossible. ■

We can now prove (3). The above remarks show that $(P, \subseteq)$ and R satisfy the assumptions of (4). It follows that R has a fixed-point $p^* \in P$. Furthermore, for every $q \in P$, if $qR \subseteq \leq q$, then $p^* \subseteq q$.

As p^*Rp^*, we have that $Dom(p^*) = \omega$ and $p^*(i) \in G(p^*\lceil i)$ for all $i \in \omega$. Putting $F := p^*$, we see that (3) holds.

Note that AC_ω is required in the proof of implication (4) ⇒ (3).

(1) ⇒ (4). This is Theorem 2.4. ■

It follows from the proof of Metatheorem 3.2 that in the presence of the Arithmetic Induction Principle, each of the conditions (1)–(3) is equivalent to the conjunction of AC_ω and Theorem 2.4. In particular, the Axiom of Dependent Choices (DC) is equivalent to the conjunction of the Principle of Countable Choice and Theorem 2.4. It is unknown whether Theorem 2.4 itself implies AC_ω. This fact would entail the effective equivalence of Theorem 2.4 with DC.

Example. We take a closer look at the proof of Baire's Theorem. This theorem is formulated here in the following form:

> *Let (X, d) be a complete metric space. Let $A_0, A_1, \ldots$ be a countable sequence of closed nowhere dense subsets of X. Then the union $\bigcup_{n\in\omega} A_n$ is a proper subset of X.*

PROOF. Let B be the set of all open balls in (X, d) with radius $< 1/2$. For each $n \in \omega$ we define

$$P_n := \{a \in B : a \cap A_n = \emptyset\}.$$

Since each set $X \setminus A_n$ is dense and open in X, it follows that $a \cap (X \setminus A_n) \neq \emptyset$ for all n and all open balls a. Hence $P_n \neq \emptyset$, for all n.

Let

$$A := \bigcup_{n\in\omega} \{n\} \times P_n.$$

Let r_a denote the radius of the ball a. We then define the relation $R \subseteq A \times A$ as follows: for $(m, a), (n, b) \in A$,

$$(m, a)R(n, b) \text{ iff } n = m + 1, a \supseteq b \text{ and } r_b < (1/2)r_a.$$

CLAIM. *R is serial.*

PROOF OF THE CLAIM. Let (m, a) be an element of A. Then $a \in P_m$, which means that $a \cap A_m = \emptyset$. As $a \cap (X \setminus A_{m+1})$ is non-empty, there exists an open ball b such that $b \subseteq a \cap (X \setminus A_{m+1}))$ and $r_b < (1/2)r_a$. So $(m, a)R(n, b)$. ■

We then select a ball $a \in P_0$. By the Axiom of Dependent Choices, there exists a mapping $f : \omega \to A$ such that $f(0) = (0, a)$ and $f(n)Rf(n + 1)$ for all $n \in \omega$. We may then write

$$f(n) = (n, a_n), \text{ for all } n.$$

The burden of the proof of the theorem rests on the definition of the above sequence of balls a_n, for all n. This definition requires DC.

Let x_n and r_a be the center and the radius of a_n, respectively, for all $n \in \omega$. The conclusion of the proof of the theorem is reached when one notices that

1. $\{x_n\}$ is a Cauchy sequence in (X, d).
2. Let x be the limit of $\{x_n\}$ in (X, d). Then $x \notin A_n$, for all $n \in \omega$.

Hence $\bigcup_{n\in\omega} A_n$ is a proper subset of X. ■

We shall discuss a certain property of relations, which is stronger than monotonicity, viz. the transfer property for upper bounds of chains. Monotonicity is a special case of the transfer property obtained by restricting the latter property to one-element chains.

Definition 3.3. Let $(P, \leq)$ be a poset and let $R \subseteq P \times P$ be a binary relation. We say that *upper bounds of directed sets in* $(P, \leq)$ *are transferable with respect to* R if the following holds:

> For every non-empty directed set D in $(P, \leq)$, for every monotone mapping $f : D \to P$ such that $xRf(x)$ for all $x \in D$, and for every upper bound a of D in $(P, \leq)$ there exists an element $b \in P$ such that b is an upper bound of the directed set $f[D]$ in $(P, \leq)$ and aRb. $\square$

In other words, under the adopted assumptions, if for every $x \in D$ the world x sees the world $f(x)$, then each upper bound of D sees a world b which is an upper bound of $f[D]$.

By confining the above property to certain types of directed subsets, further refinements of the transfer property are defined, as e.g. the *transferability of upper bounds of chains* and the *transferability of upper bounds of well-ordered chains*.

In the sequel we will be mainly concerned with chain-complete posets. Therefore the format of the transfer property will be restricted to the contexts in which inductive posets occur. But the results presented below also holds for the other types of posets we have introduced.

Corollary 3.4. *If* $(P, \leq)$ *is a chain-complete poset. Let* $R \subseteq P \times P$ *be a binary relation. The following conditions are equivalent*:

(1) *Upper bounds of arbitrary chains in* $(P, \leq)$ *are transferable with respect to* R.

(2) R *is monotone and* R *satisfies the following condition:*

> $(*)$ *For every non-empty chain* C *in* $(P, \leq)$, *for every monotone mapping* $f : C \to P$ *such that* $xRf(x)$ *for all* $x \in C$, *there exists an element* $b \in P$ *such that* $\sup(C)Rb$ *and* $\sup(f[C]) \leq b$.

The analogous fact holds for directed-complete posets and well-ordered chain-complete posets, respectively, with chains replaced by directed sets and well-ordered chains, respectively.

Proof. (1) $\Rightarrow$ (2). Assume (1) holds. We first show R is monotone. Let u, v, z be a triple of elements of P such that uRz and $u \leq v$. $C := \{u\}$ is a one-element chain and the mapping $f : C \to P$ given by $f(u) := z$ is monotone. Trivially $xRf(x)$ for all $x \in C$. As v is an upper bound of C, (1) implies that there exists $w \in P$ such that w is an upper bound of $f[C]$ and vRw (see the figure below). Hence $z \leq w$ and vRw. So R is monotone.

$$\begin{tikzcd}
v \arrow[r, "R"] & w \\
u \arrow[u, "\leq"] \arrow[r, "R"'] & z \arrow[u, "\leq"']
\end{tikzcd}$$

The second conjunct of (2) directly follows from (1), because $\sup(C)$ is trivially an upper bound of C.

(2) ⇒ (1). Assume (2) holds. Let C be a non-empty chain in $(P, \leq)$ and let $f : C \to P$ be a monotone mapping such that $xRf(x)$ for all $x \in C$. Let a be an upper bound of C. As $(P, \leq)$ is chain-complete, $\sup(C)$ exists and obviously $\sup(C) \leq a$. By (2) $(*)$ there exists $b_0 \in P$ such that $\sup(C) R b_0$ and $\sup(f[C]) \leq b_0$. Hence, by the monotonicity of R, there exists $b \in P$ such that aRb and $b_0 \leq b$ (see the figure below).

$$\begin{tikzcd}
a \arrow[r, "R"] & b \\
\sup(C) \arrow[u, "\leq"] \arrow[r, "R"'] & b_0 \arrow[u, "\leq"']
\end{tikzcd}$$

We then have that $\sup(f[C]) \leq b_0 \leq b$, which gives that b is an upper bound of the chain $f[C])$. So (1) holds. ■

If $\pi : P \to P$ is a monotone mapping (in the usual sense), then upper bounds of arbitrary chains are transferable with respect to the graph R_π. For let C be a non-empty chain in $(P, \leq)$. Assume $f : C \to P$ is a monotone mapping such that $xR_\pi f(x)$ for all $x \in C$ and let a be an upper bound of C. Evidently, $f(x) = \pi(x)$ for all $x \in C$. As $x \leq a$ for all $x \in C$, we obtain that $\pi(x) \leq \pi(a)$ for all $x \in C$, by the monotonicity of π. It follows that $b := \pi(a)$ is an upper bound of the chain $f[C]$ in $(P, \leq)$ and $aR_\pi b$.

We shall supplement the list of fixed-point theorems for relations with the following observation:

Theorem 3.5. *Assume the Axiom of Choice. Let* $(P, \leq)$ *be a chain-complete poset. Let* $R \subseteq P \times P$ *be a relation such that upper bounds of arbitrary chains in* $(P, \leq)$ *are transferable with respect to* R*. Suppose that the set* $\mathbf{0}R$ *is non-empty. Then* R *has a fixed-point* a^**. Furthermore,* a^* *can be assumed to have the property: for every* $y \in P$*, if* $yR \subseteq {\leq} y$*, then* $a^* \leq y$*.*

The analogous results hold for well-ordered chain-complete posets and directed-complete posets.

THE PROOF is omitted. ■

In comparison with Theorem 2.4, Theorem 3.5 assumes a stronger property of the poset $(P, \leq)$, viz. chain-completeness. In turn, the property of chain-σ-continuity of a relation is replaced by the transferability of upper bounds of chains.

The above theorem provides a general, order-theoretic setting of a half of the well-known back and forth method from model theory, namely of the forth part (see e.g. Czelakowski [2006]).

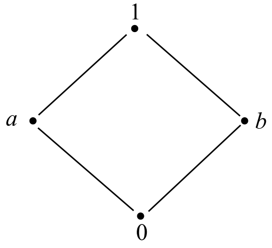
The proof, which much resembles the proof of Theorem 2.4, is omitted. It will be presented in another paper. It should be underlined that not the full strength of AC is required here but merely a stronger version of the Axiom of Dependent Choices, which is called the Ordinal Principle of Choice. (Set-theoretic aspects of Theorem 3.5 are not discussed here.)

NOTE. The fixed-point a^* need not be unambiguously defined: there may be many elements a^* satisfying the second statement of the above theorem. In order to better elucidate the above question, we introduce the following definition.

Let $(P, \leq, R)$ be an ordered frame. An element $a \in P$ is called a *strong fixed-point of* R (or a is a *strong reflexive point* of R) if aRa and $aR \subseteq \leq a$, i.e., aRz implies $z \leq a$ for all z.

It follows from Theorem 3.5 that each of the fixed-points a^* of R satisfying the last statement of the theorem is equal or smaller than all strong reflexive points of R. Indeed, suppose that a is a strong reflexive point of R. Then $aR \subseteq \leq a$. Consequently, $a^* \leq a$. But a^* itself need not be a strong reflexive point. We produce an appropriate example.

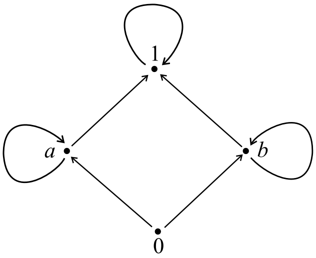
Let $P = \{0, a, b, 1\}$ be a four element poset with the order declared by $0 < a < 1$ and $0 < b < 1$. The elements a and b are incomparable — see the diagram below.



Moreover, let $R \subseteq P \times P$ be the relation defined by:

$$R := \{(0,a),(a,a),(a,1),(0,b),(b,b),(b,1),(1,1)\}$$

(see the diagram below).



We have:

(1) $(P,\leq)$ is an inductive poset.

(2) R is monotone.

(3) Upper bounds of chains in $(P,\leq)$ are transferable with respect to R.

(4) 1 is a strong reflexive point of R.

(5) a and b are incomparable reflexive points of R smaller than 1.

(6) a and b are not strong reflexive points but they satisfy the statement of Theorem 3.5.

(The proof of (2) is straightforward but tedious. Among 64 triples (x, y, z) (with repetitions) of elements of P only 17 of them satisfies the condition: xRz and $x \leq y$. Then in each such a case one finds w such that yRw and $z \leq w$.

To prove (3) one applies Corollary 3.4.(2) and examines non-empty chains C in $(P,\leq)$, viz. the subsets $\{0,a,1\}$, $\{0,b,1\}$, $\{0,a\}$, $\{0,b\}$, $\{a,1\}$, $\{b,1\}$, $\{0,1\}$, $\{0\}$, $\{a\}$, $\{b\}$, $\{1\}$, together with monotone mappings $f : C \to P$ such that $xRf(x)$ for all $x \in C$. There are 25 such situations. In each situation the statement of Corollary 3.4.(2) holds.)

If one slightly modifies the definition of R, namely one deletes the pair $(1,1)$ from R, the resulting relation is not monotone because for the triple $(1,b,1)$ it is the case that $bR1$ and $b \leq 1$, but there is no element w in P such that $1Rw$ and $1 \leq w$. This new relation does not possess strong reflexive points but a and b remain its reflexive points. □

The proof of the following result does not require the Axiom of Choice.

THEOREM 3.6. *Let $(P,\leq)$ be an inductive poset. Let $\pi : P \to P$ be a monotone mapping. Then π has the least fixed-point, i.e., there exists a^* in*

P such that $\pi(a^*) = a^*$ and for every $y \in P$, if $\pi(y) \leq y$, then $a^* \leq y$. An analogous result holds for directed-complete posets and well-ordered chain-complete posets, respectively.

PROOF. As π is monotone, the graph of π satisfies the hypotheses of Theorem 3.5. The Axiom of Choice is not needed here. ■

NOTES 3.7.

1. Theorem 2.6 proved in the previous section should not be confused with Theorem 3.6 — the latter is not a generalization of the former. The thesis of Theorem 3.6, viz, every *monotone* mapping $\pi : P \to P$ has a least fixed-point a^*, is indeed stronger that the thesis of Theorem 2.6 but Theorem 3.6 assumes *more* about the poset $(P, \leq)$ (it requires $(P, \leq)$ to be chain-complete and not merely chain-σ-complete). Thus in the narrower context formed by inductive posets, the hypothesis of σ-continuity of π, which is essential in Theorem 2.6, is dropped here altogether and replaced by the weaker condition of monotonicity.

The proof of Theorem 3.6 can be found in Chapter 8 of Davey and Priestley [2002] and in Moschovakis [1994]. Theorem 3.6 can also be found in Markowsky [1976], where in turn it is attributed to Bourbaki, based on some ideas of Zermelo.

2. In the contexts where families of related results for various notions of order completeness are presented one may use the notion of *Z-sets* in the sense of Wright, Wagner and Thatcher [1978]. They consider the following (meaningful) definition scheme "A poset $(P, \leq)$ is Z-inductive if it has a subposets C of Z-compact elements such that for every element a of P there is a Z-set $X \subseteq C$ such that $a = \sup(X)$", where the symbol Z ranges over such adjectives as "directed", "linear", "well-ordered", etc. Many theorems in the theory of posets differ only in their instantiations of Z. Similar phenomenon occurs when one considers such notions as Z-completeness of Z-continuity. Wright, Wagner and Thatcher abstracted out the essential common properties of the different instantiations of Z and proved common theorems within the resulting abstract framework. In this paper we do not make an explicit use of the above definition scheme. But some definitions and results of this paper can be uniformly formulated in terms of Z-sets as e.g. various instantiations of Definition 3.3 or Corollary 3.4, where Z ranges over the properties: *transferability of upper bounds of directed subsets, transferability of upper bounds of chains* and *transferability of well-ordered chains*. We shall devote more space to *Z-sets* in another paper devote to algebraic posets and the ways of defining various objects by means of *extending monotone* mappings from a poset $(P, \leq)$ to a directed-complete poset $(Q, \leq)$ to *order continuous* mappings from the algebraic completion of $(P, \leq)$ to $(Q, \leq)$. $\square$

4. The downward Loewenheim-Skolem-Tarski Theorem

As an illustration of the scope of Theorem 2.4 we outline here an alternative proof of the downward Loewenheim -Skolem-Tarski Theorem (LST, for short). This theorem belongs to model theory. LST states, roughly, that every infinite model A has an elementary submodel of any intermediate power between the cardinality of the language and of the cardinality of A. The standard proof of LST applies Indeterminate Arithmetic Recursion Principle (in whatever version) — see e.g. Chang and Keisler [1973, Theorem 3.1.6] The model-theoretic notions applied here are standard.

A language is a set being the union of threes sets: a set of relational symbols (predicates), a set of function symbols, and a set of constant symbols. (Constant symbols are often viewed as nullary function symbols.) If L is a language, then $For(L)$ denotes the set of first-order formulas of L.

Theorem 4.1 (Downward Loewenheim-Skolem-Tarski Theorem). *Let L be a language and let α and β be cardinal numbers such that $|For(L)| \leq \beta \leq \alpha$. Let A be a model for L of cardinality α. Then A has an elementary submodel of cardinality β. In fact, for any subset $X_0 \subseteq A$ of power $\leq \beta$, the model A has an elementary submodel of power β which contains X_0 as a subset of its universe.*

Proof. Let P be the family consisting of all subsets $X \subseteq A$ such that $|X| = \beta$ and $X_0 \subseteq X$. Evidently, P is non-empty because $|A| \geq \beta$. We have:

(A) *The family P, ordered by inclusion, is chain-σ-complete.*

Indeed, for any non-empty countable chain C (of type $\leq \omega$) of subsets of A such that $|X| = \beta$ for all $X \in C$, the union $\bigcup C$ has cardinality β. Furthermore, the set $\mathbf{0} := X_0$ is the least element of P. $\mathbf{0}$ is the supremum of the empty chain.

We define the following binary relation R on the poset $(P, \subseteq)$: for $X, Y \in P$,

> XRY iff $X \subseteq Y$ and for every formula $\Phi(x, x_1, \ldots, x_n) \in For(L)$ and any sequence $a_1, \ldots, a_n \in X$ (of length n) such that
>
> $$A \models (\exists x)\Phi[a_1, \ldots, a_n]$$

there exists $b \in Y$ such that

$$A \models \Phi[b, a_1, \ldots, a_n].$$

XRY thus says that the set Y includes X and furthermore, for each formula $\Phi(x, x_1, \ldots, x_n)$ and any n-tuple $a_1, \ldots, a_n \in X$ satisfying $(\exists x)\Phi$ in A, the set Y contains at least one $b \in A$ such that the $(n+1)$-tuple $b, a_1, \ldots, a_n$ satisfies $\Phi(x, x_1, \ldots, x_n)$ in A. The set Y may also contain some other elements of A but the cardinality of Y should not exceed β.

$(P, \subseteq, R)$ is an ordered frame. As $|For(L)| \leq \beta$, the crucial observation is that

(B) *The relation R is serial, i.e., for any $X \in P$ there exists $Y \in P$ such that XRY.*

Furthermore, the definition of R gives that for any $X, Y, X_1, X_2, Y \in P$:

(C) *If XRY then $X \subseteq Y$.*

(D) *$X_1 \subseteq X_2$ and X_2RY implies X_1RY.*

It follows from (B) and (C) that R is $\forall$-expansive. But, more interestingly,

(E) *The relation R is chain-σ-continuous (in the stronger sense).*

We first check that R is monotone in $(P, \subseteq)$. Assume $X, Y, Z \in P$ so that $X \subseteq Y$ and XRZ. Evidently $Y \cup Z$ belongs to P. By (B) and (C), there exists a set W such that $Y \cup ZRW$ and $Y \cup Z \subseteq W$. As $Y \cup ZRW$, (D) gives that YRW. Evidently $Z \subseteq W$. This proves monotonicity.

Suppose we are given two non-empty chains of elements of P,

$$Y_0 \subseteq Y_1 \subseteq \ldots \subseteq Y_n \subseteq Y_{n+1} \subseteq \ldots \text{ and } Z_0 \subseteq Z_1 \subseteq \ldots \subseteq Z_n \subseteq Z_{n+1} \subseteq \ldots,$$

the first chain of type ω , such that Y_nRZ_n for all n. It is clear that the sets $\bigcup_{n\in\omega} Y_n$ and $\bigcup_{n\in\omega} Z_n$ belong to P. We claim that

$$\bigcup_{n\in\omega} Y_n \, R \bigcup_{n\in\omega} Z_n.$$

The fact that Y_nRZ_n for all n implies that $Y_n \subseteq Z_n$ for all n. Hence $\bigcup_{n\in\omega} Y_n \subseteq \bigcup_{n\in\omega} Z_n$.

Let $Y := \bigcup_{n\in\omega} Y_n$ and $Z := \bigcup_{n\in\omega} Z_n$. Let $\Phi(x, x_1, \ldots, x_k)$ be a formula in $For(L)$ and let $a_1, \ldots, a_k$ be a sequence of elements of Y (of length k) such that

$$A \models (\exists x)\Phi[a_1, \ldots, a_k].$$

There exists $n \in \omega$ such that $a_1, \ldots, a_k \in Y_n$. Since $Y_n R Z_n$, there exists $b \in Z_n$ such that

$$A \models \Phi[b, a_1, \ldots, a_n].$$

Consequently, there exists $b \in Z$ such that

$$A \models \Phi[b, a_1, \ldots, a_n].$$

As $\Phi(x, x_1, \ldots, x_k)$ is an arbitrary formula in $For(L)$ and $a_1, \ldots, a_k$ are arbitrary elements of Y, this proves that YRZ. So R is σ-continuous.

As R is serial, then trivially the set $\mathbf{0}R$ is non-empty. The system $(P, \subseteq, R)$ thus satisfies the assumptions of Theorem 2.4. It follows that the relation R has a fixed-point in $(P, \subseteq)$, say B. It is then easy to verify that B is a universe of an elementary submodel B of A. As B belongs to P, the cardinality of B is equal to β. ■

References

Berman, J., Blok, W., [1989], 'Generalizations of Tarski's fixed-point theorem for order varieties of complete meet semilattices', *Order*, 5(4): 381–392.

Cai, J., Paige, R., [1992], 'Languages polynomial in the input plus output', in *Second International Conference on Algebraic Methodology and Software Technology (AMAST 91)*, Springer Verlag, London, pp. 287–300.

Chang, C.C., Keisler, H.J., [1973], *Model Theory*, North-Holland and American Elsevier, Amsterdam–London–New York.

Czelakowski, J., [2006], 'Fixed-points for relations and the back and forth method', *Bulletin of the Section of Logic*, 35(2/3): 63–71.

Davey, B.A., Priestley, H., [2002] *Introduction to Lattices and Order*, 2nd ed., Cambridge University Press, Cambridge.

Desharnais, J., Möller, B., [2005], 'Least reflexive points of relations', *Higher-Order and Symbolic Computation*, 18: 51–77.

Dugundji, J., Granas, A., [1982], *Fixed Point Theory*, Monografie Matematyczne, vol. 61, PWN, Warsaw.

Fujimoto, T., [1984], 'An extension of Tarski's fixed point theorem and its application to isotone complementarity problems', *Mathematical Programming*, 28: 116–118.

Goebel, K., Kirk, W.A., [1990], *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge.

Gunter, C.A., Scott, D.S., [1990], 'Semantic domains', in Van Leeuwen, J. (Managing Editor), *Handbook of Theoretical Computer Science*, The MIT Press/Elsevier, Amsterdam, New York-Oxford-Tokyo/Cambridge, Massachusetts, pp. 634–674.

Kirk, W.A., Sims, B. (eds.), [2001], *Handbook of Metric Fixed Point Theory*, Kluwer, Dordrecht, Boston–London.

Kleene, S.C., [1952], *Introduction to Metamathematics*, Van Nostrand.

Kunen, K., [1999], *Set Theory. An Introduction to Independence Proofs*, Elsevier, Amsterdam–Lausanne–New York.

Markowsky, G., [1976], 'Chain-complete posets and directed sets with applications', *Algebra Universalis*, 6: 53–68.

Moschovakis, Y.N., [1994], *Notes on Set Theory*, Springer-Verlag, New York–Berlin.

Tarski, A., [1955], 'A lattice-theoretical fixpoint theorem and its applications', *Pacific Journal of Mathematics*, 5: 285–309.

Wright, J., Wagner, E., Thatcher, J., [1978], 'A uniform approach to inductive posets and inductive closure', *Theoretical Computer Science*, 7: 57–77.

Janusz Czelakowski
Institute of Mathematics and Informatics
Opole University, Oleska 48, 45-052 Opole,
Poland
`jczel@math.uni.opole.pl`
and
Institute of Mathematics and Physics
Opole University of Technology
Luboszycka 3, 45-036 Opole

A. Avron
J. Ben-Naim
B. Konikowska

Processing Information from a Set of Sources

Abstract. We introduce a general framework for solving the problem of a computer collecting and combining information from various sources. Unlike previous approaches to this problem, in our framework the sources are allowed to provide information about complex formulae too. This is enabled by the use of a new tool — non-deterministic logical matrices. We also consider several alternative plausible assumptions concerning the framework which lead to various logics. We provide strongly sound and complete proof systems for all the basic logics induced in this way.

Keywords: Information processing, multiple sources, non-deterministic matrices, non-classical logics, paraconsistency.

1. Introduction

The idea considered in this paper has originated from Belnap, whose famous four-valued logic [9, 8][1] stemmed from considering the problem of a computer collecting and combining information from various sources. Later Belnap's approach was extended by Carnielli and Lima-Marques in their *society semantics* [10] to consider various information collecting and processing strategies applied by the computer (or some other agent). However, both works considered just the simple case of sources providing information only about *atomic* formulas of some logical language (which corresponds to the case of simple relational databases). Unfortunately, this does not capture all the situations encountered in practice, for e.g. knowledge bases and disjunctive databases can also provide information about complex formulas. Accordingly, in this paper we extend the previous approaches in an essential way by allowing the sources to provide information about complex formulae too. This is enabled by the use of a new tool — non-deterministic logical matrices (*Nmatrices* — see [5, 6, 4]), which is necessary in view of the fact that ordinary logical matrices are unable to capture the above general case.

The structure of the paper is as follows. In Section 2 we describe our general framework for processing information coming from different sources, as well as several various plausible assumptions concerning it, leading to

---

[1] Actually, this logic should be called Dunn-Belnap logic, since it was originally introduced by Dunn [12].

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

important special cases. In Section 3 we investigate the four basic logics obtained by adopting the simplest such assumption, according to which a processor accepts any proposition declared true by one of its sources (even if this leads to contradictions). Two of these logics (Dunn-Belnap logic and the basic paraconsistent 3-valued logic) are well-known. The two others are new. Section 4 shortly investigates an alternative strategy, in which a processor initially accepts a proposition only if *all* its sources declare it to be true. This strategy also leads to a famous logic: Kleene's strong 3-valued logic. In Section 5 we introduce calculi of sequents for all these logics, and prove their strong soundness and completeness, as well as a strong version of the admissibility of the cut rule in them. Finally, in Section 6 we outline directions for future research.[2]

2. The framework

2.1. Informal description

Assume we have a framework for information collecting and processing, which consists of a set of information sources S and a processor P. The sources provide information about formulas of the classical propositional logic L_C (which we take here to be based on the connectives $\{\neg, \vee, \wedge\}$). We assume that for each such formula φ, a source $s \in S$ can say that φ is true, that φ is false, or that it has no knowledge about φ. Thus, every source defines some (possibly partial) valuation (using the two classical truth-values). Note that we do not assume here that the valuations must be homomorphisms of formulas into logical values. In turn, the processor collects information from the sources, combines it according to some strategy, processes the result and finally defines its resulting combined valuation (denoted in the sequel by d) of formulas in L_C. Analogously as in case of sources, the processor's valuation need not be a homomorphism either.

Clearly, for any formula $\varphi \in L_C$, the processor can encounter at the information collecting stage four possible situations concerning the information it gets from the sources:

- It has information that φ is true but no information that φ is false
- It has information that φ is false but no information that φ is true
- It has both information that φ is true and information that φ is false
- It has no information on φ at all

---

[2] The first short description of our framework was given in [7]. In that paper, the basic proof system used here was derived using some general method, yielding a roundabout proof of a rather weak form of the soundness and completeness theorem for that system.

In view of the above, a natural logical domain for the considered framework features four logical values corresponding to the four cases above, which are usually denoted[3]

$$\mathbf{t} = \{1\},\ \mathbf{f} = \{0\},\ \top = \{0,1\},\ \bot = \emptyset,$$

Here 1 and 0 represent the classical logical values of *true* and *false* (respectively), and so $\top$ represents inconsistent information, while $\bot$ denotes absence of information. Among these four truth values, we take as designated $\mathbf{t}$ and $\top$ — the truth values whose assignment to a formula φ means that the processor has information that φ is true (even though it might also have information that φ is false). This represents the so-called *weak semantics*. Another possible option could have been to consider *strong semantics*, whereby the only designated value is $\mathbf{t}$, which means φ is deemed satisfied if the processor has information that φ is true, but has no information that φ is false. However, the consequence relation induced by the strong semantics can be simulated by the weaker one employed here (see Subsection 2.4).

2.2. Variants of the model

The general model introduced above has many variants, corresponding to various assumptions on the kind of information provided by the sources and the strategy used by the processor to combine it. Within this general framework, we can classify the resulting system under four kinds of criteria:

1. Behavior of each source.
2. Behavior of the whole set of information sources.
3. Procedure for collecting information from the sources.
4. Procedure for processing the collected information.

Exemplary basic assumptions concerning those criteria are listed below.

2.2.1. Behavior of each source

i) Scope of information provided by a source:

   (a) It provides information about all propositions (*complete knowledge*), i.e., assigns either 0 or 1 to each formula.

---

[3] Especially in the literature on bilattices — see e.g. [16, 15]. An alternate notation (see e.g. [13]) uses **b** (for *both*) to denote $\{0,1\}$, **n** (for *neither*) to denote $\emptyset$, **T** to denote $\{1\}$, and **F** to denote $\{0\}$.

   (b) It provides information about some propositions only (*partial knowledge*), i.e., assigns either 0 or 1 to some formulas only (with no particular logical restrictions).
   (c) It provides information only about (some/all) atomic propositions (*partial/complete atomic knowledge*).

ii) Logical characteristics of a source:
The assignment of values by a source can be restricted by certain logical constraints. For example, we could demand that:

   (a) For any formulas A, B such that $A \sim B$ (where $\sim$ denotes classical equivalence), each source should assign the same value to A and B. Instead of classical equivalence, other types of logical equivalence, more plausible from the implementation viewpoint, can also be considered here.
   (b) The sources should be *classically coherent*: i.e., the partial valuation provided by each of the sources should be extendable to a full classical valuation.
   (c) The sources should be *classically closed*, meaning that if φ classically follows from Γ, then any source which assigns 1 to all formulas in Γ should assign 1 to φ too, and that 1 (0) is assigned to $\neg\varphi$ iff 0 (1) is assigned to φ.

2.2.2. Behavior of the whole set of information sources

We may assume that, e.g.:

1. For each atomic proposition F, there is at least one source which provides information about F.
2. For an arbitrary proposition F, there is at least one source which provides information about F.

2.2.3. Procedure for collecting information from the sources

The processor can use various strategies in combining information from the sources. Thus it can accept a formula ψ as *true* (*false*) whenever:

Existential strategy: At least one source assigns ψ the value 1 (0). Note that in this case there is a possibility of assigning both 1 and 0 to the same formula. In such a case, the processor uses the truth value $\top$.

Note 2.1. Since a source might assign *at most one* classical truth-value to a formula, we might identify a source valuation with a *total* function from $\mathcal{F}$ to $\{\mathbf{t}, \mathbf{f}, \bot\}$ (and even change our definition of a source valuation accordingly). Hence we may view the sources as using three out of the four truth values in our many-valued framework — namely, $\mathbf{t}, \mathbf{f}$ and $\bot$, but not $\top$. However, in a more complex or hierarchic framework, a processor P provided with information by some sources could in turn supply information to another processor P' by passing to P' the values it has assigned to formulas (based on the information from its sources). In such a framework, also the sources may use all of the processor's truth-values (including $\top$). This idea has indeed been taken up and investigated by Y. Shramko and H. Wansing in [19, 20]. However, in this paper we examine just a single "information layer".

Universal strategy: All the sources assign ψ the value 1 (0). Note that in this case the processor might assign no value to a formula even if its sources are of the "know all" type. In other words, the processor might use $\bot$ even if the sources assign some classical logical value in $\{0, 1\}$ to every formula (and so implicitly use $\mathbf{t}$ and $\mathbf{f}$ only).

Unanimous voting strategy: Some sources assign ψ the value 1 (0), and no source assigns 0 (1). This amounts to the *universal* policy, but with "all sources" applied only to the sources which give a definite answer.

Preferred sources strategies: Each of the three preceding strategies can be applied using a preferred set of sources (determined by ψ) rather than the whole set of sources.

2.2.4. Procedure for information processing

After collecting the direct information from the sources, the processor processes that information to define its own valuation d of formulas in L_C. We assume that during that stage the processor derives from the above direct information at least the most basic new information provided by an inspection of the truth tables of classical logic. By this we mean that

1. If the processor has already assigned $a_{i_1}, \ldots, a_{i_k}$ to $\varphi_{i_1}, \ldots, \varphi_{i_k}$ (respectively), and according to the classical truth tables the truth-value of $\diamond(\varphi_1, \ldots, \varphi_n)$ should be b ($b \in \{0, 1\}$) whenever for every $1 \leq l \leq k$ the truth-value of φ_{i_l} is a_{i_l} , then the processor assigns b to $\diamond(\varphi_1, \ldots, \varphi_n)$ (For example: If φ is assigned 1 then $\varphi \vee \psi$ is also assigned 1).
2. If the processor has already assigned $b \in \{0, 1\}$ to $\diamond(\varphi_1, \ldots, \varphi_n)$, and according to the classical truth tables the truth-value of $\diamond(\varphi_1, \ldots, \varphi_n)$ can be b only if the truth-value of φ_{i_0} is a (where $1 \leq i_0 \leq n$), then the processor assigns a to φ_{i_0} (For example: If $\varphi \vee \psi$ is assigned 0 then both φ and ψ are assigned 0, by two applications of this principle).

Note that these two principles might again lead to 0 and 1 being both assigned to the same formula. Note also that a stronger assumption on the processor's procedure for information processing might have been that it fully respects *everything* dictated by the truth tables of classical logic (For example: If ψ is assigned 0, and $\varphi \vee \psi$ is assigned 1, then φ is assigned 1). This possibility is not investigated here.

2.3. Formal definitions

The source-processor framework is formalized as follows:[4]

DEFINITION 2.1. Let $\mathcal{A}$ and $\mathcal{F}$ be the set of all atomic formulas and the set of all formulas of the language L_C of propositional classical logic, respectively.

- By a *source valuation* we mean a partial function $s : \mathcal{F} \to \{0, 1\}$.
- By a *processor valuation* we mean a function $v : \mathcal{F} \to \mathcal{P}(\{0, 1\})$.
- By a *source-processor structure* we mean a tuple $\mathcal{S} = \langle S, g, d\rangle$, where S is a non-empty set of source valuations, g is an arbitrary processor valuation, and d is a processor valuation satisfying the following conditions:

  (d0) $g(\varphi) \subseteq d(\varphi)$ for every formula φ;

  (d1) $0 \in d(\neg\varphi)$ iff $1 \in d(\varphi)$;

  (d2) $1 \in d(\neg\varphi)$ iff $0 \in d(\varphi)$;

  (d3) $1 \in d(\varphi \vee \psi)$ if $1 \in d(\varphi)$ or $1 \in d(\psi)$;

  (d4) $0 \in d(\varphi \vee \psi)$ iff $0 \in d(\varphi)$ and $0 \in d(\psi)$;

  (d5) $1 \in d(\varphi \wedge \psi)$ iff $1 \in d(\varphi)$ and $1 \in d(\psi)$;

  (d6) $0 \in d(\varphi \wedge \psi)$ if $0 \in d(\varphi)$ or $0 \in d(\psi)$.

  (d1)–(d6) are called the *standard integrity conditions* for L_C.

NOTE 2.2. (d1)–(d6) are the rules which correspond to our above minimal assumptions concerning the processor's procedure for information processing. Note that the converses of (d3) and (d6) do *not* hold. For example, since the sources can provide information about complex formulas, the processor might e.g. be informed that $\varphi \vee \psi$ is true without being told either that φ is true or that ψ is true. As from the truth of $\varphi \vee \psi$ we cannot conclude either the truth of φ or the truth of ψ, this means that the processor cannot ascribe 1 to any of these formulas based on the information that $\varphi \vee \psi$ is true — which is why we could not assume that the converse of (d3) holds.

DEFINITION 2.2. A source-processor structure $\langle S, g, d\rangle$ for L_C is called *standard*, if d is the minimal processor valuation satisfying conditions (d0)–(d6).

---

[4] This section has benefited from discussions with David Makinson.

Each source-processor structure $\mathcal{S} = \langle S, g, d \rangle$ can be seen as a representation of an instance I of the source-processor framework defined informally in Subsection 2.1, with the two being related as follows:

- S is the set of source valuations defined by the sources present in the instance I, where, for each $s \in S$, $s(\varphi) = 1$ iff φ is true according to source s, and $s(\varphi) = 0$ iff φ is false according to source s;
- g represents the global information collected by the processor directly from the sources, i.e., $1 \in g(\varphi)$ (resp. $0 \in g(\varphi)$) iff, after information collecting, φ is accepted by the processor as true (resp. false);
- d represents the information derived by the processor from g during the information processing stage, i.e., $1 \in d(\varphi)$ (resp. $0 \in d(\varphi)$) iff, after processing the global information in g, the processor concludes that φ is true (resp. false).

From the viewpoint of the information collecting strategy, in this paper we consider the following two basic types of source-processor structures:

Definition 2.3. A source-processor structure $\mathcal{S} = \langle S, g, d \rangle$ is called:

- *existential*, iff for any $\varphi \in \mathcal{F}$,

$$1 \in g(\varphi) \text{ iff } \exists\, s \in S.s(\varphi) = 1 \quad \text{and} \quad 0 \in g(\varphi) \text{ iff } \exists\, s \in S.s(\varphi) = 0$$

- *universal* iff for any $\varphi \in \mathcal{F}$,

$$1 \in g(\varphi) \text{ iff } \forall\, s \in S.s(\varphi) = 1 \quad \text{and} \quad 0 \in g(\varphi) \text{ iff } \forall\, s \in S.s(\varphi) = 0$$

Next we turn to the *logics* induced by source-processor structures. Each such structure $\mathcal{S} = \langle S, g, d \rangle$ naturally generates a satisfaction relation on the formulas in $\mathcal{F}$ (determined by the final processor valuation d):

Definition 2.4. Let $\mathcal{S} = \langle S, g, d \rangle$ be a source-processor structure. Then $\mathcal{S}$ *satisfies* (is a *model* of):

- a formula $\varphi \in \mathcal{F}$, in symbols $\models_{\mathcal{S}} \varphi$, iff $1 \in d(\varphi)$.
- a set of formulas $F \subset \mathcal{F}$, in symbols $\models_{\mathcal{S}} F$, iff for any $\varphi \in F$, $\models_{\mathcal{S}} \varphi$.

Accordingly, each source-processor structure or, more generally, a class of source-processor structures, induces the corresponding consequence relation:

Definition 2.5. Let $\mathcal{J}$ be a class of source-processor structures. The *consequence relation* induced by $\mathcal{J}$ is the relation $\vdash_{\mathcal{J}}$ on $\mathcal{P}(\mathcal{F}) \times \mathcal{F}$ such that $T \vdash_{\mathcal{J}} \varphi$ if every $\mathcal{S} \in \mathcal{J}$ which is a model of T is also a model of φ.

2.4. The need for using sequents

In the context of source-processor structures, the expressive power of *formulas* of L_C is too weak. Thus there is no way to express that a certain formula φ is *not true* (meaning that $1 \notin d(\varphi)$). In the classical framework the fact that φ is not true is equivalent to the truth of $\neg\varphi$. However, in the present context the truth of $\neg\varphi$ means only that $0 \in d(\varphi)$, and this neither implies nor is implied by $1 \notin d(\varphi)$. Similarly, there is no way to express *disjunctive knowledge* of the form "one of the sentences φ and ψ is known to be true" (meaning that either $1 \in d(\varphi)$ or $1 \in d(\psi)$), because it is possible that $1 \in d(\varphi \vee \psi)$ but neither $1 \in d(\varphi)$ nor $1 \in d(\psi)$. These problems can be overcome by using Gentzen-type *sequents* for expressing these two types of knowledge that cannot be expressed directly in the language. The idea is that given a source-processor structure $\langle S, g, d\rangle$, a sequent $\varphi_1, \dots, \varphi_n \Rightarrow \psi_1, \dots, \psi_k$ expresses the information that either $1 \notin d(\varphi_1)$, or $1 \notin d(\varphi_2)$, or ... or $1 \notin d(\varphi_n)$, or $1 \in d(\psi_1)$, or ... or $1 \in d(\psi_k)$.

Another shortcoming of L_C is that it does not possess any implication connective that corresponds to the intended consequence relation. Again this problem is (essentially) overcome by using sequents, since sequents provide non-nestable ("first-degree", in the terminology of [2]) version of implication.

Finally, in addition to the considerable enhancement of the expressive power of our language provided by sequents, the latter will be also used in their usual role – to built sequent-based Gentzen calculi which offer an excellent mechanism for reasoning about information.

The notions of model and satisfaction, and the corresponding consequence relations are extended to the language of sequents in a straightforward way:

DEFINITION 2.6.

- A *sequent* is a structure of the form $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite *sets* of formulas. We denote by Seq the set of all sequents in the language L_C.
- Let $\mathcal{S} = \langle S, g, d\rangle$ be a source-processor structure. $\mathcal{S}$ *satisfies* (is a *model* of) a sequent $\Sigma = \Gamma \Rightarrow \Delta$, in symbols $\models_{\mathcal{S}} \Sigma$, iff either $\mathcal{S}$ is a model of some formula in Δ, or it is not a model of some formula in Γ.
- Let $\mathcal{J}$ be a class of source-processor structures. The *sequent consequence relation* induced by $\mathcal{J}$ is the relation $\vdash_{\mathcal{J}}$ on $\mathcal{P}(Seq) \times Seq$ s.t. $Q \vdash_{\mathcal{J}} \Sigma$ if every $\mathcal{S} \in \mathcal{J}$ which is a model of Q is also a model of Σ.

NOTE 2.3. It can easily be seen that if Γ is a finite subset of $\mathcal{F}$, and φ is a formula in $\mathcal{F}$, then $\Gamma \vdash_{\mathcal{J}} \varphi$ iff $\vdash_{\mathcal{J}} \Gamma \Rightarrow \varphi$ iff $\{\Rightarrow \psi \mid \psi \in \Gamma\} \vdash_{\mathcal{J}} \Rightarrow \varphi$. Hence the sequent consequence relation $\vdash_{\mathcal{J}}$ can be seen as an extension of the formula consequence relation $\vdash_{\mathcal{J}}$ defined above (Definition 2.5). This justifies the use of the same symbol to denote both.

NOTE 2.4. Given a source-processor structure $\langle S, g, d\rangle$ and a formula φ, every known basic fact about $d(\varphi)$ can be expressed by sequents as follows:

- $1 \in d(\varphi)$ iff $\models_{\mathcal{S}} \Rightarrow \varphi$
- $1 \notin d(\varphi)$ iff $\models_{\mathcal{S}} \varphi \Rightarrow$
- $0 \in d(\varphi)$ iff $\models_{\mathcal{S}} \Rightarrow \neg\varphi$
- $0 \notin d(\varphi)$ iff $\models_{\mathcal{S}} \neg\varphi \Rightarrow$

One corollary of this fact is that, given a class of source-processor structures $\mathcal{J}$, a formula φ follows from a set T of formulas according to the strong semantics (see Subsection 2.1) iff both $G(T) \vdash_{\mathcal{J}} \Rightarrow \varphi$ and $G(T) \vdash_{\mathcal{J}} \neg\varphi \Rightarrow$, where $G(T) = \{\Rightarrow \psi \mid \psi \in T\} \cup \{\neg\psi \Rightarrow \mid \psi \in T\}$. Hence the consequence relation induced by the strong semantics can be simulated by the weak one investigated in this paper.

3. Existential strategy for standard structures

In this section we assume that the existential strategy is adopted, and investigate under this assumption certain basic variants of standard source-processor structures for L_C (shortly referred to as "standard structures"). We shall consider the following four basic scenarios — the first two corresponding to well-known logics, and the other two new.

I **Dunn-Belnap's logic:** the sources provide information about atomic formulas only, but not necessarily about all of them;

II **D'Ottaviano and da Costa's basic paraconsistent logic:** Like the preceding case, but the sources taken together are required to provide some information about all atomic formulas.

III **The most general source-processor logic:** The sources provide information about arbitrary formulas, both atomic and composed ones, but not necessarily about all of them.

IV **General source-processor logic with complete information:** As the preceding case, but the sources taken together are required to provide some information about all atomic formulas.

To handle the new logics arising out of the last two cases, we need a generalization (see Note 3.2) of the notion of an ordinary logical matrix, namely non-deterministic matrices (shortly: Nmatrices) introduced in [5, 6]:

Definition 3.1.

1. A *non-deterministic matrix* (*Nmatrix* for short) for a propositional language $\mathcal{L}$ is a tuple $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:
   (a) $\mathcal{V}$ is a non-empty set of *truth values.*
   (b) $\mathcal{D}$ is a non-empty proper subset of $\mathcal{V}$ (the "designated values").
   (c) For every n-ary connective $\diamond$ of $\mathcal{L}$, $\mathcal{O}$ includes a corresponding n-ary function $\widetilde{\diamond}$ from $\mathcal{V}^n$ to $2^{\mathcal{V}} - \{\emptyset\}$.
2. Let $\mathcal{F}$ be the set of formulas of $\mathcal{L}$. A *(legal) valuation* in an Nmatrix $\mathcal{M}$ is a function $v : \mathcal{F} \to \mathcal{V}$ that satisfies the following condition for every n-ary connective $\diamond$ of $\mathcal{L}$ and $\psi_1, \dots, \psi_n \in \mathcal{L}$:

$$v(\diamond(\psi_1, \dots, \psi_n)) \in \widetilde{\diamond}(v(\psi_1), \dots, v(\psi_n))$$

3. A valuation v in an Nmatrix $\mathcal{M}$ is a:
   - *model* of (*satisfies*) a formula ψ in $\mathcal{M}$ ($v \models^{\mathcal{M}} \psi$) if $v(\psi) \in \mathcal{D}$.
   - *model* of a set $T \subseteq \mathcal{F}$ in $\mathcal{M}$ ($v \models^{\mathcal{M}} T$) if $v \models^{\mathcal{M}} \psi$ for all $\psi \in T$.
   - *model* of a sequent $\Sigma = \Gamma \Rightarrow \Delta$ ($v \models^{\mathcal{M}} \Sigma$) iff either $v \models^{\mathcal{M}} \psi$ for some $\psi \in \Delta$, or $v \not\models^{\mathcal{M}} \psi$ for some $\psi \in \Gamma$.
4. The *formula consequence relation* induced by the Nmatrix $\mathcal{M}$ (denoted by $\vdash_{\mathcal{M}}$) is defined by: $T \vdash_{\mathcal{M}} \varphi$ if every model of T in $\mathcal{M}$ is also a model of φ. The corresponding *sequent consequence relation* induced by $\mathcal{M}$ (also denoted by $\vdash_{\mathcal{M}}$) is defined similarly (compare Definition 2.6).

Note 3.1. Again, we have (see Note 2.3) that for every Nmatrix $\mathcal{M}$, every finite subset $\Gamma \subseteq \mathcal{F}$, and every formula $\varphi \in \mathcal{F}$, $\Gamma \vdash_{\mathcal{M}} \varphi$ iff $\vdash_{\mathcal{M}} \Gamma \Rightarrow \varphi$ iff $\{\Rightarrow \psi \mid \psi \in \Gamma\} \vdash_{\mathcal{M}} \Rightarrow \varphi$.

Note 3.2. An ordinary (deterministic) multiple-valued matrix can be seen as a special case of an Nmatrix, in which the interpretations of the connectives always return singletons. Accordingly, we shall identify ordinary matrices with this special case of Nmatrices (identifying a truth-value a with the singleton $\{a\}$).

3.1. Dunn-Belnap's logic

The first case we examine is when the sources provide (possibly incomplete) information about atomic formulas only, and the processor uses the existential strategy to combine the direct information from the sources and obtain the global information g. We shall show that the logic induced by this class of structures coincides with Dunn-Belnap's four-valued logic ([12, 9, 8]). This logic is induced by the following four-valued matrix $\mathcal{M}_B^4 = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where $\mathcal{V} = \{\mathbf{f}, \bot, \top, \mathbf{t}\}, \mathcal{D} = \{\top, \mathbf{t}\}, \mathcal{O} = \{\tilde{\neg}, \tilde{\vee}, \tilde{\wedge}\}$, and the interpretations of the connectives are given by the following tables (see Note 3.2):

$\tilde{\vee}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\bot$	$\bot$	$\bot$	$\mathbf{t}$	$\mathbf{t}$
$\top$	$\top$	$\mathbf{t}$	$\top$	$\mathbf{t}$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$

$\tilde{\wedge}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$
$\bot$	$\mathbf{f}$	$\bot$	$\mathbf{f}$	$\bot$
$\top$	$\mathbf{f}$	$\mathbf{f}$	$\top$	$\top$
$\mathbf{t}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$

$\tilde{\neg}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
	$\mathbf{t}$	$\bot$	$\top$	$\mathbf{f}$

Recalling what $\mathbf{f}, \bot, \top, \mathbf{t}$ stand for, these tables are best understood using the following well-known equivalent representation of $\mathcal{M}_B^4$:

Definition 3.2. Let $v_0 : \mathcal{A} \to \mathcal{P}(\{0,1\})$.

- The Belnap extension of v_0 is the function $v : \mathcal{F} \to \mathcal{P}(\{0,1\})$ defined inductively as follows:
  (b0) $v(p) = v_0(p)$ for $p \in \mathcal{A}$;
  (b1) If $1 \in v(\varphi)$, then $0 \in v(\neg\varphi)$;
  (b2) If $0 \in v(\varphi)$, then $1 \in v(\neg\varphi)$;
  (b3) If $1 \in v(\varphi)$ or $1 \in v(\psi)$, then $1 \in v(\varphi \vee \psi)$;
  (b4) If $0 \in v(\varphi)$ and $0 \in v(\psi)$, then $0 \in v(\varphi \vee \psi)$;
  (b5) If $1 \in v(\varphi)$ and $1 \in v(\psi)$, then $1 \in v(\varphi \wedge \psi)$;
  (b6) If $0 \in v(\varphi)$ or $0 \in v(\psi)$, then $0 \in v(\varphi \wedge \psi)$.
- A Belnap valuation is a function $v : \mathcal{F} \to \mathcal{P}(\{0,1\})$ being a Belnap extension of some valuation $v_0 : \mathcal{A} \to \mathcal{P}(\{0,1\})$. The set of all Belnap valuations will be denoted by $\mathcal{V}(\mathcal{M}_B^4)$.
- A *Belnap model* of $\Gamma \subseteq \mathcal{F}$ is any $v \in \mathcal{V}(\mathcal{M}_B^4)$ such that $\forall \varphi \in \Gamma . 1 \in v(\varphi)$ (note this is equivalent to taking $\mathbf{t}$ and $\top$ as the designated values).
- The Belnap formula consequence relation and the Belnap sequent consequence relation (both denoted in the sequel by $\vdash_{\mathcal{M}_B^4}$) are defined based on the notion of a Belnap model in the usual way (see Definitions 2.5, 2.6, and the end of Definition 3.1. Since ordinary matrices are a special type of Nmatrices, $\vdash_{\mathcal{M}_B^4}$ is actually an instance of the latter).

Lemma 3.1. *Each Belnap valuation satisfies the converses of* (b1)–(b6).

Proof. In the inductive process of extending $v_0 : \mathcal{A} \to \mathcal{P}(\{0, 1\})$ to a Belnap valuation v, the inclusion of 0 or 1 in the value of a composed formula can be due to exactly one of the rules (b1)–(b6). Hence the result. ■

Definition 3.3. Let $\mathcal{EA}$ denote the class of standard source-processor structures $\mathcal{S} = \langle S, g, d\rangle$, where each $s \in S$ is undefined outside $\mathcal{A}$ (i.e., the sources provide information about atomic formulas only), and the processor uses the existential strategy to obtain g out of the valuations in S.

Lemma 3.2. *We have the following correspondence between* $\mathcal{M}^4_B$ *and* $\mathcal{EA}$*:*

(1) $\forall v \in \mathcal{V}(\mathcal{M}^4_B)\ \exists \mathcal{S} \in \mathcal{EA}\ \exists S\ \exists g.\mathcal{S} = \langle S, g, v\rangle$

(2) $\forall \mathcal{S} \in \mathcal{EA}\ \forall S\ \forall g\ \forall d.\mathcal{S} = \langle S, g, d\rangle \to d \in \mathcal{V}(\mathcal{M}^4_B)$

Proof. Ad (1): Define $\mathcal{S} = \langle S_v, g_v, v\rangle$, where $S_v = \{s^0_v, s^1_v\}$, and for $\varphi \in \mathcal{F}$:

(i) $s^i_v(\varphi) = i$ if $\varphi \in \mathcal{A}$ and $i \in v(\varphi)$, undefined otherwise

(ii) $g_v(\varphi) = v(\varphi)$ if $\varphi \in \mathcal{A}$, $\emptyset$ otherwise.

Since $v \in \mathcal{V}(\mathcal{M}^4_B)$, by (ii), v is the minimal extension of g_v which satisfies (b1)-(b6). Lemma 3.1 implies that it is also the minimal extension of g_v which satisfies (d1)-(d6). Hence $\mathcal{S}$ is a standard structure. It is easy to see that $i \in g_v(\varphi) \Leftrightarrow \exists s \in S.i = s(\varphi)$ (for $\varphi \in \mathcal{A}$ and $i \in \{0, 1\}$). Since $g_v(\varphi)$ is non-empty only for $\varphi \in \mathcal{A}$, this implies that $\mathcal{S}$ is existential. Hence $\mathcal{S} \in \mathcal{EA}$. Ad (2): Assume $\mathcal{S} = \langle S, g, d\rangle \in \mathcal{EA}$. Since $s(\varphi)$ is undefined for $\varphi \notin \mathcal{A}$, also $g(\varphi) = \emptyset$ for $\varphi \notin \mathcal{A}$. Accordingly, g can be viewed as a function from $\mathcal{A}$ to $\mathcal{P}(\{0, 1\})$. As $\mathcal{S}$ is a standard structure, d is the minimal processor valuation which satisfies conditions (d0)–(d6). Let v be the Belnap extension of g. Then $v(p) = g(p)$ for $p \in \mathcal{A}$, which in view of $g(\varphi) = \emptyset$ for $\varphi \notin \mathcal{A}$ implies that $g(\varphi) \subseteq v(\varphi)$ for every $\varphi \in \mathcal{F}$. This and Lemma 3.1 imply that v satisfies conditions (d0)–(d6). Hence by the minimality of d we must have $d \subseteq v$. However, since $d(p) = v(p)$ for atomic p, and d satisfies conditions (b1)–(b6), the converse implication must also hold by Definition 3.2. Thus $d = v$, and d is a Belnap valuation. ■

Proposition 3.1. $\vdash_{\mathcal{EA}} = \vdash_{\mathcal{M}^4_B}$.

Proof. This is immediate from Lemma 3.2. ■

3.2. D'Ottaviano and da Costa's basic paraconsistent logic

Case II refers to the situation when the sources provide complete information about atomic formulas, and no information about complex ones. Since we assume here the existential strategy, this implies that $g(p) \neq \bot$ for every atomic formula p. By induction on the complexity of formulas (using (d1)–

(d6)), one can show that $d(\varphi) \neq \bot$ for every formula. Accordingly, the difference between this case and the previous one is that this time we have to do with just the three logical values $\mathbf{f}, \top, \mathbf{t}$. Therefore, this case is represented by the ordinary *three*-valued submatrix $\mathcal{M}_B^3 = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ of $\mathcal{M}_B^4$, with $\mathcal{V} = \{\mathbf{f}, \top, \mathbf{t}\}, \mathcal{D} = \{\top, \mathbf{t}\}, \mathcal{O} = \{\tilde{\neg}, \tilde{\vee}, \tilde{\wedge}\}$, and the deterministic interpretations of the connectives given by:

$\tilde{\vee}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
$\top$	$\top$	$\top$	$\mathbf{t}$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$

$\tilde{\wedge}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$
$\top$	$\mathbf{f}$	$\top$	$\top$
$\mathbf{t}$	$\mathbf{f}$	$\top$	$\mathbf{t}$

$\tilde{\neg}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
	$\mathbf{t}$	$\top$	$\mathbf{f}$

This matrix corresponds to the $\{\neg, \vee, \wedge\}$-fragment of D'Ottaviano and da Costa's logic J_3 ([17, 18, 3, 14]).

PROPOSITION 3.2. *Let $\mathcal{ECA}$ denote the class of standard source-processor structures $\mathcal{S} = \langle S, g, d \rangle$, where each $s \in S$ is undefined outside $\mathcal{A}$, the processor uses the existential strategy to obtain g out of the valuations in S, and $g(\varphi) \neq \bot$ for every $\varphi \in \mathcal{A}$ (i.e., the sources taken together provide information about all atomic formulae). Then $\vdash_{\mathcal{ECA}} = \vdash_{\mathcal{M}_B^3}$.*

The proof is similar to that of Proposition 3.1, so we omit it.

3.3. The most general source-processor logic

Now we shall discuss the most general case (III), when the sources can provide information about arbitrary formulas, including the complex ones, but that information may not cover all formulas, i.e. it may be incomplete. It is easy to see that in this case the conditions (d1)–(d6) from Subsection 2.3, obeyed by the processor in assigning values to formulas, imply that the presented setup can be described by the four-valued *Nmatrix* $\mathcal{M}_I^4 = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where $\mathcal{V} = \{\mathbf{f}, \bot, \top, \mathbf{t}\}, \mathcal{D} = \{\top, \mathbf{t}\}, \mathcal{O} = \{\tilde{\neg}, \tilde{\vee}, \tilde{\wedge}\}$, and the non-deterministic interpretations of the connectives are given by the following tables:

$\tilde{\vee}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}, \top\}$	$\{\mathbf{t}, \bot\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\bot$	$\{\mathbf{t}, \bot\}$	$\{\mathbf{t}, \bot\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$
$\top$	$\{\top\}$	$\{\mathbf{t}\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\mathbf{t}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$

$\tilde{\wedge}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$
$\bot$	$\{\mathbf{f}\}$	$\{\mathbf{f}, \bot\}$	$\{\mathbf{f}\}$	$\{\mathbf{f}, \bot\}$
$\top$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\top\}$	$\{\top\}$
$\mathbf{t}$	$\{\mathbf{f}\}$	$\{\mathbf{f}, \bot\}$	$\{\top\}$	$\{\mathbf{t}, \top\}$

$\tilde{\neg}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
	$\mathbf{t}$	$\bot$	$\top$	$\mathbf{f}$

Intuitively, any legal valuation of $\mathcal{M}_I^4$ represents possible information about values of formulas in a standard source-processor structure. To better understand this, let us examine the rather surprising entry in the table for $\widetilde{\vee}$ saying that $\bot \widetilde{\vee} \bot = \{\mathbf{t}, \bot\}$. Suppose that in a source-processor structure $\mathcal{S} = \langle S, g, d \rangle$ we have $d(\varphi) = d(\psi) = \bot$. Then $0 \notin d(\varphi)$ and $0 \notin d(\psi)$, so by (d4) $0 \notin d(\varphi \vee \psi)$. Hence two cases are possible. If also $1 \notin d(\varphi \vee \psi)$ (which is what one might expect in case $1 \notin d(\varphi)$ and $1 \notin d(\psi)$), then $d(\varphi \vee \psi) = \bot$. If $1 \in d(\varphi \vee \psi)$ (e.g. because there is a source s such that $s(\varphi \vee \psi) = 1$, in which case $1 \in g(\varphi \vee \psi)$ in view of the existential globalisation strategy used by the processor), then $d(\varphi \vee \psi) = \mathbf{t}$. This justifies the two options included in this table entry; some other entries are explained in [7].

PROPOSITION 3.3. *Let $\mathcal{E}$ denote the class of standard source-processor structures where the processor uses the existential strategy. Then* $\vdash_{\mathcal{E}} = \vdash_{\mathcal{M}_I^4}$.

PROOF. Let $\mathcal{V}(\mathcal{M}_I^4)$ be the set of legal valuations of $\mathcal{M}_I^4$. Obviously, v is in $\mathcal{V}(\mathcal{M}_I^4)$ iff it satisfies conditions (d1)-(d6). It follows that

(1) $\forall \mathcal{S} \in \mathcal{E}\ \forall S\ \forall g\ \forall d.\ \mathcal{S} = \langle S, g, d \rangle \to d \in \mathcal{V}(\mathcal{M}_I^4)$

Now assume that $v \in \mathcal{V}(\mathcal{M}_I^4)$. For $i = 0, 1$ and for every $\varphi \in \mathcal{F}$, let $s_v^i(\varphi) = i$ if $i \in v(\varphi)$, and undefined otherwise. It is easy to see that $\mathcal{S} = \langle \{s_v^0, s_v^1\}, v, v \rangle$ is an element of $\mathcal{E}$. Hence:

(2) $\forall v \in \mathcal{V}(\mathcal{M}_I^4)\ \exists \mathcal{S} \in \mathcal{E}\ \exists S\ \exists g. \mathcal{S} = \langle S, g, v \rangle$

The theorem is now immediate from (1) and (2). ■

3.4. General source-processor logic with complete information

The last case is when the sources provide complete information about all atomic formulas (but they may provide information, not necessarily complete, about other formulas too). Thus for any atomic formula p of L_C, some source in S must say either that p is true or that p is false. Like in Subsection 3.2, one can easily prove by induction that under this condition no formula is given the value $\bot$. Thus in this case too only three truth-values are employed. However, this time the scenario gives rise to a logic based on a three-valued *Nmatrix*. This is the Nmatrix $\mathcal{M}_I^3 = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where $\mathcal{V} = \{\mathbf{f}, \top, \mathbf{t}\}, \mathcal{D} = \{\top, \mathbf{t}\}, \mathcal{O} = \{\widetilde{\neg}, \widetilde{\vee}, \widetilde{\wedge}\}$, and the non-deterministic interpretations of the connectives are given by:

$\widetilde{\neg}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
	$\mathbf{t}$	$\top$	$\mathbf{f}$

$\widetilde{\vee}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}, \top\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\top$	$\{\top\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\mathbf{t}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$

$\widetilde{\wedge}$	$\mathbf{f}$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$
$\top$	$\{\mathbf{f}\}$	$\{\top\}$	$\{\top\}$
$\mathbf{t}$	$\{\mathbf{f}\}$	$\{\top\}$	$\{\mathbf{t}, \top\}$

We leave the proof of the following easy proposition to the reader:

Proposition 3.4. *Let $\mathcal{EC}$ denote the class of standard source-processor structures where the sources taken together provide some information about every atomic formula, and the processor uses the existential strategy. Then* $\vdash_{\mathcal{EC}} = \vdash_{\mathcal{M}_I^3}$.

4. The universal strategy

In this section we discuss in brief the case in which the processor applies the universal strategy in collecting information from the sources. Note first that if there are at least two sources then $g(\varphi)$ may be $\bot$ in this case, even if all the sources are of the "know all" type (because the processor will assign neither 0 nor 1 to a formula φ which is assigned different values by two sources). On the other hand, it is obvious that with the universal strategy $g(\varphi) \neq \top$ for every formula φ. Without further integrity constraints, this is not necessarily true for the standard extension d of g. One such plausible constraint is that the sources should all be *classically coherent* (see Subsection 2.2.1). Another is again that the sources provide information about atomic formulae only. It is easy to see that the resulting logic in the latter case is that induced by the famous (strong) 3-valued matrix $\mathcal{M}_K^3$ of Kleene:

$\tilde{\vee}$	**f**	$\bot$	**t**
f	**f**	$\bot$	**t**
$\bot$	$\bot$	$\bot$	**t**
t	**t**	**t**	**t**

$\tilde{\wedge}$	**f**	$\bot$	**t**
f	**f**	**f**	**f**
$\bot$	**f**	$\bot$	$\bot$
t	**f**	$\bot$	**t**

$\tilde{\neg}$	**f**	$\bot$	**t**
	t	$\bot$	**f**

Proposition 4.1. *Let $\mathcal{AA}$ be the class of standard source-processor structures $\mathcal{S} = \langle S, g, d \rangle$, where each $s \in S$ is undefined outside $\mathcal{A}$ and the processor uses the universal strategy to obtain g out of the valuations in S. Then* $\vdash_{\mathcal{AA}} = \vdash_{\mathcal{M}_K^3}$.

Proof. Similar to the proof of Proposition 3.1. ■

5. Proof systems for the existential strategy

Now we will proceed to develop proof systems for the four logics discussed in the preceding section. As we explained in Subsection 2.4, to compensate for the weakness of the language, the systems we will provide will be strongly sound and complete sequent calculi.

5.1. The most general source-processor logic

We begin with the proof system for the most general case, from which we will later derive the proof systems for the remaining cases.

Definition 5.1. Let $\mathcal{C}^4_I$ be the sequent calculus defined as follows:

Axioms: $\varphi \Rightarrow \varphi$

Structural inference rules: Weakening, Cut.

Logical inference rules:

$$(\neg\neg \Rightarrow) \quad \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma, \neg\neg\varphi \Rightarrow \Delta} \qquad\qquad (\Rightarrow \neg\neg) \quad \frac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma \Rightarrow \Delta, \neg\neg\varphi}$$

$$(\Rightarrow \vee) \quad \frac{\Gamma \Rightarrow \Delta, \varphi, \psi}{\Gamma \Rightarrow \Delta, \varphi \vee \psi}$$

$$(\neg\vee \Rightarrow) \quad \frac{\Gamma, \neg\varphi, \neg\psi \Rightarrow \Delta}{\Gamma, \neg(\varphi \vee \psi) \Rightarrow \Delta} \qquad\qquad (\Rightarrow \neg\vee) \quad \frac{\Gamma \Rightarrow \Delta, \neg\varphi \;\; \Gamma \Rightarrow \Delta, \neg\psi}{\Gamma \Rightarrow \Delta, \neg(\varphi \vee \psi)}$$

$$(\wedge \Rightarrow) \quad \frac{\Gamma, \varphi, \psi \Rightarrow \Delta,}{\Gamma, \varphi \wedge \psi \Rightarrow \Delta} \qquad\qquad (\Rightarrow \wedge) \quad \frac{\Gamma \Rightarrow \Delta, \varphi \;\; \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}$$

$$(\Rightarrow \neg\wedge) \quad \frac{\Gamma \Rightarrow \Delta, \neg\varphi, \neg\psi}{\Gamma \Rightarrow \Delta, \neg(\varphi \wedge \psi)}$$

Definition 5.2. Let $\underline{\mathcal{C}}^4_I$ be the calculus obtained from $\mathcal{C}^4_I$ by limiting the applications of the cut rule to formulas occurring in the premises of sequent derivations. In other words: If $S = \{\Gamma_1 \Rightarrow \Delta_1, \ldots, \Gamma_n \Rightarrow \Delta_n\}$ then $S \vdash_{\underline{\mathcal{C}}^4_I} \Sigma$ if there is a proof of Σ from S in $\mathcal{C}^4_I$ in which all cuts are on formulas from $\bigcup_{i=1}^n (\Gamma_i \cup \Delta_i)$ (in particular: $\vdash_{\underline{\mathcal{C}}^4_I} \Sigma$ iff Σ has a cut-free proof in $\mathcal{C}^4_I$).

Theorem 5.1. *The calculus $\underline{\mathcal{C}}^4_I$ is finitely strongly sound and complete for $\vdash_{\mathcal{M}^4_I}$, i.e., for any finite set of sequents $S \subseteq Seq$ and any sequent $\Sigma \in Seq$, $S \vdash_{\mathcal{M}^4_I} \Sigma$ iff $S \vdash_{\underline{\mathcal{C}}^4_I} \Sigma$.*

Proof. For simplicity, in what follows we drop the decorations on $\models$.

It is easy to see that $\mathcal{C}^4_I$ is strongly sound for $\vdash_{\mathcal{M}^4_I}$ (i.e. if $S \vdash_{\mathcal{C}^4_I} \Sigma$ then $S \vdash_{\mathcal{M}^4_I} \Sigma$). Hence it suffices to prove the strong completeness of $\underline{\mathcal{C}}^4_I$ for finite premise sets.

We argue by contradiction. Suppose that for a finite set of sequents S and a sequent $\Sigma_0 = \Gamma \Rightarrow \Delta$ we have $S \vdash_{\mathcal{M}^4_I} \Sigma_0$, but Σ_0 is not derivable from S in $\underline{\mathcal{C}}_I^4$. We shall construct a counter-valuation v such that $v \models S$ but $v \not\models \Sigma_0$.

Denote by $F(S)$ the set of all formulae belonging to at least one of the sides in some sequent in S. Then $F(S)$ is finite; assume it has l elements. Let $\varphi_1, \varphi_2, \ldots, \varphi_l$ be an enumeration of formulae in $F(S)$. We shall now define a sequence of sequents $\Gamma_n \Rightarrow \Delta_n, n = 0, 1, \ldots, l$, such that, for $n = 0, 1, \ldots, l$:

(i) $\Gamma \subseteq \Gamma_n$, $\Delta \subseteq \Delta_n$

(ii) If $n \neq 0$ then $\varphi_n \in (\Gamma_n \cup \Delta_n)$.

(iii) $\Gamma_n \Rightarrow \Delta_n$ is not derivable from S in $\underline{\mathcal{C}}_I^4$.

The above sequences are defined inductively as follows:

- We put $\Gamma_0 = \Gamma, \Delta_0 = \Delta$. As by our assumption $\Gamma \Rightarrow \Delta$ is not derivable from S in $\underline{\mathcal{C}}_I^4$, (i)–(iii) above are satisfied for $n = 0$.
- Suppose $n \leq l - 1$ and we have defined the sequents $\Gamma_i \Rightarrow \Delta_i$ satisfying conditions (i)–(iii) for $i \leq n$. Then the sequents $\Sigma_1 = \Gamma_n \Rightarrow \Delta_n, \varphi_{n+1}$ and $\Sigma_2 = \varphi_{n+1}, \Gamma_n \Rightarrow \Delta_n$ cannot be both derivable from S in $\underline{\mathcal{C}}_I^4$, since then $\Gamma_n \Rightarrow \Delta_n$ would be derivable from them by an allowed cut on the formula $\varphi_{n+1} \in S$. We take $\Gamma_{n+1} \Rightarrow \Delta_{n+1}$ to be Σ_1, if Σ_1 is not derivable, and Σ_2 otherwise. Then, obviously, from the inductive assumption it follows that the sequence $\Gamma_{n+1} \Rightarrow \Delta_{n+1}$ satisfies conditions (i)–(iii).

By induction, each element of the sequence $\Gamma_n \Rightarrow \Delta_n, n = 0, 1, \ldots, l$, satisfies the desired conditions (i)–(iii). What is more, from the inductive construction we can see that $\Gamma_n \subseteq \Gamma_{n+1}, \Delta_n \subseteq \Delta_{n+1}$ for $n = 1, 2, \ldots, l - 1$. Now call a sequent $\Gamma \Rightarrow \Delta$ *saturated* if it is closed under the logical rules in $\underline{\mathcal{C}}_I^4$ applied backwards (e.g., if $\varphi \wedge \psi \in \Gamma$ then both φ and ψ are in Γ, while if $\varphi \wedge \psi \in \Delta$ then either $\varphi \in \Delta$ or $\psi \in \Delta$). Let $\Gamma^* \Rightarrow \Delta^*$ be an extension of $\Gamma_l \Rightarrow \Delta_l$ to a saturated sequent which is not derivable from S in $\underline{\mathcal{C}}_I^4$ (it is easy to see that such a sequent exists). Then $\Gamma \subseteq \Gamma^*, \Delta \subseteq \Delta^*$, and $F(S) \subseteq \Gamma^* \cup \Delta^*$. Using $\Gamma^* \Rightarrow \Delta^*$ we define the desired valuation v as follows:

- For any atomic p:

  (v0) $1 \in v(p)$ iff $p \in \Gamma^*$, $0 \in v(p)$ iff $\neg p \in \Gamma^*$;
- For any formulas α, β:

  (v1) $1 \in v(\neg\alpha)$ iff $0 \in v(\alpha)$;

  (v2) $0 \in v(\neg\alpha)$ iff $1 \in v(\alpha)$;

  (v3) $1 \in v(\alpha \vee \beta)$ iff $1 \in v(\alpha)$ or $1 \in v(\beta)$ or $(\alpha \vee \beta) \in \Gamma^*$;

(v4) $0 \in v(\alpha \vee \beta)$ iff $0 \in v(\alpha)$ and $0 \in v(\beta)$;
(v5) $1 \in v(\alpha \wedge \beta)$ iff $1 \in v(\alpha)$ and $1 \in v(\beta)$;
(v6) $0 \in v(\alpha \wedge \beta)$ iff $0 \in v(\alpha)$ or $0 \in v(\beta)$ or $\neg(\alpha \wedge \beta) \in \Gamma^*$;

It can be easily checked, by considering the truth tables of the Nmatrix $\mathcal{M}_I^4$, that v defined as above is a legal valuation for that Nmatrix. It remains to prove that v is indeed the desired counter-valuation, i.e., that:

(I) $v \models \Sigma$ for each $\Sigma \in S$; (II) $v \not\models (\Gamma \Rightarrow \Delta)$;

We start with (II). As $\Gamma \subseteq \Gamma^*, \Delta \subseteq \Delta^*$, then in order to prove (II) it suffices to prove that $v \not\models (\Gamma^* \Rightarrow \Delta^*)$. To this end, we have to show that:

(A) $v \models \gamma$ for each $\gamma \in \Gamma^*$; (B) $v \not\models \delta$ for each $\delta \in \Delta^*$

We argue by induction on the complexity of formulas.

Proof of (A):

- Assume γ is atomic. Then $1 \in v(\gamma)$ by (v0) in the definition of v, whence $v \models \gamma$.
- Assume $\gamma = \neg\gamma'$. This case splits in the following four subcases:

  $\gamma' = p$ (where p is atomic): Then $\neg p = \gamma \in \Gamma^*$, and by (v0) in the definition of v we have $0 \in v(p)$, whence by (v1) of that definition we get $1 \in v(\gamma)$;

  $\gamma' = \neg\alpha$**:** Then $\neg\neg\alpha = \gamma \in \Gamma^*$. As $\Gamma^* \Rightarrow \Delta^*$ is saturated, then by rule $(\neg\neg \Rightarrow)$ we have $\alpha \in \Gamma^*$, whence by the inductive assumption $1 \in v(\alpha)$, which in turn yields $1 \in v(\neg\neg\alpha) = v(\gamma)$ by applications of (v2) and (v1);

  $\gamma' = \alpha \vee \beta$**:** Then $\neg(\alpha \vee \beta) = \gamma \in \Gamma^*$. As the sequent $\Gamma^* \Rightarrow \Delta^*$ is saturated, then by rule $(\neg\vee \Rightarrow)$ we have $\neg\alpha, \neg\beta \in \Gamma^*$, whence by the inductive assumption $1 \in v(\neg\alpha), 1 \in v(\neg\beta)$ Thus $0 \in v(\alpha), 0 \in v(\beta)$ by (v1), whence $0 \in v(\alpha \vee \beta)$ by (v4), and finally $1 \in v(\neg(\alpha \vee \beta)) = v(\gamma)$ by (v1);

  $\gamma' = \alpha \wedge \beta$**:** Then $\neg(\alpha \wedge \beta) = \gamma \in \Gamma^*$, whence by (v6) $0 \in v(\alpha \wedge \beta)$, which yields $1 \in v(\neg(\alpha \wedge \beta)) = v(\gamma)$.

- Assume $\gamma = \gamma_1 \vee \gamma_2$. Then $\gamma_1 \vee \gamma_2 = \gamma \in \Gamma^*$, so by (v3) we have $1 \in v(\gamma_1 \vee \gamma_2) = v(\gamma)$.
- Assume $\gamma = \gamma_1 \wedge \gamma_2$. As $\gamma \in \Gamma^*$ and the sequent $\Gamma^* \Rightarrow \Delta^*$ is saturated, then by rule $(\wedge \Rightarrow)$ we have $\gamma_1, \gamma_2 \in \Gamma^*$, whence by the inductive assumption $1 \in v(\gamma_1), 1 \in v(\gamma_2)$, which yields $1 \in v(\gamma_1 \wedge \gamma_2) = v(\gamma)$ by (v5).

Proof of (B):

> Assume first that $\delta = p$ where p is atomic. As $\delta \in \Delta^*$ and $\Gamma^* \Rightarrow \Delta^*$ is not derivable, then $p \notin \Gamma^*$, whence $1 \notin v(\delta)$ by (v0). In turn, if $\delta = \neg p$, then $\neg p \notin \Gamma^*$, for $\Gamma^* \Rightarrow \Delta^*$ is not derivable. Thus by (v0) $0 \notin v(p)$, whence by (v1) we have $1 \notin v(\neg p) = v(\delta)$.
>
> The proof that (B) holds for δ's which are not literals is carried out by induction, following a single schema analogous to the proof of, e.g., the last case in (A). Each time, from the fact that $\Gamma^* \Rightarrow \Delta^*$ is saturated and from the right hand side rule of the sequent calculus corresponding to the given complex formula α we conclude that the appropriate component formulas of α must also be in Δ^*, whence by the inductive assumption they are not assigned 1 by v. From the latter we deduce that $1 \notin v(\alpha)$ using the appropriate clauses of the definition of v.

This ends the proof of (II) above. It remains to prove (I), i.e., to show that $v \models \Sigma$ for each $\Sigma \in S$. So let $\Sigma \in S$. Then $\Sigma = \varphi_1, \dots, \varphi_k \Rightarrow \psi_1, \dots, \psi_m$ for some integers k, m and formulas $\varphi_i, \psi_j, i = 1, \dots, k, j = 1, \dots, m$. Clearly, we cannot have both $\{\varphi_1, \dots, \varphi_k\} \subseteq \Gamma^*$ and $\{\psi_1, \dots, \psi_m\} \subseteq \Delta^*$, for then $\Gamma^* \Rightarrow \Delta^*$ would be derivable from Σ, and hence from S, by weakening. Since $F(S) \subseteq \Gamma^* \cup \Delta^*$, this implies that either $\varphi_i \in \Delta^*$ for some i, or $\psi_j \in \Gamma^*$ for some j. Hence by (A) and (B), which we have already proved, we have either $v \not\models \varphi_i$ for some i, or $v \models \psi_j$ for some j, which implies that $v \models \Sigma$. ■

Corollary 5.1. *The calculus $\mathcal{C}^4_I$ is (weakly) sound and complete for $\vdash_{\mathcal{M}^4_I}$, and the cut rule is admissible in it. In particular: If Γ is a finite set of formulas, and φ is a formula, then $\Gamma \vdash_{\mathcal{M}^4_I} \varphi$ if the sequent $\Gamma \Rightarrow \varphi$ has a cut-free proof in $\mathcal{C}^4_I$.*[5]

Proof. This follows from the last theorem and Note 3.1. ■

Note 5.1. The finiteness assumption can in fact be dropped from the formulation of Theorem 5.1, for the theorem holds for infinite premise sets too. However, we skip the proof of this fact here, for such a generalization seems to be of little practical usefulness for the purposes of this paper.

5.2. The other logics

Strongly finitely sound and complete calculi for the other main logics investigated in this paper can be obtained by extending $\mathcal{C}^4_I$ and $\underline{\mathcal{C}}^4_I$ with appropriate rules and axiom:

---

[5] This corollary was first proved (using a different method) in [7].

General source-processor logic with complete information
Calculi corresponding to this case are obtained by augmenting $\mathcal{C}_I^4$ and $\underline{\mathcal{C}}_I^4$ with either the excluded middle axiom $\Rightarrow \varphi, \neg\varphi$, or by the left-to-right swap rule

$$\frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\varphi}$$

(However, addition of the swap rule does not allow us to eliminate any of the previous negation rules, for none of them is derivable from it).

Dunn-Belnap's logic
To obtain the calculi for Dunn-Belnap's logic, we augment $\mathcal{C}_I^4$ and $\underline{\mathcal{C}}_I^4$ by the two symmetric rules "missing" from them, i.e.

$$(\vee \Rightarrow)\ \frac{\Gamma, \varphi \Rightarrow \Delta\ \Gamma, \psi, \Rightarrow \Delta}{\Gamma, \varphi \vee \psi \Rightarrow \Delta} \qquad (\neg\wedge \Rightarrow)\ \frac{\Gamma, \neg\varphi \Rightarrow \Delta\ \Gamma, \neg\psi \Rightarrow \Delta}{\Gamma, \neg(\varphi \wedge \psi) \Rightarrow \Delta}$$

The resulting system is the L_C-fragment of the Gentzen-type system for "first degree entailments" introduced in [1]. The fact that $\mathcal{M}_B^4$ is weakly characteristic for first degree entailments (and so for this Gentzen-type system) is well-known (see [2, 12, 13]).

D'Ottaviano and da Costa's logic
The calculi for the above paraconsistent logic are obtained by adding either the excluded middle axiom $\Rightarrow \varphi, \neg\varphi$, or the left-to-right swap rule, to the calculi for Belnap's logic.

Kleene's logic As is well known, calculi for Kleene's 3-valued logic are obtained by adding to the calculi for Belnap's logic either the axiom $\varphi, \neg\varphi \Rightarrow$ (corresponding to the law of contradiction), or the right-to-left swap rule $\frac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma, \neg\varphi \Rightarrow \Delta}$.

Theorem 5.2. *The obvious analogues of Theorem 5.1 (and its Corollary 5.1) hold for each of the calculi introduced above with respect to their associate matrix/Nmatrix.*

The proofs are similar to that of Theorem 5.1, and are left to the reader.

6. Future research

One direction of future research is to explore the general case of the universal strategy, namely, one where the sources can also provide information about complex formulas. As the introduction to Section 4 implies, it will split in two subcases: one when the final processor valuation d can also take the value $\top$, and one where this is not possible due to an additional constraint, like classical coherence (see Section 2.2.1).

Another direction is to investigate other variants of the framework, especially those signalled in Subsections 2.2.1 and 2.2.3.[6]

Finally, a major goal will be to upgrade our framework and results to the first-order language, or some weaker but more manageable language above the propositional level.

Acknowledgment. This research was supported by THE ISRAEL SCIENCE FOUNDATION founded by The Israel Academy of Sciences and Humanities (grant No. 809/06).

References

[1] Anderson, A.R., Belnap, N.D., 'First degree entailments', *Mathematische Annalen*, 149: 302–319, 1963.

[2] Anderson, A.R., Belnap, N.D., *Entailment, vol. 1*, Princeton University Press, Princeton NJ, 1975.

[3] Avron, A., 'Natural 3-valued logics: Characterization and proof theory', *The Journal of Symbolic Logic*, 56: 276–294, 1991.

[4] Avron, A., 'Logical Non-determinism as a tool for logical modularity: an introduction', in Artemov, S., Barringer, H., d'Avila Garcez, A.S., Lamb, L.C., Woods, J. (eds.), *We Will Show Them: Essays in Honor of Dov Gabbay, vol. 1*, College Publications, Ithaca, NY, 2005, pp. 105–124.

[5] Avron, A., Lev, I., 'Canonical propositional Gentzen-type systems', in Goré, R., Leitsch, A., Nipkow, T., *Proceedings of the 1st International Joint Conference on Automated Reasoning (IJCAR 2001)*, LNAI, vol. 2083, Springer Verlag, 2001, pp. 529–544.

[6] Avron, A., Lev, I., 'Non-deterministic multiple-valued structures', *Journal of Logic and Computation*, 15: 241–261, 2005.

[7] Avron, A., Ben-Naim, J., Konikowska, B., 'Cut-free ordinary sequent calculi for logics having finite-valued semantics', *Logica Universalis*, 1: 41–69, 2006.

[8] Belnap, N.D., 'How computers should think', in Rylem, G. (ed.), *Contemporary Aspects of Philosophy*, Oriel Press, Stocksfield, England, 1977, pp. 30–56.

[9] Belnap, N.D., 'A useful four-valued logic', in Epstein, G., Dunn, J.M., *Modern Uses of Multiple-Valued Logic*, Reidel, Dordrecht, 1977, pp. 7–37.

---

[6] Some partial results in this direction have already been obtained concerning the important case in which all the sources are classically closed. The resulting logic is quite interesting: it is paraconsistent, but every classical tautology is valid in it. It respects many important classical equivalences, but not the distributive law. Especially important is the fact that there is no n such that a processor valuation can be obtained in a source-processor framework where all sources are classically closed only if it can be obtained in a framework of this sort in which there are at most n sources (while by our proofs just two suffice for the cases investigated in this paper).

[10] Carnielli, W.A., Lima-Marques, M., 'Society semantics for multiple-valued logics', in Carnielli, Walter A., D'Ottaviano, Itala M.L. (eds.), *Proceedings of the XII EBL—Advances in Contemporary Logic and Computer Science*, American Mathematical Society, Series Contemporary Mathematics, vol. 235, 1999, pp. 33–52.

[11] Carnielli, W.A., Marcos, J., de Amo, S., 'Formal inconsistency and evolutionary databases', *Logic and Logical Philosophy*, 8: 115–152, 2000.

[12] Dunn, J.M., 'Intuitive semantics for first degree entailments and "coupled trees"', *Philosophical Studies*, 29: 149–168, 1976.

[13] Dunn, J.M., 'Relevant logic and entailment', *Handbook of Philosophical Logic*, vol. III, Gabbay, D., Guenthner, F. (eds.), 1984.

[14] Epstein, R.L., The semantic foundation of logic, 2nd ed., vol. I: *Propositional Logics*, ch. IX, Kluwer Academic Publisher, 1995.

[15] Fitting, M., 'Kleene's three-valued logics and their children', *Fundamenta Informaticae*, 20: 113–131, 1994.

[16] Ginsberg, M.L., 'Multivalued logics: A uniform approach to reasoning in AI', *Computer Intelligence*, 4: 256–316, 1988.

[17] D'Ottaviano, I.L.M., da Costa, N.C.A., 'Sur un problème de Jaśkowski', *Comptes Rendus de l'Académie des Sciences de Paris*, 270: 1349–1353, 1970.

[18] D'Ottaviano, I.L.M., 'The completeness and compactness of a three-valued first-order logic', *Revista Colombiana de Matematicas*, XIX: 31–42, 1985.

[19] Shramko, Y., Wansing, H., 'Some useful 16-valued logics. How a computer network should think', *Journal of Philosophical Logic*, 34: 121–153, 2005.

[20] Shramko, Y., Wansing, H., 'Hyper-contradictions, generalized truth values, and logics of truth and falsehood', *Journal of Logic, Language and Information*, 15: 403–424, 2006.

Arnon Avron
School of Computer Science
Tel-Aviv University
Ramat Aviv 69978, Israel
`aa@post.tau.ac.il`

Jonathan Ben-Naim
CNRS Researcher in Computer Science
Université Paul Sabatier, IRIT,
LILaC 118 route de Narbonne
F-31062 Toulouse Cedex 9,
France
`bennaim@irit.fr`

Beata Konikowska
Institute of Computer Science
Polish Academy of Sciences
Warsaw, Poland
`beatak@ipipan.waw.pl`

JUI-LIN LEE

The Classical Model Existence Theorem in Subclassical Predicate Logics I

Abstract. We prove that in predicate logics there are some classically sound Hilbert systems which satisfy the classical model existence theorem (every $\bot$-consistent set has a classical model) but are weaker than first order logic.

Keywords: extended completeness theorem, strong completeness, prenex normal form, intuitionistic logic, three-valued logic.

1. Introduction

In classical logic the extended completeness theorem (for any Σ and φ, $\Sigma \models \varphi$ implies $\Sigma \vdash \varphi$) is obviously an extension of completeness theorem (for any φ, $\models \varphi$ implies $\vdash \varphi$). To prove the extended completeness theorem by completeness theorem, one can simply use the compactness theorem for classical semantics ($\Sigma \models \varphi$ implies that there is a finite $\Sigma_0 \subseteq \Sigma$ such that $\Sigma_0 \models \varphi$) and the semantic deduction theorem ($\Sigma \cup \{\alpha\} \models \beta$ implies $\Sigma \models \alpha \to \beta$). Since to prove the compactness theorem for classical semantics one may need some extra effort (e.g. ultraproduct), it seems more convenient to prove the extended completeness theorem directly.

The direct proof of extended completeness theorem for classical logic, in most logic textbooks, is done by proving the following two properties:

(CME) Every consistent set has a classical model (under the standard two-valued truth-functional semantics).

(RAA) If $\Sigma \not\vdash \varphi$, then $\Sigma \cup \{\neg\varphi\}$ is consistent (i.e., $\Sigma \cup \{\neg\varphi\} \not\vdash \bot$).

Let us call the first statement the classical model existence property (CME for short) and the second statement the Reducio Ad Absurdum property (RAA for short). In classical logic one can easily formulate RAA into an axiom scheme (with the aid of deduction theorem and modus ponens). Since RAA is trivial in classical logic, CME, as the major step in proving the extended completeness theorem, is called the strong completeness theorem for first order logic (see [3]).

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

Though CME holds in classical logic, it may be a mistake to call CME the strong completeness property if there are other classically sound logics which also satisfy CME (then CME does not characterize the classical logic).

Consider the $\bot$-consistency case (i.e., Σ is $\bot$-*consistent* iff $\Sigma \nvdash \bot$). One can see that $L_1 = \{(\Sigma, \bot) \mid \Sigma \text{ is not classically satisfiable}\}$ is the weakest relation which is classically sound and satisfies CME, and $L_2 = L_1 \cup \{(\emptyset, \varphi) \mid \varphi \text{ is a classical tautology}\}$ is the weakest relation which is classically sound and satisfies CME and the completeness theorem (as a property). But the problem is that L_1 (or L_2) may not be a logic. Whether L_1 (or L_2) is qualified for being a logic depends on what we think a logic is[1].

Here we take the view that a logic is a proof system. It is known that in propositional logics the classical consistency is equivalent to the intuitionistic consistency (see [12]). Therefore the intuitionistic propositional logic IPL also satisfies CME and then CME does not deserve the name "strong completeness" in propositional logics. The key step can be briefly sketched as follows: If $\Sigma \vdash_{CPL} \bot$, then (by employing Glivenko's theorem) we have $\Sigma \vdash_{IPL} \neg\neg\bot$. Since $\neg\neg\bot$ is $(\bot \to \bot) \to \bot$, $\bot \to \bot$ is provable in IPL, and modus ponens holds in IPL, we have $\Sigma \vdash_{IPL} \bot$. And IPL is not the weakest proof system satisfying CME. In [6] it is proved that CME with respect to the $\bot$-consistency holds even in some paraconsistent logic weaker than IPL.

What may be interesting is that such logics cannot be distinguished from classical logic by any example/non-example of $\bot$-consistent sets, i.e., in any of these logics (including classical logic) for any set Σ of sentences, whether Σ can derive $\bot$ does not depend on which of these logics we choose.

In this paper we consider the predicate case. We prove that there are some classically sound Hilbert systems which satisfy CME but are weaker than first order logic. This proof is based on the following three facts:

(1) In propositional logics CME holds in some weak proof systems. (See Section 2.)

(2) For any $\bot$-consistent set of prenex normal form sentences, there is a Herbrand-Henkin style extension (by adding witnesses) such that the remaining step of constructing a classical model can be done at the quantifier-free level. (See Section 3.)

---

[1] If one insists that every logic must satisfy RAA, then the classical logic becomes the only logic which satisfies both CME and RAA. On the other hand, if one takes the view that a logic is a consequence operator, the minimal consequence operator containing L_2 is not the classical logic.

(3) Converting a sentence into a prenex normal form sentence, though not allowed in intuitionistic predicate logic, can be done in some logic weaker than FOL. (See Section 4.)

This paper is organized as follows: In Section 2 we sketch the proof of CME in weak propositional logics $\mathcal{H}_1$ and $\mathcal{H}_2$, where $\mathcal{H}_1$ is the $\{\to, \bot\}$ fragment of IPL and $\mathcal{H}_2$, a sublogic of $\mathcal{H}_1$, is a paraconsistent logic. We then define the predicate logics $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$,$\mathcal{Q}_1^{=}\mathcal{H}_2$. In Section 3 we give a Herbrand-Henkin style proof of CME for $\bot$-consistent sets of prenex normal form sentences w.r.t. proof systems $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$,$\mathcal{Q}_1^{=}\mathcal{H}_2$. In Section 4 we prove that prenex normal form theorem holds in any axiomatic extension of $\mathcal{Q}_1$ (including $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$,$\mathcal{Q}_1^{=}\mathcal{H}_2$). And we sketch how these logics are weaker than FOL by three-valued semantics. In Section 5 we discuss the relationship of Kripke models of $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$ and classical models.

Remark 1.1. The $\bot$-consistency of intuitionistic predicate logic is different from that of first order logic: Since $\neg\neg\forall x[A(x) \vee \neg A(x)]$ is classically valid but not intuitionistically valid, $\{\neg\forall x[A(x) \vee \neg A(x)]\}$ is classically inconsistent but intuitionistically consistent. (See [1], pp. 48–49.)

2. Classical model existence theorem in propositional logics

In this section we sketch the proof in [6] that both $\mathcal{H}(DT1, DT2, ECQ; MP)$ (the $\{\to, \bot\}$-fragment of IPL) and $\mathcal{H}(DT1, DT2, DA^*_\bot; MP)$ (a paraconsistent logic) satisfy CME with respect to the $\bot$-consistency. At the end of this section we define the predicate proof systems which will be used later.

For convenience in this paper we denote $\mathcal{H}_1 = \mathcal{H}(DT1, DT2, ECQ; MP)$ and $\mathcal{H}_2 = \mathcal{H}(DT1, DT2, DA^*_\bot; MP)$. In these two systems, they have only one inference rule MP: A, $A \to B$ infer B, and their axiom schemes are:

$(DT1)$ $A \to (B \to A)$

$(DT2)$ $[A \to (B \to C)] \to [(A \to B) \to (A \to C)]$

(ECQ) $\bot \to A$

$(DA^*_\bot)$ $(A \to \bot) \to \{[(A \to B) \to \bot] \to \bot\}$

Note that the following proof does not rely on the syntactic translation proof of Glivenko's theorem.

Assume that Γ is $\bot$-consistent, i.e., $\Gamma \nvdash \bot$ (w.r.t. $\mathcal{H}_1$ or $\mathcal{H}_2$). We will show that Γ has a classical model. By Lindenbaum's theorem (enumerating all sentences $\psi_0, \ldots, \psi_n, \ldots$ and sequentially adding ψ_n to the set if the $\bot$-consistency is preserved) Γ can be extended to a maximal consistent set Δ.

Δ is then *negation complete*, i.e., for any sentence φ exactly one of φ and $\neg\varphi (= \varphi \to \bot)$ is in Δ: Not both $\varphi, \neg\varphi$ in Δ is clear by MP. If $\varphi = \psi_n \notin \Delta$, then $\Delta \cup \{\varphi\} \vdash \bot$. By deduction theorem $\Delta \vdash \varphi \to \bot$. Since Δ can be proved to be deductively closed (i.e., $\Delta \vdash \alpha$ implies $\alpha \in \Delta$ for any α), $\neg\varphi \in \Delta$. To show that the truth-functionality of $\to$ (i.e., $\alpha \to \beta \in \Delta$ iff $\alpha \notin \Delta$ or $\beta \in \Delta$ for any α, β) holds in Δ, it is by MP (from left to right), $(DT1)$ (from $\beta \in \Delta$ to $\alpha \to \beta \in \Delta$), and (ECQ) or $(DA^*_\bot)$ (from $\alpha \notin \Delta$ to $\alpha \to \beta \in \Delta$). Note that in $\mathcal{H}_1$ the scheme (DA): $(A \to \bot) \to (A \to B)$ is derivable and it can be used to show $\alpha \to \beta \in \Delta$ from $\alpha \notin \Delta$. In $\mathcal{H}_2$ the weaker axiom $(DA^*_\bot)$ will suffice to show the truth-functionality of $\to$ on Δ by the following argument: "if $\alpha \notin \Delta$ and $\alpha \to \beta \notin \Delta$, then $\Delta \vdash \bot$."

Note that in $\mathcal{H}_1$ the $\bot$-consistency is equivalent to the absolute consistency[2]. In $\mathcal{H}_2$ it is not so: $\{\bot\}$ is absolutely consistent in $\mathcal{H}_2$ but it does not have a classical model. However, CME with respect to the $\bot$-consistency still holds in this weaker system $\mathcal{H}_2$.

Now we define predicate proof systems which will be used in this paper.

Definition 2.1. Let $i \in \{1, 2\}$. The equality-free proof system $\mathcal{Q}_1\mathcal{H}_i$ has only one inference rule MP, and the axiom schemes of $\mathcal{Q}_1\mathcal{H}_i$ are axiom schemes from $\mathcal{H}_i$ (in the predicate sense), $(Ax1)$–$(Ax15)$, and all their universal generalizations[3]. Here all formulas used in $\mathcal{Q}_1\mathcal{H}_i$ are equality-free.

The proof system $\mathcal{Q}_1^=\mathcal{H}_i$ are defined in a similar way. Axiom schemes of $\mathcal{Q}_1^=\mathcal{H}_i$ are: the axiom schemes of $\mathcal{H}_i$, the axiom schemes $(Ax1)$–$(Ax18)$, and all their universal generalizations. Below we list the axiom schemes $(Ax1)$–$(Ax18)$:

$(Ax1)$ $\forall x\varphi(x) \to \varphi(t)$, where the term t is free for x in $\varphi(x)$.

$(Ax2)$ $\varphi(t) \to \exists x\varphi(x)$, where the term t is free for x in $\varphi(x)$.

$(Ax3)$ $\varphi \to \forall x\varphi$, where $x \notin FV(\varphi)$.

$(Ax4)$ $\forall x(\varphi \to \psi) \to [\forall x\varphi \to \forall x\psi]$

$(Ax5)$ $\forall x(\varphi \to \psi) \to [\exists x\varphi \to \exists x\psi]$

$(Ax6)$ $\forall x(\varphi \to \psi) \to [\varphi \to \forall x\psi]$, where $x \notin FV(\varphi)$.

$(Ax7)$ $[\varphi \to \forall x\psi] \to \forall x(\varphi \to \psi)$, where $x \notin FV(\varphi)$.

$(Ax8)$ $\exists x(\varphi \to \psi) \to [\varphi \to \exists x\psi]$, where $x \notin FV(\varphi)$.

$(Ax9)$ $[\varphi \to \exists x\psi] \to \exists x(\varphi \to \psi)$, where $x \notin FV(\varphi)$.

$(Ax10)$ $\forall x(\varphi \to \psi) \to [\exists x\varphi \to \psi]$, where $x \notin FV(\psi)$.

---

[2] Σ is absolutely consistent with respect to $\mathcal{Q}$ iff $\Sigma \nvdash_{\mathcal{Q}} \varphi$ for some φ.

[3] That is, if φ is an axiom of $\mathcal{Q}_1\mathcal{H}_i$, then so is $\forall\overrightarrow{x}\varphi$.

$(Ax11)$ $[\exists x\varphi \to \psi] \to \forall x(\varphi \to \psi)$, where $x \notin FV(\psi)$.

$(Ax12)$ $\exists x(\varphi \to \psi) \to [\forall x\varphi \to \psi]$, where $x \notin FV(\psi)$.

$(Ax13)$ $[\forall x\varphi \to \psi] \to \exists x(\varphi \to \psi)$, where $x \notin FV(\psi)$.

$(Ax14)$ $\forall x\varphi(x) \to \forall y\varphi(y)$, where y does not occur in $\varphi(x)$.

$(Ax15)$ $\exists x\varphi(x) \to \exists y\varphi(y)$, where y does not occur in $\varphi(x)$.

$(Ax16)$ $x = x$

$(Ax17)$ $x = y \to t(v_0 \dots v_{i-1}xv_{i+1} \dots v_n) = t(v_0 \dots v_{i-1}yv_{i+1} \dots v_n)$

$(Ax18)$ $x = y \to [R(v_0 \dots v_{i-1}xv_{i+1} \dots v_n) \to R(v_0 \dots v_{i-1}yv_{i+1} \dots v_n)]$

In $(Ax17)$,$(Ax18)$ the term t and the relation symbol R are from the predicate language in concern.

DEFINITION 2.2. $\mathcal{Q}_1$ is the proof system with inference rule MP and axiom schemes $(DT1)$, $(DT2)$, $(Ax1)$–$(Ax15)$ (and all their universal generalizations). A proof system $\mathcal{Q}$ is called an *axiomatic extension* of $\mathcal{Q}_1$ iff $\mathcal{Q}$ is obtained by adding some axiom schemes (and their universal generalizations) to $\mathcal{Q}_1$. Let $i \in \{1, 2\}$. $\mathcal{P}_i$ is the quantifier-free sublogic of $\mathcal{Q}_1\mathcal{H}_i$, i.e., $\mathcal{P}_i$ has axiom schemes of $\mathcal{H}_i$ (applying to quantifier-free, equality-free sentences). Similarly $\mathcal{P}_i^{=}$ has axiom schemes of $\mathcal{H}_i$ (applying to quantifier-free sentences) and all instantiations of $(Ax16)$, $(Ax17)$, $(Ax18)$ (replacing all variables x, y, v_j by closed terms).

3. A Herbrand-Henkin style proof of the classical model existence theorem for prenex normal form sentences

In this section[4] we prove the classical model existence theorem for prenex normal form sentences. In first order logic (FOL) it is clear that every sentence is provably equivalent to a prenex normal form sentence (w.r.t. FOL). Therefore, a proof of CME for prenex normal form sentences also shows that CME holds in FOL.

The proof idea is as follows: Given a $\perp$-consistent set Γ of prenex normal form sentences, we extend Γ into a $\perp$-consistent Γ' which fully witnesses all quantifiers in Γ by instantiating $\forall$ by *all* closed terms and instantiating $\exists$ by *new* constant symbols (new w.r.t. the construction stage). Let Γ'' be the quantifier-free part of Γ'. We then construct the canonical model of Γ'' at the quantifier-free level. Since this classical model is also a model of Γ'

---

[4] The idea of this section is originated from [5].

($\forall$ witnessed by all closed terms and $\exists$ witnessed by new constant symbols), it is a classical model of Γ and S.

For convenience we use the following setting. We assume that any first order language in concern contains at least one constant symbol. $\mathcal{L}(\Gamma)$ is the set of language symbols in Γ (including predicate symbols, function symbols, and (at least one) constant symbols). $Con_{\mathcal{Q}}(\Gamma)$ means that $\Gamma \not\vdash_{\mathcal{Q}} \perp$ in proof system $\mathcal{Q}$. $Sat(\Gamma)$ means that Γ has a classical model. "φ is a Π sentence" means that φ is of prenex normal form and it is a Π_m sentence for some $m \geq 0$. $\forall \overrightarrow{x}$ means $\forall x_1 \ldots \forall x_k$. $CT(\Gamma)$ is the set of all closed terms of $\mathcal{L}(\Gamma)$. Σ sentences, $\mathcal{L}(\varphi)$, $Con_{\mathcal{Q}}(\varphi)$, $Sat(\varphi)$, $\exists \overrightarrow{x}$, $CT(\varphi(\overrightarrow{x}))$ are defined in a similar way. The word "countable" means finite or infinitely countable.

First we consider two special cases of CME.

Theorem 3.1 (CME **for** Π_1 **and** Σ_1 **cases**). *Let* $\varphi(\overrightarrow{x})$ *be a quantifier-free formula (where* $\overrightarrow{x}$ *is* $x_1, \ldots, x_k$*) and the proof system* $\mathcal{Q}$ *be one of* $\mathcal{Q}_1\mathcal{H}_1$, $\mathcal{Q}_1^{=}\mathcal{H}_1$, $\mathcal{Q}_1\mathcal{H}_2$, $\mathcal{Q}_1^{=}\mathcal{H}_2$ *(described in Section 2).*

(a) *If* $\{\forall \overrightarrow{x} \varphi(\overrightarrow{x})\}$ *is* $\perp$*-consistent, then it has a countable classical model.*

(b) *If* $\{\exists \overrightarrow{x} \varphi(\overrightarrow{x})\}$ *is* $\perp$*-consistent, then it has a countable classical model.*

Proof. (a). Let $\Gamma = \{\forall \overrightarrow{x} \varphi(\overrightarrow{x})\}$ and $\Gamma' = \Gamma \cup \{\varphi(t_1, \ldots, t_k) \mid t_1, \ldots, t_k \in CT(\varphi(\overrightarrow{x}))\}$. Then $\Gamma'' = \Gamma' \setminus \Gamma$ is the quantifier-free part of Γ'. It is clear that $Con_{\mathcal{Q}}(\Gamma)$ implies $Con_{\mathcal{Q}}(\Gamma')$: If $\Gamma' \vdash_{\mathcal{Q}} \perp$, by using ($Ax1$) and MP, we get $\Gamma \vdash_{\mathcal{Q}} \perp$. The canonical model construction of Γ'', as presented in Theorem 3.2, has the domain consisting of all (equivalence classes of) closed terms. Since the universal quantifiers are witnessed by all closed terms (in Γ''), the canonical model of Γ'' is also a classical model of Γ.

(b). Let $\Gamma = \{\exists \overrightarrow{x} \varphi(\overrightarrow{x})\}$ and $\Gamma' = \Gamma \cup \{\varphi(d_1, \ldots, d_k)\}$, where $d_1, \ldots, d_k$ are new constant symbols. Then $Con_{\mathcal{Q}}(\Gamma)$ implies $Con_{\mathcal{Q}}(\Gamma')$: If $\Gamma' \vdash_{\mathcal{Q}} \perp$, by deduction theorem, replacing $\overrightarrow{d}$ by $\overrightarrow{x}$ (free variables not in Γ), universal generalization $\forall \overrightarrow{x}$ (a property of $\mathcal{Q}$, see Theorem 4.1), ($Ax10$), we have $\Gamma \vdash_{\mathcal{Q}} \perp$. Again the canonical model of Γ'' is also a classical model of Γ since the existential quantifiers are witnessed by $d_1, \ldots, d_k$. ∎

Theorem 3.2 (**Canonical model theorem for quantifier-free case**).
Consider proof systems $\mathcal{Q}_1\mathcal{H}_1$, $\mathcal{Q}_1^{=}\mathcal{H}_1$, $\mathcal{Q}_1\mathcal{H}_2$, $\mathcal{Q}_1^{=}\mathcal{H}_2$.

(1) *(Without equality:* $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$*) If* Γ'' *is a* $\perp$*-consistent set of quantifier-free sentences, then* Γ'' *has a classical model with domain* $CT(\Gamma'')$.

(2) *(With equality:* $\mathcal{Q}_1^{=}\mathcal{H}_1$, $\mathcal{Q}_1^{=}\mathcal{H}_2$*) If* Γ'' *is a* $\perp$*-consistent set of quantifier-free sentences, then* Γ'' *has a classical model with domain* $CT(\Gamma'')/\sim$, *where* $\sim$ *is an equivalence relation (on* $CT(\Gamma'')$*) obtained by the Lindenbaum maximal extension construction.*

Proof. Let $i \in \{1,2\}$. If $\Gamma'' \nvdash_{\mathcal{Q}_1\mathcal{H}_i} \bot$, since $\mathcal{P}_i$, the quantifier-free sublogic of $\mathcal{Q}_1\mathcal{H}_i$, is a subsystem of $\mathcal{Q}_1\mathcal{H}_i$, $\Gamma'' \nvdash_{\mathcal{P}_i} \bot$. (Similarly, $\Gamma'' \nvdash_{\mathcal{Q}_1^{=}\mathcal{H}_i} \bot$ implies $\Gamma'' \nvdash_{\mathcal{P}_i^{=}} \bot$.)

The proof is essentially the same as the propositional case. Enumerate all quantifier-free sentences of $\mathcal{L}(\Gamma'')$, say, $\psi_0, \psi_1, \ldots, \psi_n, \ldots$ Define $\Delta_0 = \Gamma''$,

$$\Delta_{n+1} = \begin{cases} \Delta_n \cup \{\psi_n\} & \text{if it is } \bot\text{-consistent,} \\ \Delta_n & \text{else.} \end{cases}$$

Let $\Delta = \bigcup_{i=0}^{\infty} \Delta_i$. One can easily show that (1) Δ is maximal $\bot$-consistent and $\Delta \vdash \alpha$ implies $\alpha \in \Delta$ for any quantifier-free α; (2) Δ is negation complete, i.e., exactly one of φ, $\varphi \to \bot$ is in Δ for any quantifier-free φ; and (3) the truth functionality of $\to$ holds in Δ. (This is because what $\mathcal{Q}_1\mathcal{H}_i$ does in the quantifier-free case is the same as what $\mathcal{H}_i$ does in the propositional case.)

In the case $\mathcal{Q}_1\mathcal{H}_i$ we define the classical model $\mathcal{M}$ with the domain $CT(\Gamma'')$ and, for any $\overrightarrow{a} \in CT(\Gamma'')$ and any relation symbol $R \in \mathcal{L}(\Gamma'')$, $\mathcal{M} \models R(\overrightarrow{x})[\overrightarrow{a}]$ iff $R(\overrightarrow{a}) \in \Delta$. In the case $\mathcal{Q}_1^{=}\mathcal{H}_i$, we take the domain to be $CT(\Gamma'')/\sim$, where $\sim$ is the equivalence relation defined by $s \sim t$ iff $(s = t) \in \Delta$. $\sim$ being an equivalence relation and both functions and relations being well-defined are proved by using the equality axioms $(Ax16)$, $(Ax17)$, $(Ax18)$. ■

Intuitively, for Σ_m with $m \geq 2$ we can introduce new constant symbols to witness the Σ_m sentences in Γ by adding some Π_{m-1} sentences. Similar to Theorem 3.1(b), the $\bot$-consistency is preserved. And for Π_m with $m \geq 2$ *it seems* that we can use "all" closed terms to witness Π_m sentences by adding some Σ_{m-1} sentences and the $\bot$-consistency is preserved (similar to Theorem 3.1(a)). *If all sentences are witnessed by quantifier-free sentences*, we can use Theorem 3.2 to construct a classical model.

But when $m \geq 2$, there are some problems in the Π_m cases. Consider the case with a Π_2 sentence $\forall \overrightarrow{x} \exists y \varphi(\overrightarrow{x}, y)$. When we first witness Π_2 sentences by all closed terms (Π_2 to Σ_1) and then witness Σ_1 sentences by new constant symbols (Σ_1 to Π_0), the closed terms used for witnessing $\forall \overrightarrow{x}$ are no longer "all the closed terms" because we add new constant symbols.

This problem can be resolved by repeating above procedure countably many times! At level 0 we define $\Gamma_{0,0} = \{\forall \overrightarrow{x} \exists y \varphi(\overrightarrow{x}, y)\}$ and $\mathcal{L}_0 = \mathcal{L}(\Gamma_{0,0})$. We enlarge $\Gamma_{0,0}$ to $\Gamma_{0,1} = \Gamma_{0,0} \cup \{\exists y\, \varphi(t_1, \ldots, t_k, y) \mid t_1, \ldots, t_k$ are closed terms of $\mathcal{L}_0\}$. $\Gamma_{0,2} = \Gamma_{0,1} \cup \{\varphi(t_1, \ldots, t_k, c_{f(t_1,\ldots,t_k)}) \mid t_1, \ldots, t_k$ are introduced at previous stage and $\exists y\, \varphi(t_1, \ldots, t_k, y) \in \Gamma_{0,1}$ and $c_{f(t_1,\ldots,t_k)}$ is a new constant symbol$\}$. Then at level 1 we set $\Gamma_{1,0} = \bigcup_{i=0}^{2} \Gamma_{0,i}$ and $\mathcal{L}_1 = \mathcal{L}(\Gamma_{1,0})$.

By repeating the same procedure (similar to level 0) we can generate $\Gamma_{n,0}$ and $\mathcal{L}_n$ for any $n \in \mathbb{N}$. Finally we take $\Gamma' = \bigcup_{n\in\mathbb{N}} \Gamma_{n,0}$ and $\mathcal{L} = \bigcup_{n\in\mathbb{N}} \mathcal{L}_n$. It is clear that in Γ' the universal quantifiers are witnessed by all closed terms of $\mathcal{L}$ and the existential quantifiers are witnessed by new constant symbols.

Note that in general the set of prenex normal form sentences Γ may have no finite upper bound on the complexity of quantifier alternation. Therefore in general we need to enlarge $\Gamma_{n,0}$ through $\Gamma_{n,i}$ for all $i \in \mathbb{N}$ and $\Gamma_{n+1,0} = \bigcup_{i\in\mathbb{N}} \Gamma_{n,i}$. Rigorously speaking, for $n \geq 0$ we define $\Gamma_{0,0} = \Gamma$, $\Gamma_{n+1,0} = \bigcup_{i\geq 0} \Gamma_{n,i}$, $\mathcal{L}_n = \mathcal{L}(\Gamma_{n,0})$, $\Gamma_{n,2m+1} = \Gamma_{n,2m} \cup \{\varphi(t_1, \ldots, t_k) \mid \varphi(t_1, \ldots, t_k)$ is a Σ sentence, $\forall \overrightarrow{x} \varphi(\overrightarrow{x}) \in \Gamma_{n,2m} \setminus \Gamma_{n,2m-1}$, and $t_1, \ldots, t_k \in CT(\mathcal{L}_n)\}$ and $\Gamma_{n,2m+2} = \Gamma_{n,2m+1} \cup \{\varphi(t_1, \ldots, t_k, c_{f_1(t_1,\ldots,t_k)}, \ldots, c_{f_l(t_1,\ldots,t_k)}) \mid$ it is a Π sentence, and $\exists \overrightarrow{x} \varphi(t_1, \ldots, t_k, \overrightarrow{x}) \in \Gamma_{n,2m+1} \setminus \Gamma_{n,2m}, t_1, \ldots, t_k$ are introduced for this sentence previously, and $f_1, \ldots, f_l$ are the corresponding Skolem function symbols for indexing new constant symbols$\}$. By repeating this procedure to generate $\mathcal{L}_n, \Gamma_{n,m}$ for $n, m \in \mathbb{N}$, we have a $\bot$-consistent set $\Gamma' = \bigcup_{n,m>0} \Gamma_{n,m}$ and Γ'' is the quantifier-free part of Γ'. Now by Theorem 3.2, we have:

THEOREM 3.3 (*CME* **for prenex normal form sentences**). *Let the proof system* $\mathcal{Q}$ *be one of* $\mathcal{Q}_1\mathcal{H}_1$, $\mathcal{Q}_1^{=}\mathcal{H}_1$, $\mathcal{Q}_1\mathcal{H}_2$, $\mathcal{Q}_1^{=}\mathcal{H}_2$. *If* Γ *is a* $\bot$*-consistent set of prenex normal form sentences, then* Γ *has a classical model.*

Remark 3.4. The difference between our approach and the standard Henkin construction (or Hintikka set construction) is that our construction from S through Γ to Γ' does not need to use *AC* (we only need the existence of inductive sets). The standard Henkin construction, due to the mix-up of quantifier cases and propositional cases, makes less clear that the quantifier cases can be managed without *AC*. However, it is not avoidable to use at least a weak version of *AC* at the quantifier-free stage for model construction.

Remark 3.5. The Skolem function indexing on constant symbols gives us not only an easy way to manage new constant symbols, but also an easy way to construct the Herbrand universe for canonical model. Surely it is not really necessary to index constant symbols this way for proving *CME* in first order logic (Henkin's method is mathematically successful). However, we should illustrate this by the following example: Assume that $\mathcal{L}_0$ has one ternary function f, one unary function g, two constant symbols d_1, d_2, and the Skolem functions introduced for indexing constant symbols are two unary function symbols h_1, h_2. Then the Herbrand universe is the free algebra D generated by $\{d_1, d_2, f, g, h_1, h_2\}$ and the domain is either D or $D/\sim$ (when we deal with logic with equality and $\sim$ is the equivalence relation obtained by the Lindenbaum maximal extension construction). Then the closed term $f(d_1, c_{h_1(g(d_2))}, c_{h_2(c_{h_2(d_1)})})$ is interpreted as the equivalence class $[f(d_1, h_1(g(d_2)), h_2(h_2(d_1)))]$ in $D/\sim$, where $f(d_1, h_1(g(d_2)), h_2(h_2(d_1))) \in D$ is obtained by removing all c and pulling up all Skolem terms from the lower right corners.

Remark 3.6. CME states that consistency implies satisfiability. Consider the sentence $\forall\overrightarrow{x}\exists y\varphi(\overrightarrow{x}, y)$. By AC, $Sat(\forall\overrightarrow{x}\exists y\varphi(\overrightarrow{x}, y))$ holds if and only if $Sat(\forall\overrightarrow{x}\varphi(\overrightarrow{x}, f(\overrightarrow{x})))$ holds. It seems that to prove CME is as easy as the Π_1 case (Theorem 3.1(a)) if we can introduce Skolem function symbols to eliminate all existential quantifiers.

However, in this approach for proving CME it seems not avoidable to prove at first the preservation of $\bot$-consistency from $\{\forall\overrightarrow{x}\exists y\varphi(\overrightarrow{x}, y)\}$ to $\{(\forall\overrightarrow{x}\varphi(\overrightarrow{x}, f(\overrightarrow{x}))\}$. A direct proof of it is not trivial, though feasible. In [10] (pp. 55–56) the "Theorem on Functional Extensions" (it states, in our terminology, that "If $T \vdash \exists y\varphi(\overrightarrow{x}, y)$, f is a new function symbol, and $f(\overrightarrow{x})$ is free for y in $\varphi(\overrightarrow{x}, y)$, then $T \cup \{\varphi(\overrightarrow{x}, f(\overrightarrow{x}))\}$ is a conservative extension of T.") implies the preservation of $\bot$-consistency.

In proving Theorem on Functional Extensions in [10], Herbrand's theorem plays a crucial role. And the proof of Herbrand's theorem seems to require some careful proof theoretic analysis (either as in [10] (pp. 48–55), or by cut-elimination theorem).

Remark 3.7. One may also try to prove CME via second order logic. Apply the axiom AC in second order logic $\forall\overrightarrow{x}\exists y\varphi(\overrightarrow{x}, y) \to \exists f\forall\overrightarrow{x}\varphi(\overrightarrow{x}, f(\overrightarrow{x}))$ and then repeat the argument in Theorem 3.1(b) by introducing new (second order) function constant f. The problem is that one needs to show that the preservation of $\bot$-consistency of S from FOL to some second order deductive system ($FOL + AC$ or some deductive system including AC, for example, $D2$ in [9]), and this seems not easier.

4. Prenex normal form theorem holds in logics weaker than first order logic

In this section we prove that the prenex normal form theorem holds in any axiomatic extension of $\mathcal{Q}_1$ (described in Section 2). Since we do not use the connective $\wedge$, we define the $\mathcal{Q}$-provable equivalence as follows: φ and ψ are *provably equivalent* in $\mathcal{Q}$ iff $\vdash_{\mathcal{Q}} \varphi \to \psi$ and $\vdash_{\mathcal{Q}} \psi \to \varphi$.

Let $FV(\Sigma)$ be the collection of all free variables in φ for some $\varphi \in \Sigma$. We also take the following convention: In any derivation of $\Sigma \vdash_{\mathcal{Q}} \varphi$, the universal generalization on variable x must satisfy that $x \notin FV(\Sigma)$. And to avoid the trouble that Σ uses all free variables, we take the view that "There are always countably many un-used free variables (to Σ)."

Theorem 4.1. *Let $\mathcal{Q}$ be any axiomatic extension of $\mathcal{Q}_1$ (described in Section 2). If $\Sigma \vdash_{\mathcal{Q}} \varphi$ and $x \notin FV(\Sigma)$, then $\Sigma \vdash_{\mathcal{Q}} \forall x\varphi$.*

Theorem 4.1 is proved by induction using axioms $(Ax3)$,$(Ax4)$ and the fact that the universal generalization of an axiom in $\mathcal{Q}$ is also an axiom in $\mathcal{Q}$. Note that this type of axiomatization is proposed in [4], [11].

We then modify the construction in [8] (pp. 48–49) to prove the prenex normal form theorem. By taking axioms $(Ax14)$,$(Ax15)$, $\mathcal{Q}_1$ allows the *schema of alphabetic change of a bound variable*: $\vdash_{\mathcal{Q}_1} \forall x\varphi(x) \to \forall y\varphi(y)$ and $\vdash_{\mathcal{Q}_1} \forall y\varphi(y) \to \forall x\varphi(x)$, where y does not occur in $\varphi(x)$ (and similarly the $\exists$ case). Again this holds for any axiomatic extension of $\mathcal{Q}_1$.

Next we have the following schema of substitutivity of equivalence.

Theorem 4.2 (**Substitutivity of Equivalence**). *Let $\mathcal{Q}$ be any axiomatic extension of $\mathcal{Q}_1$ and $\varphi(\overrightarrow{x})$, $\psi(\overrightarrow{x})$ be formulas with free variables among $\overrightarrow{x}$. Suppose that B results from A by replacing zero or more occurrences of $\varphi(\overrightarrow{x})$ in A by $\psi(\overrightarrow{x})$. Then*

$$\vdash_{\mathcal{Q}} \forall\overrightarrow{x}[\varphi(\overrightarrow{x}) \to \psi(\overrightarrow{x})] \to \{\forall\overrightarrow{x}[\psi(\overrightarrow{x}) \to \varphi(\overrightarrow{x})] \to [A \to B]\}$$

and

$$\vdash_{\mathcal{Q}} \forall\overrightarrow{x}[\varphi(\overrightarrow{x}) \to \psi(\overrightarrow{x})] \to \{\forall\overrightarrow{x}[\psi(\overrightarrow{x}) \to \varphi(\overrightarrow{x})] \to [B \to A]\}$$

Note that the proof of Theorem 4.2 is done by induction using Theorem 4.1 and $(DT1)$, $(DT2)$, $(Ax1)$, $(Ax4)$, $(Ax5)$.

Note that in $\mathcal{Q}_1$ the existential quantifier $\exists$ is not equivalent to $\neg\forall\neg$ (according to the three-valued semantics at the end of this section).

Finally we have the prenex normal form theorem.

Theorem 4.3. *Let $\mathcal{Q}$ be any axiomatic extension of $\mathcal{Q}_1$. In $\mathcal{Q}$ every formula is provably equivalent to a formula in prenex normal form.*

The proof is done by

1. Change all bound variables so that they are all distinct (using Theorem 4.2 and the schema of alphabetic change of a bound variable).
2. Use $(Ax6)$, $(Ax7)$, $(Ax8)$, $(Ax9)$, $(Ax10)$, $(Ax11)$, $(Ax12)$, $(Ax13)$ to proof-theoretically convert the formula into a prenex normal form formula.

Since $\mathcal{H}_1$ and $\mathcal{H}_2$ both satisfy propositional CME, according to Section 3 we have:

Theorem 4.4. *In the following four systems $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$,$\mathcal{Q}_1^{=}\mathcal{H}_2$, the classical model existence theorem for $\perp$-consistency holds.*

At the end of this section we show that $\mathcal{Q}_1^{=}\mathcal{H}_1$ is weaker than FOL (so are the other three systems). To do so we use the following three-valued semantics. Let the truth values be 0,1,2. We take 2 for designated value. $0 < 1 < 2$ and the value of $\forall x\varphi(x)$ is the minimum of all the values of $\varphi(d)$ with d ranging over the domain of the intended three-valued model M. $\exists$ is evaluated similarly by taking the maximum value. The truth value of $\bot$ is 0 and the truth function f for $\rightarrow$ is:

f	0	1	2
0	2	2	2
1	0	2	2
2	0	1	2

It is a tedious but feasible work to check that all axioms in $\mathcal{Q}_1$ are tautologies (always with truth value 2) with respect to above three-valued semantics. However, $\neg\forall x\neg\varphi(x) \rightarrow \exists x\varphi(x)$ and $\neg\neg\psi \rightarrow \psi$ are not tautologies. To deal with equality, we interpret $(x = y)[a, a]$ the value 2 for any a in the domain of discourse and $(x = y)[a, b]$ the value 0 for distinct a, b in the domain of discourse. With these ($Ax16$), ($Ax17$), ($Ax18$) are tautologies in above three-valued semantics.

5. Concluding remarks

In this section we discuss the (modified) Kripke models of $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$, and the relationship of the Kripke model for $\Sigma \not\models \varphi$ and the classical model for Σ.

We modify the definition of Kripke models (in [1], p. 46) by dropping the conditions of $\wedge$,$\vee$ (because we are considering $\{\rightarrow, \bot\}$-fragment), and adding conditions given by axioms ($Ax9$), ($Ax13$). Note that ($Ax9$), ($Ax13$) are axioms not true in Kripke models for first order intuitionistic logic.

Now we modify (see [1], p. 67) the definition that a set Γ is *nice* with respect to P by dropping the condition that Γ has the *Or*-property. Then the modified Lemma 10.3, Lemma 10.4 (3), (4), (5), (6), Lemma 10.5 (in [1], pp. 67–69) can be proved in $\mathcal{Q}_1\mathcal{H}_1$ and $\mathcal{Q}_1^{=}\mathcal{H}_1$. With these we can prove the extended completeness theorem for these modified Kripke semantics.

Then what is the relationship between $\Gamma \nvdash \bot$ and $\Gamma \nvdash \varphi$? Assume that Γ is finite. In the propositional case $\mathcal{H}_1$, if $\Gamma \nvdash \bot$, by filtration (considering all subformulas of $\Gamma \cup \{\varphi\}$) there is a finite frame Kripke model in which there is a w such that $w \Vdash \Gamma$ and $w \nVdash \varphi$. Since the persistency of true sentences in Kripke model, from w we can reach an end node in the finite frame Kripke model, which is a classical model of Γ (this is probably due to Jaśkowski). If $\Gamma \cup \{\neg\varphi\} \nvdash \bot$, the classical model of $\Gamma \cup \{\neg\varphi\}$ is also a Kripke model for

$\Gamma \not\models \varphi$ (with single world). However, this will not always happen: for some φ we have $\neg\neg\varphi \nvdash \varphi$ and $\{\neg\neg\varphi, \neg\varphi\} \vdash \bot$. When Γ is infinite, the classical model can be reached only through infinite process (similar to compactness theorem).

In $\mathcal{Q}_1\mathcal{H}_1$ or $\mathcal{Q}_1^{=}\mathcal{H}_1$, if $\Gamma \nvdash \varphi$, we may assume that Γ is in prenex normal form. Then by the method in Section 3, we can extend Γ to Γ' such that $\Gamma' \nvdash \varphi$. But now the quantifier-free part of Γ' can be managed as the infinite case in $\mathcal{H}_1$. The classical model of Γ can be constructed this way (with single world). If fortunately we have $\Gamma \cup \{\neg\varphi\} \nvdash \bot$, this classical model is also a Kripke model for $\Gamma \not\models \varphi$ (though this will not always work).

There are at least three more remarkable facts from the study of CME. Firstly, Glivenko's theorem holds in $\mathcal{Q}_1\mathcal{H}_1$,$\mathcal{Q}_1^{=}\mathcal{H}_1$,$\mathcal{Q}_1\mathcal{H}_2$,$\mathcal{Q}_1^{=}\mathcal{H}_2$. Let $\mathcal{Q}$ be any of them. If $\Sigma \vdash \varphi$ in FOL but $\Sigma \nvdash \neg\neg\varphi$ in $\mathcal{Q}$, then $\Sigma \cup \{\neg\varphi\} \nvdash \bot$. By CME with respect to the $\bot$-consistency, there is a classical model for $\Sigma \not\models \varphi$, a contradiction.

Secondly, for any given set Γ of sentences we can always convert it into prenex normal form and instantiate it as we do in Section 3. Since the quantifier-free part of Γ' has a classical model if and only if Γ is $\bot$-consistent, we can conclude that whether Γ is $\bot$-consistent is a matter of quantifier-free issue (with adding countably many new constant symbols).

Thirdly, we can easily extend $\mathcal{Q}_1\mathcal{H}_1$ by adding the corresponding axioms for $\wedge, \vee$ in IPL (axioms 3–8 in [1], p. 63) and the corresponding prenex-normal-form axioms for $\wedge, \vee$: $Qx(\varphi * \psi) \to [(Qx\varphi) * \psi]$ and $[(Qx\varphi) * \psi] \to Qx(\varphi * \psi)$, where $Q \in \{\forall, \exists\}$, $* \in \{\wedge, \vee\}$, and $x \notin FV(\psi)$. This logic satisfies prenex normal form theorem and CME. If we add one more axiom (A6) (in [2], p. 36) into it, it becomes an extension of Gödel logic $G\forall$ (by adding $(Ax9)$, $(Ax13)$ to $G\forall$), which is still subclassical (by the three-valued semantics in Section 4 together with interpreting $\wedge$ by *min* and $\vee$ by *max*).

Acknowledgements. I wish to thank the following people for their valuable suggestions: Arnon Avron, Petr Cintula, an anonymous referee. I also wish to thank the travel grant support of NSC in Taiwan.

References

[1] Fitting, M., *Intuitionistic Logic, Model Theory and Forcing*, North-Holland Publishing Co., Amsterdam-London, 1969.

[2] Hájek, P., *Metamathematics of Fuzzy Logic, Trends in Logic—Studia Logica Library, vol. 4*, Kluwer Academic Publishers, Dordrecht, 1998.

[3] Henkin, L., 'The discovery of my completeness proofs', *Bulletin of Symbolic Logic*, 2: 127–158, 1996.

[4] HUNTER, G., *Metalogic. An Introduction to the Metatheory of Standard First Order Logic*, University of California Press, 1973.

[5] LEE, J.-L., 'On Gödel's Completeness Theorem', presented in *12th International Congress of Logic Methodology and Philosophy of Science*, at Oviedo, Spain, August 2003.

[6] LEE, J.-L., 'Classical model existence theorem in propositional logics', in Béziau, Jean-Yves, Costa-Leite, Alexandre (eds.), *Perspectives on Universal Logic*, Polimetrica, Monza, Italy, 2007, pp. 179–197.

[7] LEE, J.-L., *Classical model existence and resolution*, manuscript, 2007.

[8] ROBBIN, J.W., *Mathematical Logic*, W.A. Benjamin, Inc., New York-Amsterdam, 1969.

[9] SHAPIRO, S., 'Classical logic II: Higher-order logic', in Goble, L. (ed.), *The Blackwell Guide to Philosophical Logic*, Blackwell Publishers Ltd., 2001, pp. 33–54.

[10] SHOENFIELD, J.R., *Mathematical Logic*, Addison-Wesley Publishing Co., Reading, MA, 1967.

[11] SUNDHOLM, G., 'Systems of deduction', in Gabbay, D.M., Guenthner, F. (eds.), *Handbook of Philosophical Logic, vol. 2*, Kluwer Acad. Publ., Dordrecht, 2001, pp. 1–52.

[12] TENNANT, N., 'Minimal logic is adequate for popperian science', *The British Journal for the Philosophy of Science*, 36: 325–329, 1985.

JUI-LIN LEE
Department of Philosophy
National Chung-Cheng University
Chia-Yi, Taiwan
`jlleelogician@gmail.com`

Wojciech Zielonka

Weak Implicational Logics Related to the Lambek Calculus — Gentzen versus Hilbert Formalisms

Abstract. It has been proved by the author that the product-free Lambek calculus with the empty string in its associative (L_0) and non-associative (NL_0) variant is not finitely Gentzen-style axiomatizable if the only rule of inference is the cut rule. We give here rather detailed outlines of the proofs for both L_0 and NL_0. In the last section, Hilbert-style axiomatics for the corresponding weak implicational calculi are given.

Keywords: Lambek calculus, implicational logics, finite axiomatizability.

1. Introduction

The calculus of syntactic types invented by Joachim Lambek [5], [6] almost fifty years ago still attracts the attention of linguists, logicians, mathematicians and computer scientists. Since then, numerous variants of the calculus were invented. In this paper, we are interested in some of them.

For the sake of uniformity, we start with considering the sequential (Gentzen-style) formulation of the calculi. Lambekian sequents have always the form $X \to x$ where the succedent x is a single type and the antecedent X is, generally speaking, a string of types. The axioms are all sequents of the form $x \to x$ and the rules of inference, following Gentzen, introduce type-forming functors either in the antecedent or in the succedent. Now, particular calculi may differ in the following respects:

- The choice of type-forming functors from among the three original Lambekian ones: \ (left division), / (right division) and · (product). In particular, product-free calculi turn out to form an important class, mainly in linguistic applications. All the systems in question will be product-free.
- The admissibility of the empty sequent antecedent (the empty string, to be short). Our systems do admit it. However, a number of results on those which do not will be mentioned below.

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*, Trends in Logic, 28

- Structural (like commutativity) and substructural (like associativity) properties of the operation of building sequent antecedents out of types. Both Lambek's systems, the associative [5] and the non-associative [6] do not employ structural rules peculiar to Gentzen. In the last section, we shall go a little beyond the substructural domain.

Various kinds of semantics have been proposed for the Lambek calculus. We shall concentrate on two of them.

Lambek himself is motivated linguistically: his types are to be interpreted as sets of expressions with the product and both divisions standing respectively for the complex concatenation and its both reverse operations. The sequent $x_1 \dots x_n \to y$ means "every concatenation of expressions of the types $x_1, \dots, x_n$ in that order belongs to the type y". Since the empty string has little linguistic application, Lambek does not allow it.

Another interpretation, of rather logical than linguistic kind, follows the spirit of the Gentzen formalism. Here, Lambekian functors (resp. primitive types, types, sequents) are interpreted as propositional connectives (resp. variables, formulas, rules of inference) of linear logic (to be more precise, its non-commutative intuitionistic multiplicative fragment; see [1]). Under this interpretation, $\backslash$ and $/$ correspond respectively to the right ($\to$) and left ($\leftarrow$) implication. In absence of commutativity, $\backslash$ and $/$ (and thus $\to$ and $\leftarrow$) are not mutually definable. The empty string is now allowed since otherwise there would be rules but no axioms.

All the syntactic calculi are known to be closed w. r. t. the (appropriately formulated) cut rule whence their decidability easily follows. For a long time, attempts were made to find for them axiomatics based on the cut as the only rule of inference, by adding sufficiently many axioms instead of Gentzenian functor introduction rules. The problem has a trivial solution: one may add all derivable sequents as new axioms. In view of the decidability, such axiomatization is even recursive. What was really looked for were *finite* cut-rule axiomatics. According to what was stated in preceding paragraphs, this research was motivated in a twofold way: logically or linguistically depending on whether the system admits the empty string or not. More about this motivation may be found in the introductory section of [12].

Before going into details, we introduce some notation. By *NL*, *L*, *LP* we denote respectively the nonassociative, associative, and associative-commutative Lambek calculus without the empty string. The subscript 0 (thus NL_0, L_0, LP_0) means that the empty string is allowed.

The first attempt was made by Cohen [3] who (wrongly) identified *L* with the system based on the cut and the following axiom schemata:

(1)	$x/y\ y \rightarrow x$	(1')	$y\ y\backslash x \rightarrow x$
(2)	$y \rightarrow x/(y\backslash x)$	(2')	$y \rightarrow (x/y)\backslash x$
(3)	$x/y\ y/z \rightarrow x/z$	(3')	$z\backslash y\ y\backslash x \rightarrow z\backslash x$
(4)	$(x\backslash y)/z \rightarrow x\backslash(y/z)$	(4')	$x\backslash(y/z) \rightarrow (x\backslash y)/z$

In [8], I gave the nonfinite-axiomatizability proof for L. The research was continued in the late 80's together with Kandulski ([4], a similar negative result for NL) and Buszkowski ([2], solutions for LP (negative) and LP_0 which was shown to be the only finitely cut-rule axiomatizable variant). In [9], negative results were established for unidirectional (i.e., involving only one division functor) fragments of L_0 and NL_0. For the whole of both latter systems, the problem turned out to be particularly difficult and it remained unsolved until the end of the millenium. The solution for L_0 was finally published in [11] and [12], and that for NL_0 in [10] and [13]. In the next two sections, we give their outlines.

2. Preliminaries

We start with presenting the calculi L_0 and NL_0 in their Gentzen-style form, as given by Lambek respectively in [5] and [6]. Next, we discuss briefly the finite cut-rule axiomatizability problem for both systems.

We define the set Tp of *types*:

- *Primitive types* $p_1, p_2, \ldots$ are types.
- If x and y are types, so are (x/y) and $(x\backslash y)$ (as usual, we omit the outermost parentheses).

We use lowercase (resp. capital) letters to denote elements of Tp (resp. Tp*). *Sequents* are expressions of the form $X \rightarrow x$ (unlike Lambek, we admit sequents with the empty antecedent).

Axioms of L_0: all sequents of the form $x \rightarrow x$.

Rules of L_0:

$$(\mathrm{R1}')\quad \frac{T \rightarrow y \quad UxV \rightarrow z}{Ux/yTV \rightarrow z} \qquad\qquad (\mathrm{R1}'')\quad \frac{T \rightarrow y \quad UxV \rightarrow z}{UTy\backslash xV \rightarrow z}$$

$$(\mathrm{R2}')\quad \frac{Ty \rightarrow x}{T \rightarrow x/y} \qquad\qquad (\mathrm{R2}'')\quad \frac{yT \rightarrow x}{T \rightarrow y\backslash x.}$$

Cut elimination theorem (Lambek [5]). L_0 *is closed w.r.t. the* cut rule*:*

$$(\mathrm{CUT})\quad \frac{T \rightarrow x \quad UxV \rightarrow y}{UTV \rightarrow y.}$$

The calculus NL_0 differs from L_0 in that the sequent antecedents are not (unstructured) strings but *bracketed strings of types* (BSTp's) defined as follows:

- The empty string is a BSTp.
- Types are BSTp's.
- If X and Y are nonempty BSTp's, then $[XY]$ is a BSTp (as usual, we omit the outermost brackets).

Substrings of a nonempty BSTp are defined in a natural way: the type x is its only substring; substrings of $[XY]$ are those of X and Y as well as $[XY]$ itself. The rules (R1) and (CUT) in NL_0 have the form

$$(\mathrm{R1}') \quad \frac{T \to y \quad Y[x] \to z}{Y[x/yT] \to z} \qquad\qquad (\mathrm{R1}'') \quad \frac{T \to x \quad Y[x] \to z}{Y[Ty\backslash x] \to z}$$

$$(\mathrm{CUT}) \quad \frac{T \to x \quad Y[x] \to y}{Y[T] \to y.}$$

Here, as well as in (R2) which do not change formally, Y and T denote BSTp's. $Y[x]$ is a BSTp in which x occurs as a substring. If X is nonempty or $Y[x] = x$, then $Y[X]$ results from $Y[x]$ by substitution of the BSTp X for one such occurrence of x. If $Y[x] \neq x$, then $Y[x]$ clearly has a substring of the form $[xZ]$ or $[Zx]$ with Z nonempty. Thus, we define $Y[]$ as a result of the substitution of Z for one such occurrence of $[xZ]$ or $[Zx]$ in $Y[x]$.

The cut elimination theorem holds also for NL_0 (see Lambek [6]).

It should be remembered that, in both L_0 and NL_0, we admit empty sequent antecedents which are not allowed in Lambek's original systems.

In the presence of (CUT), one may replace (R1) by the axioms

$$(\mathrm{A1'}) \ \ x/y\, y \to x \qquad\qquad (\mathrm{A1''}) \ \ y\, y\backslash x \to x$$

In fact, in L_0 we have

$$\frac{\dfrac{T \to y \quad x/y\, y \to x}{x/y\, T \to x} \qquad UxV \to z}{Ux/yTV \to z}$$

and similarly for (R1”). In NL_0, the proof is essentially the same.

One may ask whether it is possible to replace (R2) by finitely many axiom schemata in a similar way. The answer is negative for both L_0 and NL_0.

A standard method of nonfinite-axiomatizability proofs is based on a simple criterion due to Tarski: a deductive system is not finitely axiomatizable (under a given set of rules of inference) if it is the union of a chain (S_n) of systems such that, for every n, S_{n+1} is a proper extension of S_n. Thus, in order to prove the aforementioned results, one must do two things:

- firstly, find a (clearly infinite) axiomatics based on (CUT) as the only rule of inference for the system in question (L_0 or NL_0) and divide its axioms into "ranks" numbered with natural numbers;
- secondly, prove that the axioms of a given rank are not derivable from those of inferior ranks.

In sections 3 and 4 below, we shall show how to do that respectively for L_0 and NL_0.

3. The associative case

Let C be the formal system whose only rule is (CUT) and whose axioms are defined as follows:

- Axioms of rank 0: all sequents of the form
  (A1′) $x/y\ y \to x$ (A1″) $y\ y\backslash x \to x$
- Axioms of rank 1: all sequents of the form
  (A2′) $\to x/x$ (A2″) $\to x\backslash x$
  (A3′) $\to (x/(y\backslash x))/y$ (A3″) $\to y\backslash((x/y)\backslash x)$
  (A4′) $\to ((x/z)/(y/z))/(x/y)$ (A4″) $\to (y\backslash x)\backslash((z\backslash y)\backslash(z\backslash x))$
- Axioms of higher ranks: if $\to y$ is an axiom of rank n, then all sequents of the form
  (A5′) $\to x/(x/y)$ (A5″) $\to (y\backslash x)\backslash x.$
  are axioms of rank $n+1$.

We denote by C_n be the system C restricted to the axioms of rank $\leq n$. We write "$X \vdash_n x$" instead of "$X \to x$ is derivable in C_n".

THEOREM 3.1 ([11]). *L_0 and C are equivalent.*

PROOF. Using Lambek's decision method [5], we verify that axioms of C are derivable in L_0. In view of the cut elimination theorem, C is a subsystem of L_0. To prove the converse, we show first that the axiom-forming rules (A5) are admissible rules of C in the sense that they yield a theorem of C_{n+1} when applied to a theorem of C_n. It follows easily that (R2) have the same property. (R1) may be directly derived in C as shown in Section 1. The axiom sequents $x \to x$ of L_0 follow from (A1), (A2) and (CUT). ■

It remains to be proved that C_{n+1} is a proper extension of C_n. For this purpose, we need some intermediate results. For every $n > 1$, let C_n^R be the system whose only rule is (CUT) and whose axioms are

(3A')	$y \to x/(y\backslash x)$	(3A")	$y \to (x/y)\backslash x$
(4A')	$x/y \to (x/z)/(y/z)$	(4A")	$y\backslash x \to (z\backslash y)\backslash(z\backslash x)$
(5A')	$x/y \to x$ whenever $\to y$ is an axiom of C_{n-1}	(5A")	$y\backslash x \to x$ whenever $\to y$ is an axiom of C_n
(4B)	$y/z \to x/z$ whenever $y \to x$ is an axiom of C_n^R.		

For $n = 1$, the axioms are (3A'), (4A'), (4B), and

(3B) $y\backslash x \to x$ whenever $\to y$ is derivable in C_n.

THEOREM 3.2 ([12]). *Let* $n > 0$. *If* $\vdash_n u/v$ *and* $u \neq v$, *then* $v \to u$ *is derivable in* C_n^R.

PROOF. Induction on derivations in C_n^R and in some auxiliary systems related to C_n and C_n^R. The details may be found in [12] (Theorems 4, 5 and 6). ■

We define the type y/X by induction on the length of the string X of types: $y/() = y$; $y/(Xx) = (y/x)/X$. An axiom of C_n^R is said to be *Montagovian* if it has the form $y/Z \to x/Z$ with $y \to x$ an instance of (3A). Clearly, a Montagovian axiom is (3A) if Z is empty and (4B) otherwise.

We define the *head* of the type x as follows: head$(p) = p$ for p primitive; head$(x/y) =$ head$(y\backslash x) =$ head(x). Now, Montagovian axioms have the property that the antecedent and the succedent may have different heads. This is the main source of technical difficulties in the proof of Theorem 3.3 below.

For $n = 1, 2, \ldots$, we define S_n to be the least set such that

- all primitive types are in S_n,
- if $x \in S_n$ and $\vdash_{n-1} y$, then $x/y \in S_n$,
- if $x \in S_n$ and $\vdash_n y$, then $y\backslash x \in S_n$.

THEOREM 3.3 ([12]). *If* $n > 1$, $\vdash_n s/t$ *and* s *is primitive, then* $t \in S_n$.

PROOF. It is easy to see that the derivation of $x \to y$ in C_n^R may be represented in the form of a *reduction*, i.e., in the form

$$x_0 \to x_1 \to \ldots \to x_m$$

where $x_0 = x$, $x_m = y$ and $x_{k-1} \rightarrow x_k$ $(0 < k \leq m)$ is an axiom of C_n^R. The axiom $x_{k-1} \rightarrow x_k$ is the k-th *stage* of the reduction and the number m is its *length.* By Theorem 3.2, there exists a reduction of t to s in C_n^R. There are two cases:

(1) The reduction has no Montagovian stages. Then the proof proceeds easily by induction on its length.

(2) The contrary holds. Define the *complexity* of an axiom (5A) to be the number of occurrences of "/" and "\" in the type denoted by y in the axiom schema (5A), and the complexity of a reduction to be the sum of the complexities of all its stages (5A). We prove that whenever the reduction has a Montagovian stage, its complexity may be lowered. Thus, (2) may be reduced to (1). ■

COROLLARY 3.1. *Let s be primitive. If $n > 1$ and $\vdash_n s/(s/x)$, then $\vdash_{n-1} x$.*

Let s be primitive. Define the sequence (y_n) of types as follows: $y_0 = s/s$, $y_{n+1} = s/(s/y_n)$. it is easy to show that $\not\vdash_0 y_0$ and $\not\vdash_1 y_1$. Hence $\not\vdash_n y_n$ by the corollary and induction on n. On the other hand, $\rightarrow y_n$ is an axiom of C_{n+1}. It follows that C_{n+1} is a proper extension of C_n which was to be proved.

4. The non-associative case

Let NC be the formal system whose only rule is (CUT) and whose axioms are defined as follows:

- Axioms of rank 0 an 1: like in C_n, with (A4) replaced by

  (A4') $\rightarrow (((x/y)\backslash x)/z)/(y/z)$ (A4") $\rightarrow (z\backslash y)\backslash(z\backslash(x/(y\backslash x)))$

  (A5') $\rightarrow (z/((x/y)\backslash x))\backslash(z/y)$ (A5") $\rightarrow (y\backslash z)/((x/(y\backslash x))\backslash z)$

- If $\rightarrow y$ is an axiom of rank n, so are

  (A6') $\rightarrow (x/z)\backslash((y\backslash x)/z)$ (A6") $\rightarrow (z\backslash(x/y))\backslash(z\backslash x)$

  (A7') $\rightarrow (z/x)\backslash(z/(y\backslash x))$ (A7") $\rightarrow ((x/y)\backslash z)/(x\backslash z)$

- If $\rightarrow y$ is an axiom of rank n, then

  (A8') $\rightarrow x/(x/y)$ (A8") $\rightarrow (y\backslash x)\backslash x$

  are axioms of rank $n+1$.

We denote by NC_n the system NC restricted to the axioms of rank $\leq n$. We write "$X \vdash_n x$" instead of "$X \rightarrow x$ is derivable in NC_n".

THEOREM 4.1 ([10]). *NL_0 and NC are equivalent.*

PROOF. Analogous to that of Theorem 3.1. In view of (A6) and (A7), every level of the hierarchy (S_n) has its own hierarchical structure. Consequently, in order to prove that NC is closed w. r. t. (R2), one must prove that (A6) and (A7) are admissible rules of NC_n for every n. This requires some auxiliary results and makes the proof technically very complicated when compared to that of Theorem 3.1. ■

We define the *reflection* $\overline{x}$ of the type x as follows: $\overline{s} = s$ for s primitive; $\overline{y/z} = \overline{z}\backslash\overline{y}$; $\overline{z\backslash y} = \overline{y}/\overline{z}$. Clearly $\overline{\overline{x}} = x$.

For a fixed n, we define inductively the sequences (F'_k), (F''_k), (G'_k), (G''_k) of axiom schemata as follows:

F'_0: $y \to (x/y)\backslash x$

G'_0: $y\backslash x \to x$ whenever $\to y$ is an axiom of NC_n

F''_0: $y \to x/(y\backslash x)$

G''_0: $x/y \to x$ whenever $\to y$ is an axiom of NC_{n-1}

F'_{k+1} (resp. G'_{k+1}): $\overline{v}\backslash z_k \to \overline{u}\backslash z_k$ whenever $u \to v$ is F'_k (resp. G'_k)

F''_{k+1} (resp. G''_{k+1}): $z_k/\overline{v} \to z_k/\overline{u}$ whenever $u \to v$ is F''_k (resp. G''_k)

For every n, let NC^R_n be the system whose only rule is (CUT) and whose axioms are all the F_k and G_k as well as

F': $y/z \to ((x/y)\backslash x)/z$

G': $(y\backslash x)/z \to x/z$ whenever $\to y$ is an axiom of NC_n

F'': $z\backslash y \to z\backslash(x/(y\backslash x))$

G'': $z\backslash(x/y) \to z\backslash x$ whenever $\to y$ is an axiom of NC_{n-1}

The axioms F_0 and F are said to be *Montagovian.*

THEOREM 4.2 ([13]). *Let $n > 0$. If $\vdash_n u/v$ and $u \neq v$, then $v \to u$ is derivable in NC^R_n.*

PROOF. Induction on derivations in NC^R_n, NC_n and in the system NC'_n related to NC_n. The details may be found in [13] (Theorems 4 and 5) ■

THEOREM 4.3 ([13]). *If $n > 0$, $\vdash_n s/t$ and s is primitive, then $t \in S_n$.*

PROOF. We define reductions in NC^R_n analogously to those in C^R_n. By Theorem 4.2, there exists a reduction of t to s in NC^R_n. Here again, we distinguish two cases (1) and (2) depending on whether it has a Montagovian stage or not.

(1) We proceed exactly as in the associative case.

(2) Define the complexity of an axiom G_0 like that of an axiom (5A) of C^R_n. An axiom G_{k+1} of the form $\overline{v}\backslash z_k \to \overline{u}\backslash z_k$ or $z_k/\overline{v} \to z_k/\overline{u}$ has the same complexity as $u \to v$. An axiom G of the form $u/z \to v/z$ has the same

complexity as $u \to v$. The complexity of a reduction in NC_n^R is the sum of the complexities of all its stages G and G_k.

The *height* of an axiom F_k or G_k is the number k. Axioms F and G have height 0. The height of a reduction in NC_n^R is the sum of heights of all its stages.

We prove by induction on the height of a reduction that whenever it has a Montagovian stage, either its complexity or the number of its stages F_k may be lowered. Thus, (2) may be reduced to (1). ■

COROLLARY 4.1. *Let s be primitive. If $n > 0$ and $\vdash_n s/(s/x)$, then $\vdash_{n-1} x$.*

We prove now in the same way as in the associative case that NC_{n+1} is a proper extension of NC_n.

Let us emphasize that, in spite of some apparent similarity, there is no far-reaching analogy between the proofs of case (2) in Theorems 3.3 and 4.3. They involve quite different technical means. This lack of a general method of type-raising elimination is peculiar to the nonfinite-axiomatizability proofs for various systems related to the Lambek calculus.

5. Hilbert-style formalism

We adopt here the "logical" interpretation of the Lambek calculus presented in Section 1. Consequently, there are two variants of *modus ponens*:

$$\text{(MP)}\quad \mathbf{A}\ \mathbf{A}{\to}\mathbf{B} \vdash \mathbf{B} \qquad\text{and}\qquad \text{(PM)}\quad \mathbf{B}{\leftarrow}\mathbf{A}\ \mathbf{A} \vdash \mathbf{B}$$

corresponding to the derivable sequents $y\ y\backslash x \vdash x$ and $x/y\ y \vdash x$ of L_0 (and NL_0).

We are concerned with the Hilbert-style axiomatization of implicational logics related to L_0 and NL_0 in the above-mentioned sense. As far as possible, we look for axiomatics which (1) do not employ rules other than *modus ponens* and (2) are finite.

The cut-rule axiomatics for L_0 and NL_0 given respectively in sections 3 and 4 have the property that all their axiom sequents have empty antecedents except for $y\ y\backslash x \vdash x$ and $x/y\ y \vdash x$ which, under our interpretation, are (MP) and (PM). In [9], unidirectional (i.e., \-free or /-free) parts of both systems were axiomatized in a similar way. Since, under the same interpretation, the cut rule is nothing more than the metarule of proof-tree construction, it follows that the corresponding implicational calculi are not finitely detachment-rule axiomatizable.

We define the *reflection* $\overline{\mathbf{A}}$ of an implicational formula $\mathbf{A}$ by induction: $\overline{\mathbf{A}} = \mathbf{A}$ if $\mathbf{A}$ is a propositional variable; $\overline{\mathbf{A} \to \mathbf{B}} = \overline{\mathbf{B}} \leftarrow \overline{\mathbf{A}}$; $\overline{\mathbf{B} \leftarrow \mathbf{A}} = \overline{\mathbf{A}} \to \overline{\mathbf{B}}$.

Now, the implicational logic for L_0 may be axiomatized by means of (MP), (PM), and the following axiom schemata:

(1) $\mathbf{A} \to \mathbf{A}$
(2) $\mathbf{A} \to ((\mathbf{B} \leftarrow \mathbf{A}) \to \mathbf{B})$,
(3) $(\mathbf{A} \to \mathbf{B}) \to \mathbf{B}$ whenever $\mathbf{A}$ is an axiom,
(4) $(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))$.

together with their reflections. To get the $\leftarrow$-free fragment, corresponding to the system L of [9], delete everything except for (MP), (1), (3), and (4).

It was proved in [11] that (3) may be treated as an inference rule rather than an axiom-forming rule. In this case, we clearly get finite axiom systems.

The implicational logic for NL_0 may be axiomatized by means of (MP), (PM), (1), (2), (3) as above, and the following axiom schemata:

(4) $(\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to (\mathbf{B} \leftarrow (\mathbf{A} \to \mathbf{B})))$,
(5) $(\mathbf{A} \to \mathbf{C}) \leftarrow ((\mathbf{B} \leftarrow (\mathbf{A} \to \mathbf{B})) \to \mathbf{C})$,
(6) $(\mathbf{C} \to (\mathbf{B} \leftarrow \mathbf{A})) \to (\mathbf{C} \to \mathbf{B})$ whenever $\mathbf{A}$ is an axiom,
(7) $((\mathbf{B} \leftarrow \mathbf{A}) \to \mathbf{C}) \leftarrow (\mathbf{B} \to \mathbf{C})$ whenever $\mathbf{A}$ is an axiom.

together with their reflections. To get the $\leftarrow$-free fragment, corresponding to the system NL of [9], delete everything except for (MP), (1), and (3) and then add

(8) $(\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B})$ whenever $\mathbf{A} \to \mathbf{B}$ is an axiom,
(9) $(\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C})$ whenever $\mathbf{A} \to \mathbf{B}$ is an axiom.

Here again, (6), (7), (8), (9), and (3) may be treated as inference rules rather than axiom-forming rules, as proved in [10], in order to obtain finite axiom systems.

Both L_0 and NL_0 are *substructural*, i.e., their sequential formulation does not involve Gentzen's "structural" rules which operate on sequent antecedents without introducing any functor. However, variants of L_0 augmented with structural rules like

$$\text{(P)}\ \frac{UxyV \vdash z}{UyxV \vdash z} \qquad \text{and} \qquad \text{(C)}\ \frac{Uxx \vdash y}{Ux \vdash y}$$

have also been considered. We denote by LP_0 (resp. LPC_0) the system L_0 with (P) (resp. (P) and (C)) added.

In LP_0 (and LPC_0), both $y\backslash x \vdash x/y$ and $x/y \vdash y\backslash x$ are derivable. As a consequence, $\to$ and $\leftarrow$ may be reduced to a single implication, to be denoted $\to$. An appropriate axiomatics for LP_0 found by Buszkowski [2] consists of (MP) and the following schemata:

(1′) $\mathbf{A} \to \mathbf{A}$,
(2′) $(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{B} \to (\mathbf{A} \to \mathbf{C}))$,
(3′) $(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))$.

If we add the schema

(4′) $(\mathbf{A} \to (\mathbf{A} \to \mathbf{B})) \to (\mathbf{A} \to \mathbf{B})$,

we get an axiomatics for LPC_0 which, curiously, is equivalent to the well-known "weak theory of implication" W (see, e.g., [7] and references therein). In both cases, (2′) may be replaced by

(2″) $\mathbf{A} \to ((\mathbf{A} \to \mathbf{B})) \to \mathbf{B})$.

which is equivalent to (2′) modulo (1′), (3′), and (MP).

Acknowledgements. I wish to express my gratitude to Professor Wojciech Buszkowski for suggesting me the general idea of the paper an for his further helpful hints, as well as to Professor Jacek Malinowski for his encouragement.

References

[1] Abrusci, V.M., 'A comparison between Lambek syntactic calculus and intuitionistic linear propositional logic', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 36: 11–15, 1990.

[2] Buszkowski, W., 'The logic of types', in Srzednicki, J. (ed.), *Initiatives in Logic*, M. Nijhoff, Amsterdam, 1987, pp. 180–206.

[3] Cohen, J.M., 'The equivalence of two concepts of categorial grammar', *Information and Control*, 10: 475–484, 1967.

[4] Kandulski, M., 'The non-associative Lambek calculus', in Buszkowski, W., Marciszewski, W., van Benthem, J. (eds.), *Categorial Grammar*, J. Benjamins, Amsterdam, 1988, pp. 141–151.

[5] Lambek, J., 'The mathematics of sentence structure', *American Mathematical Monthly*, 5: 154–170, 1958.

[6] Lambek, J., 'On the calculus of syntactic types', in Jakobson, R. (ed.), *Structure of Language and Its Mathematical Aspects*, AMS, Providence, 1961, pp. 166–178.

[7] Urquhart, A., 'Completeness of weak implication', *Theoria*, 3: 274–282, 1971.

[8] Zielonka, W., 'Axiomatizability of Ajdukiewicz-Lambek calculus by means of cancellation schemes', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 27: 215–224, 1981.

[9] Zielonka, W., 'Cut-rule axiomatization of product-free Lambek calculus with the empty string', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 34: 135–142, 1988.

[10] Zielonka, W., 'Cut-rule axiomatization of the syntactic calculus NL_0', *Journal of Logic, Language and Information*, 9: 339–352, 2000.

[11] Zielonka, W., 'Cut-rule axiomatization of the syntactic calculus L_0', *Journal of Logic, Language and Information*, 10: 233–236, 2001.

[12] Zielonka, W., 'On reduction systems equivalent to the Lambek calculus with the empty string', *Studia Logica*, 71: 31–46, 2002.

[13] Zielonka, W., 'On reduction systems equivalent to the non-associative Lambek calculus with the empty string', *Journal of Logic and Computation*, 17: 299–310, 2007.

Wojciech Zielonka
Faculty of Mathematics and Computer Science
University of Warmia and Mazury
Żołnierska 14a
10-561 Olsztyn, Poland
zielonka@uwm.edu.pl

Daniele Mundici

Faithful and Invariant Conditional Probability in Łukasiewicz Logic

in memoriam Sauro Tulipani

Abstract. To every consistent finite set Θ of conditions, expressed by formulas (equivalently, by one formula) in Łukasiewicz infinite-valued propositional logic, we attach a map $\mathcal{P}_\Theta$ assigning to each formula ψ a rational number $\mathcal{P}_\Theta(\psi) \in [0,1]$ that represents "the conditional probability of ψ given Θ". The value $\mathcal{P}_\Theta(\psi)$ is effectively computable from Θ and ψ. The map $\Theta \mapsto \mathcal{P}_\Theta$ has the following properties: (i) (Faithfulness): $\mathcal{P}_\Theta(\psi) = 1$ if and only if $\Theta \vdash \psi$, where $\vdash$ is syntactic consequence in Łukasiewicz logic, coinciding with semantic consequence because Θ is finite. (ii) (Additivity): For any two formulas ϕ and ψ whose conjunction is falsified by Θ, letting χ be their disjunction we have $\mathcal{P}_\Theta(\chi) = \mathcal{P}_\Theta(\phi) + \mathcal{P}_\Theta(\psi)$. (iii) (Invariance): Whenever Θ' is a finitely axiomatizable theory and ι is an isomorphism between the Lindenbaum algebras of Θ and of Θ', then for any two formulas ψ and ψ' that correspond via ι we have $\mathcal{P}_\Theta(\psi) = \mathcal{P}_{\Theta'}(\psi')$. (iv) If $\theta = \theta(x_1, \ldots, x_n)$ is a tautology, then for any formula $\psi = \psi(x_1, \ldots, x_n)$, the (now unconditional) probability $\mathcal{P}_{\{\theta\}}(\psi)$ is the Lebesgue integral over the n-cube of the McNaughton function represented by ψ.

Keywords: Conditional, conditional probability, de Finetti coherence criterion, Dutch Book, many-valued logic, Łukasiewicz logic, infinite-valued logic, MV-algebra, state, finitely additive measure, subjective probability, invariant measure, faithful state.

Introduction: Conditionals and de Finetti coherence criterion

Probability measures over sets of events were characterized by de Finetti as coherent betting systems as follows: Two players, Ada (the palindromic, or reversible bookmaker) and Blaise (the philosopher-mathematician bettor) wager money on the possible occurrence of events $\phi_1, \phi_2, \ldots$, given a set Θ of conditions, all represented by formulas. By $\mathrm{Mod}(\Theta)$ we shall denote the set of all "possible worlds" satisfying Θ. Every possible world $V \in \mathrm{Mod}(\Theta)$ assigns to each event ϕ_i a "truth-value" $V(\phi_i)$, which is always assumed to be a real number in the real interval $[0,1]$. In particular, the truth-value of a boolean (yes-no) event is either 1 or 0. Ada assigns to each event ϕ_i a

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

"betting odd" $\mathcal{B}(\phi_i) \in [0,1]$. Then Blaise chooses a finite subset $\{\psi_1, \ldots, \psi_n\}$ of $\{\phi_1, \phi_2, \ldots\}$, and for each $j = 1, \ldots, n$ he fixes a "stake" $\sigma_j \in \mathbb{R}$ for his bet on event ψ_j. As an effect of his choice, $\sigma_j \mathcal{B}(\psi_j)$ euros are instantly exchanged between Blaise and Ada, with the understanding that $-\sigma_j V(\psi_j)$ euros shall be exchanged in the opposite direction when the truth-value $V(\psi_j)$ is known in the possible world $V \in \mathrm{Mod}(\Theta)$. *Money transfers are so oriented that "positive" means Blaise-to-Ada.* In particular, if Blaise chooses a *negative stake* σ_j then we have a "reverse bet" on event ψ_j, where the roles of Ada and Blaise are interchanged: Ada pays now $|\sigma_j|\, \mathcal{B}(\psi_j)$ euros to Blaise, and Blaise will pay off to Ada $|\sigma_j|\, V(\psi_j)$ euros.[1]

By definition, Ada's "book" $\mathcal{B}$ is *coherent* if it has the following property:

> For any events $\psi_1, \ldots, \psi_n \in \{\phi_1, \phi_2, \ldots\}$ and stakes $\sigma_1, \ldots, \sigma_n \in \mathbb{R}$ there is a possible world $V \in \mathrm{Mod}(\Theta)$ such that
>
> $$\sum_{j=1}^{n} \sigma_j(\mathcal{B}(\psi_j) - V(\psi_j)) \geq 0. \tag{1}$$

In other words, Ada's book is incoherent if Blaise can choose events and stakes for his (possibly reverse) bets, ensuring him a sure profit regardless of the outcome of the gamble.

For the case of boolean events ϕ_i, where $V(\phi_i) \in \{0,1\}$ for any possible world V, de Finetti showed that condition (1) is necessary and sufficient for $\mathcal{B}$ to be a finitely additive probability measure on the (Lindenbaum) boolean algebra of Θ. In this way he derived the main properties of conditional probability from principles that are applicable well beyond boolean semantics.[2] In fact, de Finetti's characterization was extended by Paris [18] to various modal logics, by Gerla [8] to finite-valued Łukasiewicz logics, and by the present author [15] to infinite-valued Łukasiewicz logic Ł_∞.

Specifically, in [15, 5.6, 6.1] it is proved that, given any consistent set of formulas Θ, a [0,1]-valued map $\mathcal{B}$ on formulas satisfies condition (1) if and only if $\mathcal{B}$ is a *state of* Θ, i.e., $\Theta \vdash \psi \Rightarrow \mathcal{B}(\psi) = 1$, and whenever Θ proves the incompatibility of the conjunction $\phi \odot \psi$ of two formulas ϕ, ψ then $\mathcal{B}$ evaluates their disjunction $\phi \oplus \psi$ as $\mathcal{B}(\phi) + \mathcal{B}(\psi)$. This shows that states are a natural generalization of de Finetti's coherent betting systems for continuously-valued events.

---

[1] Needless to say, real bookmakers never accept reverse bets.

[2] See, e.g., [4, pp. 311–312], [5, pp. 34–35], [6, pp. 85–90]. See [11] for further developments of the logical theory of conditionals vs. conditional bets. See [1] for deductive-algorithmic applications.

From a different, but no less important viewpoint, in [17, Proposition 1.1] it is proved that states of (the Lindenbaum algebra of) Θ are in one-one correspondence with Borel probability measures on the set $\mathrm{Mod}(\Theta)$, equipped with the natural (spectral) topology. Thus states also take care of probability distributions over $\mathrm{Mod}(\Theta)$ different from the even distribution corresponding to Lebesgue measure.

Having thus seen that states provide a convincing *philosophical* justification of probability in systems of $[0,1]$-valued events obeying Łukasiewicz semantics, we define a *conditional* of Ł_∞ to be a map $\mathcal{P}\colon \Theta \mapsto \mathcal{P}_\Theta$ assigning to every consistent finite set Θ of formulas a state $\mathcal{P}_\Theta$ of Θ.[3] For every formula ψ, the number $\mathcal{P}_\Theta(\psi)$ is meant to represent the probability of ψ given Θ.[4] A conditional $\mathcal{P}$ is *faithful* if the only $\mathcal{P}$-negligible events are those which Θ proves to be impossible. $\mathcal{P}$ is *invariant* if $\mathcal{P}_\Theta(\psi)$ only depends on the logical-semantical relations between the events described by Θ and ψ. See 3.2 for a precise definition.

The invariance property of a conditional $\mathcal{P}$ captures the natural desideratum that the probability of an event ψ given a set of Θ conditions, should not depend on our choice of which events are to be thought of as "primitive" and as such, will be coded by propositional variables. Invariance is a very strong property: Panti [17, Theorem 2.3] characterizes the Lebesgue integral as the only invariant state s on the Lindenbaum algebra of m-variable formulas in Łukasiewicz logic, such that, letting $\phi^n = \phi \odot \cdots \odot \phi$ (n times), $\lim_{n\to\infty} s(\phi^n) = 0$ for all formulas with $\dim(\mathrm{Mod}(\phi)) < m$.

In Theorem 3.3 we show that Ł_∞ has a faithful invariant conditional $\mathcal{P}$. In particular, if the conditions Θ are tautological, the (now "unconditional") probability $\mathcal{P}_\Theta(\phi)$ of any event ϕ coincides with the Lebesgue integral of the McNaughton function f_ϕ represented by ϕ (see Corollary 5.2).

1. The i-dimensional volume of a formula

Preliminaries on MV-algebras and Łukasiewicz logic [2]. Throughout, $[0,1]^n$ denotes the n-cube, equipped with the product topology of $\mathbb{R}^n$. Further, $\mathcal{M}([0,1]^n)$ shall denote the MV-algebra of all continuous piecewise (affine) linear functions $f\colon [0,1]^n \to [0,1]$ such that each piece of f has integer coefficients. Any such f is called a ($[0,1]$-valued) *McNaughton function.*

---

[3] In MV-algebraic probability theory, ([19, 3.2] and references therein) conditional expectations are defined for σ-complete MV-algebras with a product operation. No such assumption is made here.

[4] Following Makinson's remarks on conditionalization in [10], we do not assume that "ψ given Θ" is a formula in any logic.

We say that $P \subseteq [0,1]^n$ is a *McNaughton set* if $P = g^{-1}(1)$ for some McNaughton function $g \in \mathcal{M}([0,1]^n)$. McNaughton's theorem [2, 9.1.5], states that $\mathcal{M}([0,1]^n)$ is the free MV-algebra over the free generating set whose members are the coordinate functions $\pi_i \colon [0,1]^n \to [0,1]$.

More generally, for any closed set $\emptyset \neq X \subseteq [0,1]^n$ we shall denote by $\mathcal{M}(X)$ the MV-algebra of restrictions to X of the functions in $\mathcal{M}([0,1]^n)$. Further, $\mu(\mathcal{M}(X))$ shall denote the space of maximal ideals of $\mathcal{M}(X)$, i.e., kernels of homomorphisms of $\mathcal{M}(X)$ into the standard MV-algebra $[0,1]$. The space $\mu(\mathcal{M}(X))$ comes equipped with the *spectral* topology: a basis of closed sets for $\mu(\mathcal{M}(X))$ is given by the *zerosets* Z_f of all elements $f \in \mathcal{M}(X)$, i.e., by the sets $Z_f = \{\mathfrak{m} \in \mu(\mathcal{M}(X)) \mid f \in \mathfrak{m}\}$. As is well known, $\mu(\mathcal{M}(X))$ is a nonempty compact Hausdorff space. In fact, much more is true:

PROPOSITION 1.1. *For $n = 1, 2, \ldots$ let X be a nonempty closed subset of $[0,1]^n$. We then have:*

(i) *The map $x \in X \mapsto \mathfrak{m}_x = \{f \in \mathcal{M}(X) \mid f(x) = 0\}$ is a homeomorphism of X onto the maximal ideal space $\mu(\mathcal{M}(X))$. The inverse map $\mathfrak{m} \mapsto x_{\mathfrak{m}}$ sends every $\mathfrak{m} \in \mu(\mathcal{M}(X))$ to the only element $x_{\mathfrak{m}}$ of the set $\bigcap\{g^{-1}(0) \mid g \in \mathfrak{m}\}$.*

(ii) *For every $\mathfrak{m} \in \mu(\mathcal{M}(X))$ there is a unique pair $(\iota_{\mathfrak{m}}, I_{\mathfrak{m}})$ where $I_{\mathfrak{m}}$ is a subalgebra of the standard MV-algebra $[0,1]$, and $\iota_{\mathfrak{m}}$ is an isomorphism of the quotient $\mathcal{M}(X)/\mathfrak{m}$ onto $I_{\mathfrak{m}}$.*

(iii) *For every $x \in X$ and $f \in \mathcal{M}(X)$, $f(x) = \iota_{\mathfrak{m}_x}(f/\mathfrak{m}_x)$.*

PROOF. (i) The proof of [13, 4.17] shows that for any two distinct points $x, y \in X$ there is $f \in \mathcal{M}(X)$ such that $f(x) \neq f(y)$. By [2, 3.4.3] the map $x \mapsto \mathfrak{m}_x$ is a one-one correspondence between X and $\mu(\mathcal{M}(X))$. The definition of spectral topology ensures that this map is a homeomorphism.

(ii) By [2, 1.2.10, 3.5.1], $\mathcal{M}(X)/\mathfrak{m}$ is isomorphic to a subalgebra A of $[0,1]$. By [2, 7.2.6] A is uniquely determined, and so is the isomorphism of $\mathcal{M}(X)/\mathfrak{m}$ onto A.

(iii) Two functions $f, g \in \mathcal{M}(X)$ are mapped to the same element by the quotient map $f \mapsto f/\mathfrak{m}_x$ iff $|f - g| \in \mathfrak{m}_x$ iff $f(x) = g(x)$. The map $\beta \colon f/\mathfrak{m}_x \mapsto f(x)$ is an isomorphism of $\mathcal{M}(X)/\mathfrak{m}_x$ onto the MV-algebra $V = \{v \in [0,1] \mid v = f(x)$ for some $f \in \mathcal{M}(X)\}$. By (ii), $\beta = \iota_{\mathfrak{m}_x}$. ■

Preliminaries on polyhedral topology [7, 22]. For $0 \leq m \leq n$, an *m-simplex* in $\mathbb{R}^n$ is the convex hull $S = \mathrm{conv}(v_0, \ldots, v_m)$ of $m+1$ affinely independent points in $\mathbb{R}^n$. The *vertices* $v_0, \ldots, v_m$ are uniquely determined by S. We say that an m-simplex $S = \mathrm{conv}(v_0, \ldots, v_m) \subseteq [0,1]^n$ is *rational* if the coordinates of each vertex v_j of S are rational numbers. Let Σ be a simplicial complex. A simplex $S \in \Sigma$ is *maximal* if it is not properly contained in any simplex of Σ. We denote by $\Sigma^{\max}(d)$ the set of maximal d-simplexes of Σ. The point-set union of the simplexes in Σ is called the *support* of Σ, and is denoted $|\Sigma|$. Σ is said to be a *triangulation* of $|\Sigma|$.

Let $S \subseteq [0,1]^n$ be a rational m-simplex with vertices $v_0, \ldots, v_m$. For each $j = 0, \ldots, m,$ the vertex v_j has the form $(r_{j1}/s_{j1}, \ldots, r_{jn}/s_{jn})$, for uniquely determined integers $0 \leq r_{jt} \leq s_{jt}$ $(t = 1, \ldots, n)$ such that $s_{jt} > 0$, and $\gcd(r_{jt}, s_{jt}) = 1$. The least common multiple of $\{s_{j1}, \ldots, s_{jt}\}$ is called the *denominator* of v_j, and is denoted $\operatorname{den}(v_j)$. The *denominator* $\operatorname{den}(S)$ *of* S is defined by $\operatorname{den}(S) = \operatorname{den}(v_0) \cdots \operatorname{den}(v_m)$. The *homogeneous correspondent* of v_j is the integer vector $\tilde{v}_j = \operatorname{den}(v_j)\,(r_{j1}/s_{j1}, \ldots, r_{jn}/s_{jn},\, 1) \in \mathbb{Z}^{n+1}$. This vector is *primitive*, i.e., minimal (as an integer nonzero vector) along its *ray* $\{\mu\tilde{v}_j \in \mathbb{R}^{n+1} \mid \mu \geq 0\}$. Conversely, v_j is said to be the *affine correspondent* of $\tilde{v}_j$.

An m-simplex $S = \operatorname{conv}(v_0, \ldots, v_m) \subseteq [0,1]^n$ is said to be *unimodular* if it is rational and the set of homogeneous correspondents $\{\tilde{v}_0, \ldots, \tilde{v}_m\}$ of its vertices can be extended to a basis of the free abelian group $\mathbb{Z}^{n+1}$. A simplicial complex is said to be a *unimodular triangulation* (of its support) if all its simplexes are unimodular.

As usual [7, 22], by a *subdivision* of a simplicial complex Σ we mean a simplicial complex Σ' with the same support of Σ, such that every simplex of Σ' is contained in some simplex of Σ. Let Σ be a simplicial complex and $a \in |\Sigma| \subseteq [0,1]^n$. Following [22, p. 376], or [7, III, 2.1], by the *blow up* $\Sigma_{(a)}$ *of* Σ *at* a we mean a subdivision which is obtained from Σ by the following procedure: replace every simplex $S \in \Sigma$ containing a by the set of all simplexes of the form $\operatorname{conv}(a, F)$, where F is any face of S that does not contain a. Note that $\Sigma_{(a)}$ is a simplicial complex. The inverse of a blow-up is called a *blow down*.

For any $m \geq 1$ and unimodular m-simplex $T = \operatorname{conv}(w_0, \ldots, w_m) \subseteq [0,1]^n$ the *Farey mediant of* T is the affine correspondent of the vector $\tilde{w}_0 + \cdots + \tilde{w}_m \in \mathbb{Z}^{n+1}$, where each $\tilde{w}_i$ is the homogeneous correspondent of w_i. In the particular case when Σ is a unimodular triangulation and a is the *Farey mediant* of a simplex S of Σ, the blow-up $\Sigma_{(a)}$ is still unimodular.

In Proposition 1.2 below we shall describe the basic relations between McNaughton sets, unimodular triangulations, and formulas in the infinite-valued propositional calculus $Ł_\infty$ of Łukasiewicz, [20]. For notation and terminology we shall follow [2, Chapter 4]. Accordingly, negation, disjunction and conjunction are denoted by $\neg, \oplus$ and $\odot$. The implication $\phi \to \psi$ is thought of as an abbreviation of $\neg\phi \oplus \psi$. For each $n = 1, 2, \ldots$ we let Form_n denote the set of formulas $\psi = \psi(x_1, \ldots, x_n)$ whose variables are contained in the set $\{x_1, \ldots, x_n\}$. The map $x_i \mapsto \pi_i$ uniquely extends to an interpretation $\psi \mapsto f_\psi$ of every formula $\psi(x_1, \ldots, x_n)$ as a McNaughton function $f_\psi \colon [0,1]^n \to [0,1]$. The McNaughton set $f_\psi^{-1}(1)$ is a closed subset of the n-cube, denoted $\mathrm{Oneset}(\psi)$. By McNaughton theorem [2, 9.1], every $f \in \mathcal{M}([0,1]^n)$ has the form f_ϕ for some $\phi(x_1, \ldots, x_n)$.

PROPOSITION 1.2. *Let X be a nonempty subset of $[0,1]^n$. Then the following conditions are equivalent:*

(i) *X coincides with the support of some unimodular triangulation Δ.*

(ii) *X is a McNaughton set.*

(iii) *$X = \mathrm{Oneset}(\psi)$ for some formula $\psi(x_1, \ldots, x_n)$.*

PROOF. (iii)→(ii) is trivial. (ii)→ (iii) follows from McNaughton theorem [2, 3.1, 9.1.5]. (ii)→ (i) Suppose $X = f^{-1}(1)$ for some McNaughton function f. By [2, 9.1.2] there is a unimodular triangulation Υ of the n-cube such that f is linear over each simplex of Υ. Let $\Delta \subseteq \Upsilon$ be the sub-complex of Υ given by those simplexes which are contained in X. Then Δ is the required unimodular triangulation of X.

Finally, to prove (i)→(ii), let $S_1, \ldots, S_u$ be the maximal simplexes of Δ. Let $H_1, \ldots, H_k$ be a list of closed half-spaces in $\mathbb{R}^n$, each H_i of the form

$$H_i = \{(x_1, \ldots, x_n) \in [0,1]^n \mid p_1 x_1 + \cdots + p_n x_n \geq q\}$$

for some $p_1, \ldots, p_n, q \in \mathbb{Q}$, such that for each $j = 1, \ldots, u$ the simplex S_j is the intersection of halfspaces taken from the set $\{H_1, \ldots, H_k\}$. Let $e_1, \ldots, e_n$ be the standard basis vectors in $\mathbb{R}^n$. For each permutation σ of the set $\{1, \ldots, n\}$ let S_σ be the n-simplex in $\mathbb{R}^n$ whose vertices are $0, e_{\sigma(1)}, e_{\sigma(1)} + e_{\sigma(2)}, \ldots, e_{\sigma(1)} + \cdots + e_{\sigma(n)}$. Let K^* be the simplicial complex consisting of the n-simplexes S_σ's together with their faces. (In [21, pp. 60–61], the set of S_σ's is denoted $T(Q)$ and is called the standard triangulation of the n-cube.) Direct inspection shows that K^* is a unimodular triangulation of the n-cube. Using the affine version of De Concini-Procesi theorem [3], [7, p. 252] as in [16, 2.2], we construct a sequence of unimodular triangulations $K_0 = K^*, K_1, \ldots, K_r$ such that (i) each K_{t+1} is obtained by blowing-up K_t at the Farey mediant of some 1-simplex of K_t, (ii) for each $i = 1, \ldots, k$, the convex polyhedron $H_i \cap [0,1]^n$ is a union of simplexes of K_r, whence (iii) for each $j = 1, \ldots, u$, S_j is a union of simplexes of K_r. Let the McNaughton function $g \in \mathcal{M}([0,1]^n)$ be uniquely determined by the following properties: g is linear over every simplex of K_r, $g(v) = 1$ for each vertex $v \in X$ of a simplex of K_r, and $g(w) = 0$ for any other vertex w of K_r; such g is obtainable as a suitable sum of the "Schauder hat" functions h_v corresponding to Δ as in [2, 9.1.4]. The coefficients of each linear piece of g are integers as a consequence of the unimodularity of K_r. Direct inspection shows that $X = g^{-1}(1)$. ■

Theorem 1.3. *Let $P \subseteq [0,1]^n$ be a McNaughton set and Δ a unimodular triangulation of P. For each $i = 0, 1, \ldots$ let $\lambda_i(P, \Delta)$ be defined by*

$$\lambda_i(P, \Delta) = \sum_{S \in \Delta^{\max}(i)} \frac{1}{i!\ \mathrm{den}(S)}\ , \tag{2}$$

*where the sum equals zero if $\Delta^{\max}(i) = \emptyset$. Then for any unimodular triangulation Δ' of P we have $\lambda_i(P, \Delta) = \lambda_i(P, \Delta')$. The rational number $\lambda_i(P) = \lambda_i(P, \Delta) = \lambda_i(P, \Delta')$ is called the i-*dimensional volume of P.

Proof. By Włodarczyk's solution of the weak Oda conjecture [22, 13.3] (also see [12]), there is a sequence of unimodular triangulations

$$\Delta_0 = \Delta, \Delta_1, \ldots, \Delta_{n-1}, \Delta_n = \Delta'$$

such that Δ_{t+1} is obtained from Δ_t by a blow-up at the Farey mediant of some simplex $S \in \Delta_t$, or vice versa, Δ_t is obtained from Δ_{t+1} by a similar blow-up. Thus we have only to prove $\lambda_i(P, \Delta_t) = \lambda_i(P, \Delta_{t+1}), \quad \forall t = 0, \ldots, n-1$ and $i = 0, 1, \ldots$. Without loss of generality we may suppose that Δ_{t+1} is obtained from Δ_t by a blow-up at the Farey mediant a of some j-simplex $S = \mathrm{conv}(v_0, \ldots, v_j) \in \Delta_t$ (otherwise we interchange the roles of Δ_t and Δ_{t+1}). From the unimodularity of S we get $\mathrm{den}(a) = \mathrm{den}(v_0) + \cdots + \mathrm{den}(v_j)$. Let M be the set of all m such that a is contained in some maximal m-simplex of Δ_t. Let $T = \mathrm{conv}(v_0, \ldots, v_j, \ldots, v_m)$ be a maximal simplex of Δ_t containing a. For each $u = 0, \ldots, j$ let

$$F_u = \mathrm{conv}(v_0, \ldots, v_{u-1}, a, v_{u+1}, \ldots, v_j, v_{j+1}, \ldots, v_m).$$

The unimodularity of T ensures that each F_u is unimodular, and $\mathrm{den}(F_u) = \mathrm{den}(T) \cdot \mathrm{den}(a)/\mathrm{den}(v_u)$. By definition of blow-up, T is replaced in Δ_{t+1} by the simplicial complex whose maximal simplexes are the m-simplexes $F_0, \ldots, F_j$. From the identity $1/\mathrm{den}(T) = \sum_{u=0}^{j} 1/\mathrm{den}(F_u)$ we obtain

$$\sum\{(m!\,\mathrm{den}(T))^{-1} \mid T \in \Delta_t^{\max}(m)\} = \sum\{(m!\,\mathrm{den}(F))^{-1} \mid F \in \Delta_{t+1}^{\max}(m)\}.$$

We have just shown that $\lambda_m(P, \Delta_t) = \lambda_m(P, \Delta_{t+1})$ for each $m \in M$. For $i \notin M$ the maximal i-simplexes of Δ_t are exactly the same as those of Δ_{t+1}. Thus $\lambda_i(P, \Delta_t) = \lambda_i(P, \Delta_{t+1})$ for all $i = 0, 1, \ldots$. ■

Remark. In particular, when P is a unimodular d-simplex,

$$\lambda_d(P) = \frac{1}{d!\ \mathrm{den}(P)}, \text{ and } \lambda_i(P) = 0 \text{ for } i \neq d. \tag{3}$$

By abuse of terminology, for any formula $\psi(x_1,\dots,x_n)$ and $i = 0,1,\dots$ we define the *i-dimensional volume* $\lambda_i(\psi)$ by $\lambda_i(\psi) = \lambda_i(\mathrm{Oneset}(\psi),\Delta)$, where Δ is an arbitrary unimodular triangulation of Oneset(ψ). Direct inspection shows that the map $(\psi,i) \mapsto \lambda_i(\psi)$ is effectively computable. Trivially, $\lambda_i(\psi) = 0$ for all integers $i > n$. On the other hand, for $i < n$, $\lambda_i(\psi)$ can be arbitrarily high. Finally, for $i = n$, the following result can be taken as a justification of our terminology:

PROPOSITION 1.4. *For every formula $\psi(x_1,\dots,x_n)$, $\lambda_n(\psi)$ is equal to the n-dimensional Lebesgue measure $\lambda(\mathrm{Oneset}(\psi))$.*

PROOF. Without loss of generality we can assume Oneset(ψ) to be n-dimensional, and $\lambda_m(\psi) = 0$ for all $m \neq n$. Let Δ be a unimodular triangulation of Oneset(ψ) as given by Proposition 1.2. Let $S_1,\dots,S_r$ be the n-simplexes of Δ. For each $k = 1,\dots,r$, let $v_{k0},\dots,v_{kn}$ be the vertices of S_k, with their respective denominators $m_{k0},\dots,m_{kn}$. It suffices to show that the n-dimensional Lebesgue volume $\lambda(S_k)$ coincides with $\lambda_n(S_k)$. To this purpose, let $S^*_k \subseteq \mathbb{R}^{n+1}$ be the $(n+1)$-simplex whose vertices are 0, $(v_{k0},1)$, $\dots$, $(v_{kn},1)$. S^*_k is contained in the $(n+1)$-dimensional parallelepiped $P_k = \{\mu_0(v_{k0},1) + \cdots + \mu_n(v_{kn},1) \mid \mu_0,\dots,\mu_n \in [0,1]\}$. By definition of homogeneous correspondent, P_k is included in the parallelepiped $U_k = \{\mu_0\tilde{v}_{k0} + \cdots + \mu_n\tilde{v}_{kn} \mid \mu_0,\dots,\mu_n \in [0,1]\} \subseteq \mathbb{R}^{n+1}$. The unimodularity of S_k means that U_k has unit Lebesgue volume. From $\tilde{v}_{k0} = m_{k0}(v_{k0},1),\dots,\tilde{v}_{kn} = m_{kn}(v_{kn},1)$ it follows that the ($(n+1)$-dimensional) Lebesgue volume $\lambda(P_k)$ equals $(m_{k0}\cdots m_{kn})^{-1}$. It is easy to see that the Lebesgue volume $\lambda(S^*_k)$ satisfies the identities

$$\frac{\lambda(S_k)\times 1}{n+1} = \lambda(S^*_k) = \frac{\lambda(P_k)}{(n+1)!} \ .$$

The first identity is the classical formula for the volume of the $(n+1)$-dimensional pyramid; the second follows from the observation that P_k can be triangulated with $(n+1)!$ simplexes, all having the same Lebesgue volume as S^*_k. By (3) we conclude that $\lambda(S_k) = \lambda(P_k)/n! = (n!\,m_{k0}\cdots m_{kn})^{-1} = \lambda_n(S_k)$. ■

2. Conditionals in Łukasiewicz propositional logic $Ł_\infty$

Intuitively, a “possible world” assigns a “truth-value” to certain “events”. When events are taken care of by infinite-valued Łukasiewicz propositional logic $Ł_\infty$, ([2, Chapter 4]), the imprecise notion of possible world is captured by the following

Definition 2.1. Fix $n = 1, 2, \ldots$. Then a *valuation* (over Form_n) is a function $V \colon \mathrm{Form}_n \to [0,1]$ such that $V(\neg\phi) = 1 - V(\phi)$, $V(\phi \oplus \psi) = \min(1, V(\phi) + V(\psi))$, and $V(\phi \odot \psi) = \max(0, V(\phi) + V(\psi) - 1)$.

Because every valuation is uniquely determined by its restriction to the propositional variables, the map $V \mapsto w_V = (V(x_1), \ldots, V(x_n))$ is a one-one correspondence between valuations over Form_n and points in the n-cube $[0,1]^n$. For arbitrary $w \in [0,1]^n$ we denote by V_w the valuation corresponding to w. V_w is the only valuation such that $w = (V_w(x_1), \ldots, V_w(x_n))$.

We say that formulas $\phi, \psi \in \mathrm{Form}_n$ are (logically) *equivalent* if $V(\phi) = V(\psi)$ for all valuations V. By [2, 4.4.1,4.5.2], this is the same as saying that $\phi \leftrightarrow \psi$ is a tautology. We denote by $|\psi|$ the equivalence class of ψ. Upon equipping the set of equivalence classes of formulas $\psi(x_1, \ldots, x_n)$ with the MV-algebraic operations inherited from the connectives $\neg, \oplus, \odot$, we get the free n-generated MV-algebra $\mathcal{L}_n$, [2, 4.5.5]. Let the map β send the equivalence class $|x_i|$ of each propositional variable x_i to the ith coordinate function $\pi_i \colon [0,1]^n \to [0,1]$. Then by McNaughton theorem [2, 9.1], β uniquely extends to an isomorphism $\eta \colon |\psi| \mapsto f_\psi$ of $\mathcal{L}_n$ onto the MV-algebra $\mathcal{M}([0,1]^n)$. Therefore, for every $\psi \in \mathrm{Form}_n$, η yields a concrete representation of $|\psi| \in \mathcal{L}_n$ as the McNaughton function $f_\psi \in \mathcal{M}([0,1]^n)$. By induction on the number of occurrences of connectives in ψ we get

$$f_\psi(w) = V_w(\psi) \quad \text{for all } w \in [0,1]^n. \tag{4}$$

For every formula $\theta \in \mathrm{Form}_n$ let $\mathrm{Mod}(\theta)$ denote the set of valuations $V \in [0,1]^{\mathrm{Form}_n}$ such that $V(\theta) = 1$. Any such V is said to *satisfy* θ.

The following proposition describes the tight relation between $\mathrm{Oneset}(\theta)$ and $\mathrm{Mod}(\theta)$:

Proposition 2.2. *The range of the one-one map* $V \in \mathrm{Mod}(\theta) \mapsto w_V = (V(x_1), \ldots, V(x_n)) \in [0,1]^n$ *coincides with* $\mathrm{Oneset}(\theta)$.

Proof. From (4) we have: $w = w_V$ for some $V \in \mathrm{Mod}(\theta)$ iff $V_w \in \mathrm{Mod}(\theta)$ iff $V_w(\theta) = 1$ iff $f_\theta(w) = 1$ iff $w \in \mathrm{Oneset}(\theta)$. ■

Logical equivalence of two formulas $\psi, \phi \in \mathrm{Form}_n$ is generalized to logical equivalence $\equiv_\theta$ as follows: $\psi \equiv_\theta \phi$ iff $\theta \vdash \psi \leftrightarrow \phi$, where $\vdash$ is *syntactic consequence* in Łukasiewicz logic Ł_∞, [2, 4.3.2]. For each formula $\phi(x_1, \ldots, x_n)$, the $\equiv_\theta$-equivalence class of ϕ shall be denoted $|\phi|_\theta$. The set of $\equiv_\theta$-equivalence classes of Form_n forms an MV-algebra, called the *Lindenbaum algebra* of θ and denoted $\mathcal{L}_\theta$ (see [2, 4.6.8]).

We write $\theta \models \psi$, and we say that ψ is a *semantic consequence* of θ, iff $\mathrm{Mod}(\theta) \subseteq \mathrm{Mod}(\psi)$ (iff $\mathrm{Oneset}(\theta) \subseteq \mathrm{Oneset}(\psi)$). Wójcicki's theorem [23], [2, 4.6.7] states that $\theta \vdash \psi$ iff $\theta \models \psi$.[5]

The following generalization of McNaughton theorem allows us to identity $\mathcal{L}_\theta$ with $\mathcal{M}(\mathrm{Oneset}(\theta))$:

PROPOSITION 2.3. *For* $\theta \in \mathrm{Form}_n$ *with* $\mathrm{Oneset}(\theta) \neq 0$, *the map* $\rho_\theta \colon |\phi|_\theta \mapsto f_\phi \restriction \mathrm{Oneset}(\theta)$ *is an isomorphism of* $\mathcal{L}_\theta$ *onto* $\mathcal{M}(\mathrm{Oneset}(\theta))$.

PROOF. To prove that ρ_θ is a well defined homomorphism, let O be short for $\mathrm{Oneset}(\theta)$. Then by (4) and [2, 4.5.1] we have:

$$\begin{aligned} f_\psi \restriction O \neq f_\phi \restriction O \quad &\Rightarrow \quad \exists w \in [0,1]^n \text{ with } f_\theta(w) = 1 \text{ and } f_\psi(w) \neq f_\phi(w) \\ &\Rightarrow \quad V_w(\theta) = 1, V_w(\psi) \neq V_w(\phi) \text{ for some } w \in [0,1]^n \\ &\Rightarrow \quad \theta \not\models \psi \leftrightarrow \phi \\ &\Rightarrow \quad \theta \nvdash \psi \leftrightarrow \phi \\ &\Rightarrow \quad |\psi|_\theta \neq |\phi|_\theta. \end{aligned}$$

To prove that ρ_θ is onto $\mathcal{M}(O)$ we will use McNaughton theorem: for every $g \in \mathcal{M}(O)$, letting $h = h_\chi \in \mathcal{M}([0,1]^n)$ be such that $g = h \restriction O$, it follows that $g = \rho_\theta(|\chi|_\theta)$. Finally, in order to prove that ρ_θ is one-one, suppose $|\psi|_\theta$ is not the zero element of $\mathcal{L}_\theta$, i.e., $\theta \nvdash \neg\psi$. By Wójcicki's theorem, $\neg\psi$ is not a semantic consequence of θ, $\theta \not\models \neg\psi$. In other words, for some valuation V we have $V(\theta) = 1$ and $V(\neg\psi) \neq 1$. Thus $V(\psi) > 0$. The point $w = w_V$ satisfies $f_\theta(w) = 1$ and $f_\psi(w) > 0$. Thus, $f_\psi \restriction f_\theta^{-1}(1) \neq 0$, i.e., $f_\psi \restriction O \neq 0$. ■

3. A faithful invariant conditional for Ł$_\infty$

Let $\Theta = \{\theta_1, \ldots, \theta_u\}$ be a nonempty finite subset of Form_n, for some $n = 1, 2, \ldots$. Again by Wójcicki's theorem, the set

$$\Theta^\vdash = \{\psi \in \mathrm{Form}_n \mid \Theta \vdash \psi\}$$

of syntactic consequences of Θ coincides with the set

$$\Theta^\models = \{\psi \in \mathrm{Form}_n \mid V(\psi) = 1 \text{ for all } V \text{ with } 1 = V(\theta_1) = \cdots = V(\theta_u)\}$$

of semantic consequences. Further, letting $\theta = \theta_1 \odot \ldots \odot \theta_u$ be the conjunction of the formulas in Θ, we have $\Theta^\vdash = \{\theta\}^\vdash$. We shall always assume that Θ is *consistent*, i.e., $\Theta^\vdash$ does not coincide with Form_n. Throughout this paper we

---

[5] As is well known, the identity between syntactic and semantic consequence ceases to hold in general if θ is replaced by an infinite set of formulas, [2, p. 99].

shall identify Θ with θ without any danger of confusion. Wójcicki's theorem then becomes the identity

$$\theta^{\models} = \theta^{\vdash}. \tag{5}$$

By Proposition 2.3, the consistency of θ is equivalent to $\mathrm{Oneset}(\theta)$ being nonempty.

Recall [14, 15] that a *faithful state* of an MV-algebra A is a $[0,1]$-valued map s on A such that, for all $x, y \in A$, (i) $s(x) = 1 \Leftrightarrow x = 1$, and (ii) $s(x \oplus y) = s(x) + s(y)$ whenever $x \odot y = 0$. Similarly, by a *faithful state* of a consistent formula $\theta \in \mathrm{Form}_n$ we mean a map $\mathcal{S}\colon \mathrm{Form}_n \to [0,1]$ such that (i) $\mathcal{S}(\psi) = 1$ if and only if $\theta \vdash \psi$; and (ii) $\mathcal{S}(\psi \oplus \phi) = \mathcal{S}(\psi) + \mathcal{S}(\phi)$ whenever $\theta \vdash \neg(\psi \odot \phi)$; in other words, $\mathcal{S}$ is *additive* on pairs of formulas whose conjunction is falsified by θ. It follows that $\mathcal{S}(\neg\psi) = 1 - \mathcal{S}(\psi)$, and $\psi \equiv_\theta \psi' \Rightarrow \mathcal{S}(\psi) = \mathcal{S}(\psi')$, whence the proof of the following proposition is a routine verification:

Proposition 3.1. *Let θ be a fixed, but otherwise arbitrary formula in* Form_n. *Let us define the map*

$$\epsilon\colon \textit{ faithful states of } \mathcal{M}(\mathrm{Oneset}(\theta)) \to \textit{ faithful states of } \theta$$

by the following stipulation: for any such state s, letting $\mathcal{S} = \epsilon(s)$,

$$\mathcal{S}(\psi) = s(f_\psi \restriction \mathrm{Oneset}(\theta)) \quad \forall \psi \in \mathrm{Form}_n.$$

Then $\epsilon\colon s \mapsto \mathcal{S}$ maps faithful states of $\mathcal{M}(\mathrm{Oneset}(\theta))$ *one-one onto faithful states of θ.*

Definition 3.2. A *faithful conditional* is a map $\mathcal{P}\colon \theta \mapsto \mathcal{P}_\theta$ such that, for every $n = 1, 2, \ldots$ and every consistent formula $\theta \in \mathrm{Form}_n$, $\mathcal{P}_\theta$ is a faithful state of θ. We say that $\mathcal{P}$ is *invariant* if it has the following property: whenever $\theta \in \mathrm{Form}_n$, $\theta' \in \mathrm{Form}_{n'}$, and η is an isomorphism of the Lindenbaum algebras $\mathcal{L}_{\theta'}$ and $\mathcal{L}_\theta$, then $\mathcal{P}_\theta(\psi) = \mathcal{P}_{\theta'}(\psi')$ for any two formulas ψ and ψ' such that $|\psi|_\theta = \eta(|\psi'|_{\theta'})$.

When $\mathcal{P}$ is clear from the context, for every $\psi \in \mathrm{Form}_n$ the quantity $\mathcal{P}_\theta(\psi)$ is said to be the *conditional probability of ψ given θ.* As noted in the introduction, invariance ensures that this quantity does not depend on the syntactical details of our transcription of event ψ and condition θ, but only on their mutual logical relations.

The main result of this paper is the following

Theorem 3.3. *Łukasiewicz propositional logic has a faithful invariant conditional $\mathcal{P}$.*

4. Proof: construction of a faithful conditional $\mathcal{P}$

Given a consistent formula $\theta = \theta(x_1, \dots, x_n)$ let the McNaughton set $P \subseteq [0,1]^n$ be defined by $P = \mathrm{Oneset}(\theta)$. By hypothesis, P is nonempty. Let us identify $\mathcal{L}_\theta$ and $\mathcal{M}(P)$ via the isomorphism ρ_θ of Proposition 2.3. Let $\mathcal{D}$ be the set of those indexes j such that $\lambda_j(P) > 0$. Then $\mathcal{D} \subseteq \{0, \dots, n\}$ and its number $|\mathcal{D}|$ of elements is ≥ 1. Let us say that $\mathcal{D}$ is the *dimensional spectrum of* P. With reference to Theorem 1.3, for each $j \in \mathcal{D}$ let the rational coefficient c_j be defined by

$$c_j = c_{j,\theta} = \frac{1}{|\mathcal{D}| \cdot \lambda_j(P)} \, . \tag{6}$$

For any $\psi \in \mathrm{Form}_n$ let $f = f_\psi \restriction P = |\psi|_\theta$. Let Δ be a unimodular triangulation of P. The proof of Proposition 1.2 shows that it is no loss of generality to assume that f is linear over every simplex of Δ, for short Δ is an (always unimodular) f*-triangulation:* As a matter of fact, if Δ is not an f-triangulation, a suitable subdivision $\Delta^\natural$ of Δ via blow-ups will be an f-triangulation, by the affine version of De Concini-Procesi theorem.

For each $i \in \mathcal{D}$ and i-simplex $T = \mathrm{conv}(v_0, \dots, v_i) \in \Delta^{\max}(i)$, the average value $\overline{f \restriction T}$ of f over T is given by

$$\overline{f \restriction T} = \frac{f(v_0) + \cdots + f(v_i)}{i+1} \, . \tag{7}$$

Letting 1_P denote the constant function 1 over P, by Theorem 1.3 we have

$$\sum_{T \in \Delta^{\max}(i)} c_i \cdot \lambda_i(T) \cdot \overline{1_P \restriction T} = \sum_{T \in \Delta^{\max}(i)} \frac{\lambda_i(T)}{|\mathcal{D}| \cdot \lambda_i(P)} = \frac{1}{|\mathcal{D}|}. \tag{8}$$

Let the rational number $\varsigma_T(f)$ be defined by

$$\varsigma_T(f) = c_i \cdot \lambda_i(T) \cdot \overline{f \restriction T}. \tag{9}$$

Intuitively, $\varsigma_T(f)$ is the normalized $(i+1)$-dimensional volume of the portion of space below the graph of $f \restriction T$. We now define

$$\varsigma_{\theta,\Delta}(f) = \sum_{i \in \mathcal{D}} \sum_{T \in \Delta^{\max}(i)} \varsigma_T(f). \tag{10}$$

This is the total normalized sum of the volumes below the graph of $f \restriction P$ in all dimensions where P has maximal simplexes. (In Claim 1 below we shall see that this volume only depends on θ.) From (8) we obtain

$$\varsigma_{\theta,\Delta}(1_P) = 1. \tag{11}$$

For $S = \mathrm{conv}(v_0, \ldots, v_j)$ an arbitrary simplex of Δ, let a be the Farey mediant of S. Then the blow-up $\Delta_{(a)}$ of Δ at a is a unimodular f-triangulation.

We shall prove invariance under blow-up,

$$\varsigma_{\theta,\Delta}(f) = \varsigma_{\theta,\Delta_{(a)}}(f). \tag{12}$$

As a consequence of the unimodularity of S we get $\mathrm{den}(a) = \mathrm{den}(v_0) + \cdots + \mathrm{den}(v_j)$. Let $R = \mathrm{conv}(v_0, \ldots, v_j, \ldots, v_m)$ be a maximal m-simplex of Δ containing a. Note that m belongs to $\mathcal{D}$. For each $u = 0, \ldots, j$, let the m-simplex F_u of $\Delta_{(a)}$ be defined by

$$F_u = \mathrm{conv}(v_0, \ldots, v_{u-1}, a, v_{u+1}, \ldots, v_j, v_{j+1}, \ldots, v_m).$$

Then F_u is unimodular and $\mathrm{den}(F_u) = \mathrm{den}(R) \cdot \mathrm{den}(a)/\mathrm{den}(v_u)$. In $\Delta_{(a)}$ the m-simplex R is replaced by (the complex generated by) the maximal m-simplexes $F_0, \ldots, F_j$. Each F_u is maximal in $\Delta_{(a)}$. Since $f \upharpoonright R$ is linear and its coefficients are integers, for each $k = 0, \ldots, m$ there is an integer n_k such that $0 \leq n_k \leq \mathrm{den}(v_k)$ and $f(v_k) = n_k/\mathrm{den}(v_k)$. Similarly, there is an integer $0 \leq n_a \leq \mathrm{den}(a)$ such that $f(a) = n_a/\mathrm{den}(a)$. From the unimodularity of $S \subseteq R$, together with the linearity of f over R, it follows that $n_a = n_0 + \cdots + n_j$. Next we prove

$$\varsigma_R(f) = \sum_{u=0}^{j} \varsigma_{F_u}(f). \tag{13}$$

To this purpose, recalling (3), we first write

$$\lambda_m(F_u) = \frac{1}{m!\,\mathrm{den}(F_u)} = \frac{\mathrm{den}(v_u)}{m!\,\mathrm{den}(R)\,\mathrm{den}(a)}\ , \tag{14}$$

and recalling (7),

$$\overline{f \upharpoonright F_u} = \frac{f(a) - f(v_u) + \sum_{k=0}^{m} f(v_k)}{m+1}\ . \tag{15}$$

A tedious but straightforward computation now yields:

$$\begin{aligned}
\sum_{u=0}^{j} \varsigma_{F_u}(f) &= \sum_{u=0}^{j} c_m \cdot \lambda_m(F_u) \cdot \overline{f \upharpoonright F_u} \\
&= \frac{c_m \sum_u \mathrm{den}(v_u)\,[f(a) - f(v_u) + \sum_{k=0}^{m} f(v_k)]}{(m+1)!\,\mathrm{den}(R)\,\mathrm{den}(a)} \\
&= \frac{c_m \sum_u \mathrm{den}(v_u)\left[\frac{n_a}{\mathrm{den}(a)} - \frac{n_u}{\mathrm{den}(v_u)} + \sum_{k=0}^{m} f(v_k)\right]}{(m+1)!\,\mathrm{den}(R)\,\mathrm{den}(a)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{c_m\left[(\sum_u \operatorname{den}(v_u))\,\frac{n_a}{\operatorname{den}(a)} - \sum_u n_u + (\sum_u \operatorname{den}(v_u))\sum_{k=0}^{m} f(v_k)\right]}{(m+1)!\ \operatorname{den}(R)\ \operatorname{den}(a)}\\
&= \frac{c_m\left[n_a - n_a + \operatorname{den}(a)\sum_{k=0}^{m} f(v_k)\right]}{(m+1)!\ \operatorname{den}(R)\ \operatorname{den}(a)}\\
&= \frac{c_m}{m!\ \operatorname{den}(R)} \cdot \frac{1}{m+1} \cdot \sum_{k=0}^{m} f(v_k) \;=\; c_m \cdot \lambda_m(R) \cdot \overline{f\restriction R} \;=\; \varsigma_R(f).
\end{aligned}
$$

This settles (13) as well as (12).

Having thus proved invariance under blow-up, we now prove invariance under unimodular triangulations:

Claim 1. If Δ' is an f-triangulation of P then $\varsigma_{\theta,\Delta}(f) = \varsigma_{\theta,\Delta'}(f)$.

In order to use (12), we must construct a path from Δ to Δ' only consisting of f-triangulations. To this purpose, let Δ^* be a subdivision of Δ' which is obtained from Δ via a sequence of blow-ups. The existence of Δ^* again follows from the affine version of the De Concini-Procesi theorem [3], [7, p. 252], arguing as in [16, 2.2] (compare with the proof of Proposition 1.2). All triangulations in the path leading from Δ to Δ^* are f-triangulations. By (12) we have

$$\varsigma_{\theta,\Delta}(f) = \varsigma_{\theta,\Delta^*}(f). \tag{16}$$

For each $i \in \mathcal{D}$ every i-simplex L of Δ' is a union of i-simplexes $L_1, \ldots, L_q$ of Δ^*. Let $\Delta'_L \subseteq \Delta'$ be the sub-complex given by all simplexes of Δ' contained in L. Then Δ'_L consists of L together with its faces. Let $\Delta^*_L \subseteq \Delta^*$ be the sub-complex given by all simplexes of Δ^* contained in L. One more application of the solution of the weak Oda conjecture [22, 13.3], [12], shows that Δ'_L and Δ^*_L are connected by a path of blow-ups and blow-downs: each complex obtained in this path is a unimodular triangulation of L, and is also an f-triangulation, because f is linear over L. The same calculation of the proof of (13) yields $\varsigma_L(f) = \varsigma_{L_1}(f) + \cdots + \varsigma_{L_q}(f)$, whence $\varsigma_{\theta,\Delta'}(f) = \varsigma_{\theta,\Delta^*}(f)$. Recalling (16), Claim 1 is settled.

The definition of $\mathcal{P}$. We are now in a position to define the map $\mathcal{P}\colon \theta \mapsto \mathcal{P}_\theta$ by stipulating that, for each formula $\psi = \psi(x_1, \ldots, x_n)$,

$$\mathcal{P}_\theta(\psi) = \varsigma_{\theta,\Delta}(f_\psi \restriction P), \tag{17}$$

where Δ is an arbitrary triangulation of $P = \operatorname{Oneset}(\theta)$ such that f_ψ is linear over each simplex of Δ . By Claim 1, $\mathcal{P}$ is well defined.

Claim 2. $\mathcal{P}_\theta$ is a faithful state of θ.

Suppose $\theta \vdash \psi$, i.e., $f_\psi = 1$ over P. Let Ω be an arbitrary unimodular triangulation of P. From (11) we can write $\varsigma_{\theta,\Omega}(1_P) = 1$, whence $\mathcal{P}_\theta(\psi) = 1$.

In order to prove that $\mathcal{P}_\theta$ has the additivity property, suppose $\theta \vdash \neg(\psi \odot \phi)$. By (5) and Proposition 2.3 we can write $f_{\psi\odot\phi} \restriction P = (f_\psi \odot f_\phi) \restriction P = 0$. In other words, the sum $f_\psi + f_\phi$ is ≤ 1 over P. Let Φ be a unimodular triangulation of P such that both f_ψ and f_ϕ (whence their sum) are linear over each simplex of Φ. Trivially, $(f_\psi + f_\phi) \restriction P$ is uniquely determined by its values at the vertices of Φ. By (7)-(9), for any maximal simplex $T \in \Phi$ we have the identity $\varsigma_T((f_\psi + f_\phi) \restriction P) = \varsigma_T(f_\psi \restriction P) + \varsigma_T(f_\phi \restriction P)$. From (10) and (17) it follows that

$$\mathcal{P}_\theta(\psi \oplus \phi) = \varsigma_{\theta,\Phi}(f_{\psi\oplus\phi} \restriction P) = \varsigma_{\theta,\Phi}((f_\psi \oplus f_\phi) \restriction P) = \varsigma_{\theta,\Phi}((f_\psi + f_\phi) \restriction P)$$

$$= \varsigma_{\theta,\Phi}(f_\psi \restriction P) + \varsigma_{\theta,\Phi}(f_\phi \restriction P) = \mathcal{P}_\theta(\psi) + \mathcal{P}_\theta(\phi).$$

We have just proved that $\mathcal{P}_\theta$ is a state of θ.

There remains to be proved that $\mathcal{P}_\theta$ is faithful. To this purpose, suppose $\theta \nvdash \psi$, i.e., f_ψ is not constantly equal to 1 over P. As above, let Ψ be a triangulation of P such that f_ψ is linear over each simplex of Ψ. For some maximal simplex S of Ψ and some vertex v of S we must have $f_\psi(v) < 1$, whence $\overline{f_\psi \restriction S} < 1$. By (6) and (8) we conclude that $\mathcal{P}_\theta(\psi) < 1$. This settles Claim 2.

5. Conclusion of the proof: $\mathcal{P}$ is invariant

Given $\theta' \in \mathrm{Form}_{n'}$ and an isomorphism $\eta\colon \mathcal{L}_{\theta'} \cong \mathcal{L}_\theta = \mathcal{M}(P)$, let $P' = \mathrm{Oneset}(\theta') \subseteq [0,1]^{n'}$. By Proposition 2.3 we may write without loss of generality $\mathcal{L}_{\theta'} = \mathcal{M}(P')$. For each $i = 1, \ldots, n'$ let $\pi_i \restriction P'$ be the restriction to P' of the ith coordinate function $\pi_i\colon [0,1]^{n'} \to [0,1]$. Then $\eta(\pi_i \restriction P')$ is the restriction f_i to P of some (possibly not unique) McNaughton function $\tilde{f}_i \in \mathcal{M}([0,1]^n)$. Let the continuous function $\tilde{\mathbf{f}}\colon [0,1]^n \to [0,1]^{n'}$ be defined by $x \mapsto \tilde{\mathbf{f}}(x) = (\tilde{f}_1(x), \ldots, \tilde{f}_{n'}(x))$. Let $\mathbf{f} = (f_1, \ldots, f_{n'})$ be the restriction of $\tilde{\mathbf{f}}$ to P. By McNaughton theorem, every function $g \in \mathcal{M}(P')$ has the form $f_\chi \restriction P'$ for some formula $\chi(x_1, \ldots, x_{n'})$. By induction on the number of connectives occurring in χ we obtain

$$\eta(g) = \eta(f_\chi \restriction P') = f_\chi \circ \mathbf{f}, \tag{18}$$

where $\circ$ denotes composition. Suppose $z' \in \mathrm{range}(\mathbf{f}) \setminus P'$ (absurdum hypothesis). Then by [13, 4.17] there is a formula ξ with n' variables, such that f_ξ vanishes over P' and $f_\xi(z') = 1$. By (18), $\eta(f_\xi \restriction P') = f_\xi \circ \mathbf{f} \neq 0$, whereas

$f_\xi \restriction P'$ is the zero element of $\mathcal{M}(P')$, against our assumption about η. We have just proved that $\mathbf{f}$ maps P into P'.

Symmetrically, we have a function $\mathbf{f}'\colon P' \to P$ such that $\eta^{-1}(f_\rho \restriction P) = f_\rho \circ \mathbf{f}'$, for every formula $\rho \in \mathrm{Form}_n$. Since η is an isomorphism, $\mathbf{f}' = \mathbf{f}^{-1}$. Since P is compact, $\mathbf{f}$ is a piecewise linear homeomorphism of P onto P', and each piece of both $\mathbf{f}$ and $\mathbf{f}^{-1}$ has integer coefficients.

By Proposition 1.1, for each $h \in \mathcal{M}(P)$ and $x \in P$, $\iota_{\mathfrak{m}_x}(h/\mathfrak{m}_x) = h(x)$. Since η is an isomorphism, $\mathcal{M}(P')/\mathfrak{m}_{\mathbf{f}(x)} \cong \mathcal{M}(P)/\mathfrak{m}_x$. Once these two quotients are uniquely embedded into $[0,1]$, they coincide. For every rational point $z \in P$, identifying the quotient $\mathcal{M}(P)/\mathfrak{m}_z$ with the MV-algebra $\{h(z) \mid h \in \mathcal{M}(P)\}$, it follows that $\mathcal{M}(P)/\mathfrak{m}_z$ is the subalgebra of $[0,1]$ generated by $\mathrm{den}(z)$. Thus a point $z \in P$ is rational iff so is the point $\mathbf{f}(z) \in P'$, and for all $z \in P \cap \mathbb{Q}^n$ we have

$$\mathrm{den}(\mathbf{f}(z)) = \mathrm{den}(z). \tag{19}$$

Let Σ be a unimodular triangulation of P such that every function $f_1, \ldots, f_{n'}$ is linear over every simplex of Σ. The existence of Σ follows by the same argument as in the proof of Proposition 1.2. Let Σ' be the $\mathbf{f}$-image of Σ. Since each linear piece of every f_i has integer coefficients, Σ' is a rational triangulation of P'.

Fix a simplex $S = \mathrm{conv}(v_0, \ldots, v_j)$ of Σ, with its $\mathbf{f}$-image $S' \in \Sigma'$. Then the (affine) linear map $\mathbf{f}\colon x \in S \mapsto y \in S'$ determines the homogeneous linear map $\mathbf{f}^{\uparrow}_S\colon (x,1) \mapsto (y,1)$. In more detail, let M_S be the $(n'+1) \times (n+1)$ integer matrix whose ith row ($i = 1, \ldots, n'$) is given by the coefficients of the linear (affine) polynomial $f_i \restriction S$, and whose bottom row has the form $(0,0,\ldots,0,0,1)$, with n zeros. Then $M_S(x,1) = (y,1) = \mathbf{f}^{\uparrow}(x,1)$. Let

$$S^{\uparrow} = \mathbb{R}_{\geq 0}\, \tilde{v}_0 + \cdots + \mathbb{R}_{\geq 0}\, \tilde{v}_j \subseteq \mathbb{R}^{n+1}$$

be the positive span of the homogeneous correspondents $\tilde{v}_0, \ldots \tilde{v}_j$ of the vertices of S. In a similar way, let $S'^{\uparrow}$ be the positive span of the vectors $M_S\, \tilde{v}_0, \ldots, M_S\, \tilde{v}_j$. Then M_S sends the set of integer points of $S^{\uparrow}$ one-one into the set of integer points of $S'^{\uparrow}$. Interchanging the roles of P and P' one sees that M_S actually sends integer points of $S^{\uparrow}$ one-one *onto* integer points of $S'^{\uparrow}$. From Blichfeldt's theorem in the Geometry of Numbers [9, p.35] we get the following characterization: S is unimodular iff the half-open parallelepiped $Q_S = \{\mu_0\tilde{v}_0 + \cdots + \mu_j\tilde{v}_j \mid 0 \leq \mu_0, \ldots, \mu_j < 1\}$ contains no nonzero integer points, iff so does its M_S-image $Q_{S'}$, iff S' is unimodular. Since S is unimodular we have proved

Claim 3. For any unimodular triangulation Σ of P such that $\mathbf{f}$ is linear over every simplex of Σ, the image $\Sigma' = \mathbf{f}(\Sigma)$ is a unimodular triangulation of P'.

To conclude the proof of invariance, suppose $\psi' \in \mathrm{Form}_{n'}$, and $|\psi|_\theta = \eta(|\psi'|_{\theta'})$, with the intent of proving $\mathcal{P}_\theta(\psi) = \mathcal{P}_{\theta'}(\psi')$. By Proposition 2.3 we can write $|\psi|_\theta = f_\psi \restriction P$ and $|\psi'|_{\theta'} = f_{\psi'} \restriction P'$. Let Σ satisfy the hypothesis of Claim 3, and have the additional property that f_ψ is linear over every simplex of Σ. The existence of Σ again follows from the affine version of De Concini-Procesi theorem. By Claim 3 and (18), the image $\Sigma' = \mathbf{f}(\Sigma)$ is a unimodular triangulation of P' such that $f_{\psi'}$ is linear over every simplex of Σ'. For each d-simplex $S \in \Sigma$ its correspondent $\mathbf{f}(S) = S'$ is a d-simplex of Σ'. By (3) and (19) we can write

$$\lambda_d(S) = (d!\,\mathrm{den}(S))^{-1} = (d!\,\mathrm{den}(\mathbf{f}(S)))^{-1} = \lambda_d(S'). \tag{20}$$

For all $e \neq d$ both $\lambda_e(S)$ and $\lambda_e(S')$ vanish. Maximal simplexes of Σ correspond via $\mathbf{f}$ to maximal simplexes of Σ'. Computing $\lambda_i(P')$ with the help of the unimodular triangulation Σ' in the light of Theorem 1.3, we obtain $\lambda_i(P) = \lambda_i(P')$ for all $i = 0, 1, 2, \ldots$. Since the dimensional spectra $\mathcal{D}$ and $\mathcal{D}'$ of P and P' are equal, from (6) we get $c_{j,\theta} = c_{j,\theta'}$, $\forall j \in \mathcal{D} = \mathcal{D}'$. Using (18), for all $x \in P$ we can write $f_{\psi'}(\mathbf{f}(x)) = (\eta^{-1}(f_\psi))(\mathbf{f}(x)) = f_\psi(\mathbf{f}^{-1}(\mathbf{f}(x))) = f_\psi(x)$, whence the values of f_ψ and $f_{\psi'}$ at corresponding points $x \in P$ and $x' = \mathbf{f}(x) \in P'$ coincide. For every maximal i-simplex T of Σ, letting $T' = \mathbf{f}(T)$, by (7) and (20) we obtain the identity $\lambda_i(T) \times \overline{f_\psi \restriction T} = \lambda_i(T') \times \overline{f_{\psi'} \restriction T'}$. It follows that $\varsigma_T(f_\psi \restriction T) = \varsigma_{T'}(f_{\psi'} \restriction T')$ and

$$\mathcal{P}_\theta(\psi) = \varsigma_{\theta,\Sigma}(f_\psi \restriction P) = \varsigma_{\theta',\Sigma'}(f_{\psi'} \restriction P') = \mathcal{P}_{\theta'}(\psi').$$

Thus $\mathcal{P}$ is invariant, and the proof of Theorem 3.3 is complete. ■

Since finitely presented MV-algebras coincide with isomorphic copies of the Lindenbaum algebras $\mathcal{L}_\theta$ for some formula θ, combining Propositions 2.3 and 3.1 with Theorem 1.3 we obtain from the foregoing theorem:

COROLLARY 5.1. *Let $A \cong \mathcal{L}_\theta \cong \mathcal{M}(\mathrm{Oneset}(\theta))$ be a finitely presented MV-algebra. Then A has a faithful state, which is also invariant under all automorphisms of A. Further, the sequence of rational numbers*

$$\Lambda(A) = \lambda_0(\mathrm{Oneset}(\theta)),\ \lambda_1(\mathrm{Oneset}(\theta)), \ldots$$

is an invariant *of A, in the sense that $A' \cong A \Rightarrow \Lambda(A') = \Lambda(A)$.*

Recalling Proposition 1.4 we have the following representation of the unconditional fragment of $\mathcal{P}$:

COROLLARY 5.2. *Let $\tau = \tau(x_1, \ldots, x_n)$ and $\psi = \psi(x_1, \ldots, x_n)$ be formulas, with τ a tautology. Then $\mathcal{P}_\tau(\psi)$ is the Lebesgue integral over the n-cube of* *the McNaughton function* f_ψ *,*

$$\mathcal{P}_\tau(\psi) = \int_{[0,1]^n} f_\psi \; .$$

In particular, $\mathcal{P}_\tau(\psi) = 0$ *if and only if* ψ *is the negation of a tautology.*

Final Remarks.

(i) The map $(\theta, \psi) \mapsto \mathcal{P}_\theta(\psi) \in \mathbb{Q}$ is effectively computable.

(ii) The conditional probability $\mathcal{P}_\theta$ is defined even if $\text{Oneset}(\theta) \subseteq [0,1]^n$ is Lebesgue negligible. Our hypothesis that θ is consistent ensures that for some $0 \leq i \leq n$ the i-dimensional volume of θ does not vanish. The consistency of θ is all we need to define $\mathcal{P}_\theta(\psi)$. Note that $\mathcal{P}_\theta = \mathcal{P}_{\theta\odot\theta} = P_{\theta\odot\theta\odot\theta} = \ldots$. Thus $\mathcal{P}_\theta(\psi)$ actually evaluates "the conditional probability of ψ within the set $\mathcal{W}$ of possible worlds assigning the truth-value 1 to θ."

(iii) In case θ is inconsistent, $\mathcal{W}$ is empty; still one might be interested, e.g., in assessing the conditional probability p of ψ within the set of all possible worlds V such that $V(\theta) \geq 1/2$. One then immediately sees that $p = \mathcal{P}_{\theta\oplus\theta}(\psi)$. This is so because $V(\theta) \geq 1/2 \Leftrightarrow V(\theta \oplus \theta) = 1$.

(iv) More generally, let $Q \subseteq [0,1]$ be a finite union of closed intervals with rational endpoints. Let σ be a one-variable formula such that $\text{Oneset}(\sigma) = Q$, as given by Proposition 1.2. As a generalization of (1), Ada's book might refer to the set $\mathcal{W}$ all possible worlds V such that $V(\theta) \in Q$. Thus it makes sense to speak of the probability q of ψ within all possible worlds in $\mathcal{W}$. A moment's reflection shows that q coincides with $\mathcal{P}_{\sigma(\theta)}(\psi)$.

(v) Last, but not least, for any $\psi, \theta, \tau \in \text{Form}_n$ with τ a tautology and θ different from the negation of a tautology (which is a more general condition than θ being consistent) our conditional $\mathcal{P}$ allows the introduction of the quantity $\mathcal{P}_\tau(\psi \wedge \theta)/\mathcal{P}_\tau(\theta)$, measuring the average truth-value of $\psi \wedge \theta$ relative to the average truth-value of θ. Here, as usual, $\wedge$ denotes the derived idempotent conjunction of Ł$_\infty$, as in [2, p. 89]. By Corollary 5.2, $\mathcal{P}_\tau(\theta) \neq 0$. A discussion of the properties of this quantity and its relationship with de Finetti's theory are left for future work.

Acknowledgements. The author is very grateful to the two referees for their competent and careful reading, and for their corrections and suggestions for improvement.

References

[1] Baioletti, M., Capotorti, A., Tulipani, S., Vantaggi, B., 'Simplification rules for the coherent probability assessment problem', *Annals of Mathematics and Artificial Intelligence*, 35: 11–28, 2002.

[2] Cignoli, R.L.O., D'Ottaviano, I.M.L., Mundici, D., *Algebraic Foundations of Many-Valued Reasoning*, Trends in Logic, vol. 7, Kluwer Academic Publishers, Dordrecht, 2000.

[3] De Concini, C., Procesi, C., 'Complete symmetric varieties. II. Intersection theory', in *Algebraic Groups and Related Topics*, Kyoto/Nagoya, 1983, North-Holland, Amsterdam, 1985, pp. 481–513.

[4] de Finetti, B., 'Sul significato soggettivo della probabilitá', *Fundamenta Mathematicae*, 17: 298–329, 1931.

[5] de Finetti, B., 'La prévision: ses lois logiques, ses sources subjectives', *Annales de l'Institut H. Poincaré*, 7: 1–68, 1937. Translated into English by Kyburg, Jr., Henry E., as 'Foresight: Its logical laws, its subjective sources', in: Kyburg, Jr., Henry E., Smokler, Howard E. (eds.), *Studies in Subjective Probability*, Wiley, New York, 1964. Second edition published by Krieger, New York, 1980, pp. 53–118.

[6] de Finetti, B., *Theory of Probability, vol. 1*, John Wiley and Sons, Chichester, 1974.

[7] Ewald, G., *Combinatorial Convexity and Algebraic Geometry*, Springer-Verlag, New York, 1996.

[8] Gerla, B., 'MV-algebras, multiple bets and subjective states', *International Journal of Approximate Reasoning*, 25: 1–13, 2000.

[9] Lekkerkerker, C.G., *Geometry of Numbers*, Wolters-Noordhoff, Groningen and North-Holland, Amsterdam, 1969.

[10] Makinson, D., *Bridges from Classical to Nonmonotonic Logic*, King's College Texts in Computing, vol. 5, 2005.

[11] Milne, P., 'Bruno de Finetti and the logic of conditional events', *Brit. J. Phil. Sci.*, 48: 195–232, 1997.

[12] Morelli, R., 'The birational geometry of toric varieties', *Journal of Algebraic Geometry*, 5: 751–782, 1996.

[13] Mundici, D., 'Interpretation of AF C^*-algebras in Łukasiewicz sentential calculus', *Journal of Functional Analysis*, 65(1): 15–63, 1986.

[14] Mundici, D., 'Averaging the truth value in Łukasiewicz sentential logic', *Studia Logica, Special issue in honor of Helena Rasiowa*, 55: 113–127, 1995.

[15] Mundici, D., 'Bookmaking over infinite-valued events', *International Journal of Approximate Reasoning*, 43: 223–240, 2006.

[16] Panti, G., 'A geometric proof of the completeness of the Łukasiewicz calculus', *Journal of Symbolic Logic*, 60(2): 563–578, 1995.

[17] Panti, G., 'Invariant measures in free MV-algebras', *Communications in Algebra*, to appear. Available at http://arxiv.org/abs/math.LO/0508445v2.

[18] Paris, J., 'A note on the Dutch Book method', in De Cooman, G., Fine, T., Seidenfeld, T. (eds.), *Proc. Second International Symposium on Imprecise Probabilities and their Applications*, ISIPTA 2001, Ithaca, NY, USA, Shaker Publishing

Company, 2001, pp. 301–306. Available at http://www.maths.man.ac.uk/DeptWeb/Homepages/jbp/.

[19] Riečan, B., Mundici, D., 'Probability on MV-algebras', in: Pap, E. (ed.), *Handbook of Measure Theory, vol. II*, North-Holland, Amsterdam, 2001, pp. 869–909.

[20] Tarski, A., Łukasiewicz, J., 'Investigations into the sentential calculus', in *Logic, Semantics, Metamathematics*, Oxford University Press, Oxford, 1956, pp. 38–59. Reprinted by Hackett Publishing Company, Indianapolis, 1983.

[21] Semadeni, Z., *Schauder Bases in Banach Spaces of Continuous Functions*, Lecture Notes in Mathematics, vol. 918, Springer-Verlag, Berlin, 1982.

[22] Włodarczyk, J., 'Decompositions of birational toric maps in blow-ups and blow-downs', *Transactions of the American Mathematical Society*, 349: 373–411, 1997.

[23] Wójcicki, R., 'On matrix representations of consequence operations of Łukasiewicz sentential calculi', *Zeitschrift für Math. Logik und Grundlagen der Mathematik*, 19: 239–247, 1973. Reprinted in Wójcicki, R., Malinowski, G. (eds.), *Selected Papers on Łukasiewicz Sentential Calculi*, Ossolineum, Wrocław, 1977, pp. 101–111.

D. Mundici
Department of Mathematics "Ulisse Dini"
University of Florence
Viale Morgagni 67/A
50134 Florence, Italy
`mundici@math.unifi.it`

Stephan van der
Waart van Gulik

A Fuzzy Logic Approach to Non-scalar Hedges

Abstract. In [4], George Lakoff proposes a fuzzy semantics for the non-scalar hedges *technically*, *strictly speaking*, and *loosely speaking*. These hedges are able to modify the meaning of a predicate. However, Lakoff's proposal is problematic. For example, his semantics only contains interpretations for hedged predicates using semantic information provided by selection functions. What kind of information these functions should provide for non-hedged predicates remains unspecified. This paper presents a solution for this deficit and other problems by means of a generic first-order fuzzy logic $\mathbf{FL_h}$. A wide range of fuzzy logics can be used as a basis for $\mathbf{FL_h}$. Next to a fully specified semantics, this solution also incorporates a proof theory for reasoning with these hedges. $\mathbf{FL_h}$ makes use of a special set of selection functions. These functions collect the kind of information a reasoner can retrieve from concepts in his or her memory when interpreting a (non-)hedged predicate. Despite this non-standard element, $\mathbf{FL_h}$ remains a conservative modification of its underlying fuzzy logic.

Keywords: fuzzy logic, non-scalar hedges, fuzzy concepts, cognitive science.

1. Introduction

In [4], George Lakoff analyzes the semantics of a specific kind of linguistic modifiers, namely those terms or phrases which modify the meaning of a lexical item. He calls these modifiers *hedges*. Consider the following examples.

(1.a) "Technically*, it's a bird.*"
(1.b) "*It's a* regular *bird.*"
(1.c) "Loosely speaking*, it's a bird.*"

On the basis of his analysis, Lakoff proposes a formal fuzzy semantics for the hedges *technically*, *strictly speaking*, and *loosely speaking*. His proposal however is problematic. For example, his semantics does not deal with non-hedged predicates. It only contains interpretations for hedged predicates using semantic information provided by special selection functions. What kind of information these functions should provide for non-hedged predicates remains unspecified. In this paper, I present a solution for this deficit and other problems by means of a generic first-order fuzzy logic $\mathbf{FL_h}$ based upon

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*, Trends in Logic, 28 © Springer Science + Business Media B.V. 2009

a fuzzy logic **FL**. A wide range of fuzzy logics can be used as **FL**. Moreover, not only does this solution include a fully specified fuzzy semantics for the hedges *technically*, *strictly speaking*, and *loosely speaking*, it also incorporates a proof theory for reasoning with these hedges.

The paper is structured as follows. In section 2, I first give an outline of Lakoff's proposal and clarify its relation with research in cognitive science and fuzzy logic. Next, I discuss its problems. In section 3, I introduce and explain some new formal machinery that plays a special role in $\mathbf{FL_h}$ (including a way of representing logically relevant information stored in the fuzzy concepts of predicates). I also discuss some intuitive relations between the hedged and non-hedged usage of predicates that should hold in a proper logic. Section 4 contains the formal characterization of $\mathbf{FL_h}$ as well as several meta-theorems which confirm the intuitive relations from section 3. In section 5, I sum up the main results and present some open research questions.

2. Lakoff's proposal

2.1. An outline

Consider the following sentences.

(2.a) *"Technically, Richard Nixon is a Quaker."*
(2.b) *"Strictly speaking, Richard Nixon is a Quaker."*
(2.c) *"Strictly speaking, a whale is a mammal."*
(2.d) *"Loosely speaking, a whale is a fish."*

According to Lakoff, sentence (2.a) is true. Nixon is a Quaker 'in some definitional sense' because a definitional criterium associated with the predicate Quaker, i.e. the predicate *Born-into-Quaker-Family*, can be applied to Nixon. However, people also tend to associate the embracement of pacifism with the predicate *Quaker*, as this is also characteristic of most Quakers. However, the predicate *Pacifist* cannot be applied to Nixon without great controversy. This is the reason why sentence (2.b) is not true. Given these insights, Lakoff concludes that a predicate is used in a technical sense iff all its associated predicates of definitional importance are applicable and at least one associated predicate of primary importance is not. A predicate is used in a strict sense iff all its associated predicates of definitional and primary importance are applicable.

Lakoff analyzes the difference between the semantics of the hedges *strictly speaking* and *loosely speaking* in a similar way. In his view, sentences (2.c) and (2.d) are both true. Sentence (2.c) states that whales can be classified as mammals when both definitional and other important predicates for distinguishing mammals are taken into account (e.g. *Gives-life-birth* and *Breathes-air*). Interestingly, sentence (2.d) states that whales can also be classified as fish in a loose sense. When associated predicates of secondary importance are taken into account (e.g. *Lives-in-water*), whales can be interpreted as a kind of fish, despite the obvious fact that many predicates of definitional and primary importance associated with fish (e.g. *Gills*) cannot be applied to whales. Hence, Lakoff concludes that a predicate is used in a loose sense iff all its associated predicates of secondary importance are applicable and at least one associated predicate of definitional or primary importance is not.

By means of these insights, Lakoff suggests a formal fuzzy semantics for the hedges *technically*, *strictly speaking*, and *loosely speaking*.[1] Let π be a unary predicate and let $\mu_\pi : D \rightarrow [0,1]$ be the membership function characterizing its fuzzy extension, where D is a non-empty set and $[0,1] = \{x \mid 0 \leq x \leq 1 \text{ and } x \in \mathbb{R}\}$. Now presuppose a predicate π that is linked with a set of associated predicates $\{\pi_1, ..., \pi_n\}$. Let $\overline{\pi} = \{\mu_{\pi_1}, ..., \mu_{\pi_n}\}$ be the set of the membership functions of the extensions of these associated predicates. Furthermore, let *def*(*initional*), *prim*(*ary*) and *sec*(*ondary*) be selection functions which select the appropriate membership functions from $\overline{\pi}$ in order to generate the membership function of a new fuzzy set. Take the minimum function *min* to construct this function. For example, if *def* picks out μ_{π_1}, μ_{π_2} and μ_{π_3} in the case of π, then $def(\pi) = min(\mu_{\pi_1}, min(\mu_{\pi_2}, \mu_{\pi_3}))$. Note that *min* captures the truth-functionality of a fuzzy conjunction. Finally, let *neg* be a standard order inversion $1 - x$, where $x \in [0,1]$. Remark that *neg* captures the truth-functionality of a fuzzy negation.

(1) $\mu_{tech(\pi)} =_{df} min(def(\overline{\pi}), neg(prim(\overline{\pi})))$
(2) $\mu_{strict(\pi)} =_{df} min(def(\overline{\pi}), prim(\overline{\pi}))$
(3) $\mu_{loos(\pi)} =_{df} min(sec(\overline{\pi}), neg(min(def(\overline{\pi}), prim(\overline{\pi}))))$

It is easy to see that these definitions perfectly correspond to Lakoff's conclusions concerning the semantics of *technically*, *strictly speaking*, and *loosely speaking*. For example, in line with Lakoff's conclusion regarding the semantics of *technically* in sentence (2.a), definition (1) demands that a predicate is used in a technical sense iff all its associated definitional pred-

---

[1] Lakoff also discusses the semantics of the hedge *regularly*. I do not discuss this hedge as it is not clear at the moment how it can be integrated in a truth-functional logic like $\mathbf{FL_h}$.

icates collected by *def* are applicable and at least one associated predicate of primary importance selected by *prim* is not.

2.2. Its relation with cognitive science and fuzzy logic

I first clarify the relation between Lakoff's proposal and research in cognitive science. In Lakoff's proposal, the type of predicate that may be modified by hedges like *technically*, *strictly speaking* and *loosely speaking* is necessarily linked with a set of associated predicates. This set is often interpreted as the *concept* of the predicate, cf., for instance, [5]. I call this type of predicate a *complex predicate*. It stands in contrast to what I call a *scalar predicate*. This type of predicate cannot be analyzed in terms of other predicates and is simply linked to some scale on which its applicability is set out (e.g. the application of the primitive predicate *Red* in function the level of perceived 'reddishness'). Lakoff also presupposes that each predicate in the concept of a complex predicate π has a specific level of importance. This level is based on how characteristic the predicate in casu is for things that are known to be π. This idea is related with classic models in cognitive science concerning the development and structure of fuzzy concepts.

A good example is Fintan Costello's Diagnostic Evidence Model, cf. [2]. I briefly illustrate the main idea behind this model. Consider the predicates *Fly* and *Bird*. The more birds are observed flying, and the less other things are observed flying, the more characteristic and, hence, important *Fly* becomes in the concept of *Bird*. The fuzzy truth of the expression α *is a Bird*, i.e. the level up to which *Bird* can be applied to an instance α, directly depends on the levels up to which associated predicates like *Fly*, *Feathered*, *Beak*, etc. are applicable to a (possibly weighted in function of their relative importance). Note that most concepts do not contain predicates with absolute, or *definitional* importance, i.e. they do not contain a *definitional core*.[2]

With respect to research in fuzzy logic, it is important to keep in mind that the type of hedges analyzed by Lakoff are not the same as those discussed by Lofti Zadeh in [10]. Zadeh focuses on hedges which operate exclusively on scalar predicates. Consequentially, these hedges can be best distinguished as *scalar hedges*. Examples are intensifiers and de-intensifiers like *very*, *sort of*, etc. Lakoff instead focuses on hedges that operate on

---

[2] In many (natural language) categories, each member of the category has at least one, and probably several, properties in common with one or more other members, but often no, or few, properties are common to all members, cf. Eleanor Rosch's well-known research in, for instance, [7] and [8]. Remark that Lakoff himself explicitly uses Rosch's experimental research in [6] to frame his proposal.

complex predicates, i.e. *non-scalar hedges*. This type of hedges can be understood best as narrowing down or loosening the meaning of a complex predicate, or even generating a kind of meaning shift (see also subsections 3.3 and 4.6).

2.3. Problems

Lakoff's formal fuzzy semantics for the hedges *technically*, *strictly speaking*, and *loosely speaking* is problematic. First of all, his semantics is not fully specified. As already mentioned, it does not deal with non-hedged complex predicates. It is not specified what kind of semantic information the selection functions should provide when interpreting non-hedged predicates. Note that this makes it impossible to specify and check intuitions concerning the logical relations between the hedged and non-hedged usage of complex predicates. Secondly, the application radius of *strictly speaking*, and *loosely speaking* is too much restricted. Reconsider definitions (2) and (3) in subsection 2.1. Remark that the selection function *def* is used in both definitions to generate an argument for the binary function *min*. If *def* would be allowed to put in nothing, it would become possible for *min* to lack an argument. Hence, only those predicates that have at least one predicate of definitional importance in their concept are within the range of application. This is strange, as *strictly speaking* and *loosely speaking* clearly may also operate on predicates of which the concept does not own predicates of definitional importance. Take for instance the predicate *Game*. As Ludwig Wittgenstein has argued extensively, this is a good example of a predicate for which it is very hard, maybe even impossible, to conceive a set of singly necessary and jointly sufficient criteria, cf. §3 in [9]. In other words, the concept of *Game* lacks a definitional core. Yet, it is intuitively correct to state phrases like *"Strictly speaking, it's a game."* or *"Loosely speaking, it's a game."* Hence, Lakoff's proposal applies to an unrealistically small set of complex predicates.

In the next sections, I present a solution for these problems by means of the logic $\mathbf{FL_h}$. In section 3, I introduce some new formal machinery that has a special role in $\mathbf{FL_h}$. Section 4 contains the actual characterization of $\mathbf{FL_h}$.

3. Some new machinery

3.1. Selection functions

As already mentioned, there are two main types of predicates: scalar predicates and complex predicates. In contrast to Lakoff, I further divide the latter type into those predicates of which the concept possesses a definitional core, i.e. those of which the concept owns at least one predicate of definitional importance, and those that do not. In sum, I use the following sets of predicates: the set of scalar unary predicates $\mathcal{P}^s$, the set of complex unary predicates with a definitional core $\mathcal{P}^d$, and the set of complex unary predicates without a definitional core $\mathcal{P}^h$. The union of all previous sets $\mathcal{P}^s \cup \mathcal{P}^d \cup \mathcal{P}^h$ is called $\mathcal{P}$.

I also introduce a *set of selection functions* S. Informally speaking, these functions collect the kind of information a reasoner can retrieve from fuzzy concepts in his or her memory when interpreting a (non-)hedged complex predicate. For every complex predicate, the selection functions d, h, p, and s respectively select the associated predicates of definitional or high importance, primary importance and secondary importance. The set is defined as follows.

Definition 3.1. A set of selection functions S is a non-empty set $\{d, h, p, s\}$ which complies with the following conditions (ρ, ρ_i, $\rho_j \in S$ and $\pi \in \mathcal{P}$):

(a) $\rho : \mathcal{P}^d \cup \mathcal{P}^h \to \wp(\mathcal{P})$,
(b) for each $\pi \in \mathcal{P}^d$: $\rho(\pi) \neq \emptyset$,
(c) for each $\pi \in \mathcal{P}^h$: $d(\pi) = \emptyset$,
(d) for each $\pi \in \mathcal{P}^h$: $\rho(\pi) \neq \emptyset$ if $\rho \neq d$,
(e) for each π: $\rho_i(\pi) \cap \rho_j(\pi) = \emptyset$ where $\rho_i \neq \rho_j$,
(f) π is of type 0 iff $\pi \in \mathcal{P}^s$; π is of type $n + 1$ iff the maximum type of the predicates in $d(\pi) \cup h(\pi) \cup p(\pi) \cup s(\pi)$ equals n.

I briefly explain each condition. Condition (a) demands that the concept of each complex predicate only consists of complex predicates and scalar predicates. Condition (b) states that for every predicate with a definitional core, each selection function should pick up a non-empty set of predicates. Conditions (c) and (d) together state that for every predicate without a definitional core, all selection functions should pick up a non-empty set of predicates except for selection function d, which should pick up the empty set (given the absence of a core). The motivation behind the conditions (b)–(d) is technical. They make sure that it is impossible that complex predicates are defined in terms of empty predicate sets in the definitions (Dh1)–(Dh7) in subsection 3.2. Condition (e) demands that every two different predicate sets selected in function of a complex predicate are disjoint. To let these sets overlap is pointless and overly complex, both from a logical and a cognitive science perspective. Condition (f) establishes a recursive structure. Scalar predicates are of type 0, complex predicates are of type $n > 0$ and the concept of a predicate of type n may only consist of predicates of type $n-1$ or smaller. This condition makes sure that the analysis of a concept cannot go on infinitely (but eventually has to end with a set of primitive scalar predicates). Given the finite nature of our cognitive machinery, this seems a reasonable demand. The condition also avoids conceptual circularity. The concept of a complex predicate π cannot be based on π itself.

For some readers this type of selection functions might suggest a rather idealized kind of concept structures. However, for the present purpose, it does the job. Moreover, the set-up can still be modified in the future to fit more realistic standards.

3.2. Modified definitions

I now define the interpretations of (non-)hedged complex predicates (t, s and l respectively stand for *technically*, *strictly speaking* and *loosely speaking*. π_i, $\pi_j \in \mathcal{P}^d \cup \mathcal{P}^h$. α is a term, and $\neg$ and $\&$ are the fuzzy negation and conjunction of the **FL** of choice, see also subsection 4.2).[3]

(Dh1) $\pi\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in d(\pi)\}$, $(\pi \in \mathcal{P}^d)$
(Dh2) $\pi\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in h(\pi)\}$, $(\pi \in \mathcal{P}^h)$
(Dh3) $\pi^t\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in d(\pi)\}\&\neg\boldsymbol{\&}\{\pi_j\alpha \mid \pi_j \in p(\pi)\}$, $(\pi \in \mathcal{P}^d)$
(Dh4) $\pi^s\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in d(\pi) \cup p(\pi)\}$, $(\pi \in \mathcal{P}^d)$
(Dh5) $\pi^s\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in h(\pi) \cup p(\pi)\}$, $(\pi \in \mathcal{P}^h)$
(Dh6) $\pi^l\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in s(\pi)\}\&\neg\boldsymbol{\&}\{\pi_j\alpha \mid \pi_j \in d(\pi) \cup p(\pi)\}$, $(\pi \in \mathcal{P}^d)$
(Dh7) $\pi^l\alpha =_{df} \boldsymbol{\&}\{\pi_i\alpha \mid \pi_i \in s(\pi)\}\&\neg\boldsymbol{\&}\{\pi_j\alpha \mid \pi_j \in h(\pi) \cup p(\pi)\}$, $(\pi \in \mathcal{P}^h)$

The definitions (Dh3)–(Dh7) are based upon Lakoff's original proposal. There is only one major difference. In the cases of *strictly speaking* and *loosely speaking* I differentiate between predicates that own a definitional core, cf. (Dh4) and (Dh6), and those that do not, cf. (Dh5) and (Dh7). In the latter type, the highly important predicates take over the role of the definitional core. In contrast to Lakoff, I also provide interpretations for non-hedged complex predicates, cf. (Dh1) and (Dh2). In (Dh1) the definitional core is used, in (Dh2) the highly important predicates are selected. Another possibility would be to simply collect all predicates associated with the complex predicate in question. However, it is well-known that the most important predicates in a concept are also the most easily accessible ones, cf., for instance, [1]. Hence, it is more likely that only the definitional or highly important predicates are taken into account during interpretation, as

---

[3] I use **&** for the &-conjunction of all formulas in a given set.

these are retrieved relatively fast and unhindered.[4] The presence of a hedge implies some extra cognitive effort during interpretation because also less important predicates need to be retrieved from the concept.

Remark that by means of the definitions (Dh1)–(Dh7) and the recursive structure established by S, every expression $\pi^{(h)}\alpha$ is eventually equated with a formula that only consists of scalar predicates, where $\pi \in \mathcal{P}^d \cup \mathcal{P}^h$, $h \in \mathcal{H}$ and (h) denotes the possible presence of a hedge. In other words, I presuppose that the interpretation of a (non-)hedged complex predicate is eventually based upon a set of primitive scalar predicates.

3.3. Some logical intuitions

Given definitions (Dh1) and (Dh2), it becomes possible to specify some intuitions concerning the logical relations between the hedged and non-hedged usage of complex predicates. A proper logic should affirm these relations. First of all, both *technically* as well as *strictly speaking* evidently narrow down the meaning of a complex predicate as they invoke some extra conditions on top of those used in the non-hedged case. Hence, it is correct to expect that $\pi^h\alpha$ implies $\pi\alpha$, where h is t or s, but not vice versa. For example, if someone is technically a Quaker, that person can also be called a Quaker, but not vice versa. Secondly, the hedge *loosely speaking* seems to generate some kind of meaning shift. The focus lies on the predicates of secondary importance and it is possible that predicates critical for the non-hedged case cannot be applied. Hence, it is correct to expect that $\pi^l\alpha$ cannot imply $\pi\alpha$, nor vice versa. For example, it is not correct to state that a whale is a fish because it is a fish in some loose sense. Likewise, it is not correct to state that because Nixon is a Quaker, Nixon is also a Quaker in a loose sense. In subsection 4.6, I show that $\mathbf{FL_h}$ confirms these relations.

4. The generic fuzzy logic for non-scalar hedges $\mathbf{FL_h}$

4.1. Preliminaries

In this section, I characterize $\mathbf{FL_h}$. This logic is based upon a first-order fuzzy logic **FL**. Many fuzzy logics can serve as **FL**. One possibility is to use a t-norm logic, cf. Petr Hájek's [3]. A t-norm logic is characterized by a t-norm operator $\otimes$ which determines the truth-functionality of the conjunction. It is defined as follows ($x, y, z \in [0, 1]$).

---

[4] This makes even more sense when taking into account the fact that people often need to interpret statements under time pressure (e.g. during fast and sloppy communication).

Definition 4.1. A t-norm $\otimes$ is a binary operator that satisfies the following properties:

1. $\otimes : [0,1]^2 \rightarrow [0,1]$,
2. $x \otimes y = y \otimes x$ (commutativity),
3. $x \otimes (y \otimes z) = (x \otimes y) \otimes z$ (associativity),
4. If $x \leq y$, then $x \otimes z \leq y \otimes z$ (monotonicity),
5. $1 \otimes x = x$ (neutral element).

The t-norm of choice determines a specific residuation operator $\Rightarrow$ which determines the truth-functionality of the implication. It is defined as follows.

Definition 4.2. A residuation operator $\Rightarrow$ is a binary operator that satisfies the following properties:

1. $\Rightarrow: [0,1]^2 \rightarrow [0,1]$,
2. $x \Rightarrow y = max\{z \mid z \otimes x \leq y\}$ (residuation).

Note that the residuation operator also has the following derivable properties (capturing classical behavior for 0 and 1): (a) $x \Rightarrow y = 1$ iff $x \leq y$, (b) if $x = 1$ and $y = 0$, then $x \Rightarrow y = 0$.

In the rest of the section, I use the concrete logic $\mathbf{BL}\forall_\mathbf{h}$ to illustrate $\mathbf{FL_h}$. $\mathbf{BL}\forall_\mathbf{h}$ is based on the t-norm logic $\mathbf{BL}\forall$, which is a well-known logic with several interesting extensions (e.g. Łukasiewicz logic **Ł**$\forall$, Goguen logic $\mathbf{\Pi}\forall$ and Gödel logic $\mathbf{G}\forall$).

4.2. The language schema of $\mathbf{FL_h}$

The language schema $\mathcal{L}_h$ of $\mathbf{FL_h}$ makes use of the following types of non-logical symbols (I do not use n-ary predicates with $n > 1$, identity or functions because they are not important in this context):

- $\mathcal{C}$: the set of constants,
- $\mathcal{V}$: the set of variables,
- $\mathcal{P}^s$: the set of scalar predicates,
- $\mathcal{P}^d$: the set of complex predicates with a definitional core,
- $\mathcal{P}^h$: the set of complex predicates without a definitional core,
- $\overline{0}, \overline{1}$: false and true,
- $\mathcal{H}$: the set of the non-scalar hedges t(*echnically*), s(*trictly*) and l(*oosely*).

The schema is closed under the connectives $\neg$, &, $\vee$, $\wedge$ $\rightarrow$, $\leftrightarrow$ and the quantifiers $\exists$ and $\forall$. The set of $\mathbf{FL_h}$-formulas $\mathcal{L}_h$ is the smallest set which satisfies the following conditions ($\pi \in \mathcal{P}$, $\alpha \in \mathcal{C} \cup \mathcal{V}$):

1. $\pi\alpha, \overline{0}, \overline{1} \in \mathcal{L}_h$,
2. if $\pi \in \mathcal{P}^d$, then $\pi^t\alpha \in \mathcal{L}_h$,
3. if $\pi \in \mathcal{P}^d \cup \mathcal{P}^h$, then $\pi^s\alpha, \pi^l\alpha \in \mathcal{L}_h$,
4. if $A \in \mathcal{L}_h$, then $\neg A \in \mathcal{L}_h$,
5. if $A, B \in \mathcal{L}_h$, then $A\&B$, $A \to B$, $A \wedge B$, $A \vee B$, $A \leftrightarrow B \in \mathcal{L}_h$,
6. if $A \in \mathcal{L}_h$ and $\alpha \in \mathcal{V}$, then $(\forall\alpha)A$, $(\exists\alpha)A \in \mathcal{L}_h$.

$\mathcal{W}_h$ is the set of closed $\mathcal{L}_h$-formulas.

This linguistic set-up is supplemented with the definitions (Dh1)–(Dh7). Let me stress again that by means of these definitions and the recursive structure established by S, every expression $\pi^{(h)}\alpha$, where $\pi \in \mathcal{P}^d \cup \mathcal{P}^h$, is eventually equated with a formula that only consists of scalar predicates.

4.3. The structure of a $\mathbf{FL_h}$-theory

DEFINITION 4.3. A $\mathbf{FL_h}$-theory is a couple $\langle\Gamma, S\rangle$. The first element Γ is a set of $\mathcal{W}_h$-formulas. The second element is a selection function set S.

4.4. The proof theory of $\mathbf{FL_h}$

The proof theory of $\mathbf{FL_h}$ is a conservative modification of the proof theory of **FL**. There is only one difference. $\mathbf{FL_h}$ implements alternative definitions of the notions *theoremhood* and *derivability* that take into account the presence of a selection function set S. Consider the proof theory of $\mathbf{BL\forall_h}$ based on $\mathbf{BL\forall}$. The axioms, rules and definitions of $\mathbf{BL\forall_h}$ are simply those of $\mathbf{BL\forall}$ $(\beta \in \mathcal{C})$.

(A1) $(A \to B) \to ((B \to C) \to (A \to C))$
(A2) $(A\&B) \to B$
(A3) $(A\&B) \to (B\&A)$
(A4) $(A\&(A \to B)) \to (B\&(B \to A))$
(A5) $(A \to (B \to C)) \to ((A\&B) \to C)$
(A6) $((A\&B) \to C) \to (A \to (B \to C))$
(A7) $((A \to B) \to C) \to (((B \to A) \to C) \to C)$
(A8) $\overline{0} \to A$
(∀1) $(\forall\alpha)A(\alpha) \to A(\beta)$ (β substitutable for α in $A(\alpha)$)
(∃1) $A(\beta) \to (\exists\alpha)A(\alpha)$ (β substitutable for α in $A(\alpha)$)
(∀2) $(\forall\alpha)(B \to A) \to (B \to (\forall\alpha)A)$ (α not free in B)
(∃2) $(\forall\alpha)\ (A \to B) \to ((\exists\alpha)A \to B)$ (α not free in B)
(∀3) $(\forall\alpha)(A \vee B) \to ((\forall\alpha)A \vee B)$ (α not free in B)

(MP) From A and $A \rightarrow B$ derive B
(UG) From A, derive $(\forall\alpha)A$

(D1) $\neg A =_{df} A \rightarrow \overline{0}$
(D2) $A \wedge B =_{df} A\&(A \rightarrow B)$
(D3) $A \vee B =_{df} ((A \rightarrow B) \rightarrow B)\&((B \rightarrow A) \rightarrow A)$
(D4) $A \leftrightarrow B =_{df} (A \rightarrow B)\&(B \rightarrow A)$

In order to complete the proof theory of $\mathbf{BL}\forall_{\mathbf{h}}$, the following $\mathbf{FL_h}$-definitions for *theoremhood* and *derivability* are added.

DEFINITION 4.4. $\vdash_{\mathbf{FL_h}} A$ iff there is a proof of A from $\langle\varnothing, S\rangle$ for each possible S: i.e., for each possible set S, there exists a sequence of formulas that ends with A in which every member is either an axiom or follows from previous members of the sequence by means of a rule.

DEFINITION 4.5. $\langle\Gamma, S\rangle \vdash_{\mathbf{FL_h}} A$ iff there is a proof of A from $\langle\Gamma, S\rangle$: i.e., there exists a sequence of formulas that ends with A in which every member is either an axiom, a member of Γ or follows from previous members of the sequence by means of a rule.

Remark that definition 4.4 makes it impossible that derivations that are based solely on the information in a particular S are interpreted as theorems of $\mathbf{FL_h}$. For example, if $Bird \in \mathcal{P}^h$ and Fly is selected by $h(Bird)$ in some S, it holds that $\langle\emptyset, S\rangle \vdash_{\mathbf{FL_h}} (\forall x)Birdx \rightarrow Flyx$. However, this expression has nothing to do with tautological truths in $\mathbf{FL_h}$. It is only a conceptual truth (or 'analytical truth', as some like to say). It is always possible to use some other S' in which $Fly \notin h(Bird)$. The same idea holds for the semantic notion *validity*, cf. definition 4.7.

4.5. The semantics of $\mathbf{FL_h}$

The semantics of $\mathbf{FL_h}$ is also a conservative modification of the semantics of $\mathbf{FL}$. There are only two differences: (1) the $\mathbf{FL_h}$-valuation function is both determined by a model M as well as a selection function set S, and (2) $\mathbf{FL_h}$ implements alternative definitions for the notions *truth in a model*, *validity* and *semantic consequence* that take into account the presence of a selection function set S. Consider the semantics of $\mathbf{BL}\forall_{\mathbf{h}}$. A model in $\mathbf{BL}\forall_{\mathbf{h}}$ is build up in the same way as a model in the original semantics of $\mathbf{BL}\forall$. In both cases, a model M is defined by a couple $\langle D, v\rangle$. D is a non-empty set. v is an assignment function that complies with the following conditions.

(i) $v : \mathcal{C} \cup \mathcal{V} \rightarrow D$
(ii) $v : \mathcal{P}^s \rightarrow (D \rightarrow [0,1])$

As already mentioned, in $\mathbf{FL_h}$, the valuation function $v_{MS} : \mathcal{L}_h \rightarrow [0,1]$ is determined by a model M and a selection function set S. In the case of $\mathbf{BL\forall_h}$ this function complies with the following conditions.

S.1 $v_{MS}(\pi\alpha) = v(\pi)(v(\alpha))$, where $\pi \in \mathcal{P}^s$
S.2 $v_{MS}(\overline{0}) = 0$
S.3 $v_{MS}(\overline{1}) = 1$
S.4 $v_{MS}(A\&B) = v_{MS}(A) \otimes v_{MS}(B)$
S.5 $v_{MS}(A \rightarrow B) = v_{MS}(A) \Rightarrow v_{MS}(B)$
S.6 $v_{MS}((\exists\alpha)A) = \mathit{Sup}\{v_{M'S}(A) \mid M' = \langle D, v'\rangle$ differs at most from M in that possibly $v'(\alpha) \neq v(\alpha)\}$
S.7 $v_{MS}((\forall\alpha)A) = \mathit{Inf}\{v_{M'S}(A) \mid M' = \langle D, v'\rangle$ differs at most from M in that possibly $v'(\alpha) \neq v(\alpha)\}$

The set of designated values W of $\mathbf{FL_h}$ is identical to the one of $\mathbf{FL}$. In the case of $\mathbf{BL\forall_h}$ this means that $W = \{1\}$. In order to complete the semantics of $\mathbf{BL\forall_h}$, the following $\mathbf{FL_h}$-definitions for the notions *truth in a model*, *validity* and *semantic consequence* are added.

Definition 4.6. $\langle M, S\rangle \models_{\mathbf{FL_h}} A$ iff $v_{MS}(A) \in W$.

Definition 4.7. $\models_{\mathbf{FL_h}} A$ iff $v_{MS}(A) \in W$ in all possible M, under all possible S.

Definition 4.8. $\langle \Gamma, S\rangle \models_{\mathbf{FL_h}} A$ iff $v_{MS}(A) \in W$ in all M where $v_{MS}(B) \in W$, for all $B \in \Gamma$.

Note that v does not assign fuzzy extensions to complex predicates. Only scalar predicates are assigned an extension. It are the definitions (Dh1)–(Dh7), combined with the recursive structure established by S, that make sure that the valuation of every expression $\pi^{(h)}\alpha$, where $\pi \in \mathcal{P}^d \cup \mathcal{P}^h$, ultimately depends on the extensions of a set of scalar predicates. In this way, the semantics directly reflects the philosophical idea that the interpretation of a (non-)hedged complex predicate ultimately depends on the primitive meaning of some related set of scalar predicates.

4.6. Some interesting meta-theorems of $\mathbf{FL_h}$

I first prove two general properties of the t-norm $\otimes$ ($x, y, z \in [0,1]$).

Lemma 4.9. $x \otimes y \leq x$

Proof. Property (4) in definition 4.1 states that $x \otimes z \leq y \otimes z$, if $x \leq y$. Property (5) in definition 4.1 states that $1 \otimes x = x$. Suppose $y = 1$. Hence, $x \otimes z \leq z$. Given property (2) in definition 4.1, this implies $z \otimes x \leq z$. ■

Lemma 4.10. $x \otimes 0 = 0$

Proof. Given lemma 4.10, $y \otimes x \leq y$. Hence, if $y = 0$, then $0 \otimes x = 0$ and, given property (2) of $\otimes$ in definition 4.1, $x \otimes 0 = 0$. ■

I now prove some meta-theorems of $\mathbf{FL_h}$ which confirm the logical relations discussed in subsection 3.3 ($\pi_i, \pi_j \in \mathcal{P}$ and *max* is a metavariable for the functions d and h in those cases where both functions are possible).

Theorem 4.11. $\models_{\mathbf{FL_h}} \pi^t\alpha \rightarrow \pi\alpha$

Proof. Observe that $v_{MS}(\pi^t\alpha \rightarrow \pi\alpha) = v_{MS}(\pi^t\alpha) \Rightarrow v_{MS}(\pi\alpha)$. Now presuppose that $v_{MS}(\pi^t\alpha) = 1$ and $v_{MS}(\pi\alpha) = 0$, for some model M and some selection function set S. This would make $v_{MS}(\pi^t\alpha) \Rightarrow v_{MS}(\pi\alpha) = 0$ because of property (b) of $\Rightarrow$. However, given definitions (Dh3) and (Dh1), $v_{MS}(\pi^t\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\}) \otimes v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\})$ and $v_{MS}(\pi\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\})$. Given lemma 4.9, $v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\}) \otimes v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\}) \leq v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\})$. Hence, the presupposition is impossible and, given property (a) of $\Rightarrow$, $\models_{\mathbf{FL_h}} \pi^t\alpha \rightarrow \pi\alpha$ holds. ■

Theorem 4.12. $\models_{FL_h} \pi^s\alpha \rightarrow \pi\alpha$

Proof. Similar to the proof of theorem 4.11. ■

Theorem 4.13. $\not\models_{FL_h} \pi\alpha \rightarrow \pi^t\alpha$

Proof. Given definitions (Dh1) and (Dh3), $v_{MS}(\pi\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\})$ and $v_{MS}(\pi^t\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\}) \otimes v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\})$. Let $v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in d(\pi)\}) = 1$ and $v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\}) = 0$, for some model M and some selection function set S. Hence, $v_{MS}(\pi^t\alpha) = 1 \otimes 0$. Given lemma 4.10, this implies $v_{MS}(\pi^t\alpha) = 0$. So, $v_{MS}(\pi\alpha \rightarrow \pi^t\alpha) = v_{MS}(\pi\alpha) \Rightarrow v_{MS}(\pi^t\alpha) = 1 \Rightarrow 0$. Hence, because of property (b) of $\Rightarrow$, $v_{MS}(\pi\alpha \rightarrow \pi^t\alpha) = 0$ and $\not\models_{FL_h} \pi\alpha \rightarrow \pi^t\alpha$ holds. ■

Theorem 4.14. $\not\models_{FL_h} \pi\alpha \rightarrow \pi^s\alpha$

Proof. Similar to the proof of theorem 4.13. ■

Theorem 4.15. $\not\models_{FL_h} \pi^l\alpha \rightarrow \pi\alpha$

Proof. Given definitions (Dh6) and (Dh7) as well as (Dh1) and (Dh2), $v_{MS}(\pi^l\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in s(\pi)\}) \otimes v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\})$ and $v_{MS}(\pi\alpha) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in max(\pi)\})$. Given definition (D1), $v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\}) = v_{MS}(\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\}) \Rightarrow v_{MS}(\overline{0})$. Note that $v_{MS}(\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\}) = v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in max(\pi)\}) \otimes v_{MS}(\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\})$. Now let $v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in max(\pi)\}) = 0$, $v_{MS}(\&\{\pi_j\alpha \mid \pi_j \in p(\pi)\}) = 1$ and $v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in s(\pi)\}) = 0$, for some model M and some selection function S. Given lemma 4.10, this implies $v_{MS}(\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\}) = 0$ and, because of property (a) of $\Rightarrow$, $v_{MS}(\neg\&\{\pi_j\alpha \mid \pi_j \in max(\pi) \cup p(\pi)\}) = 1$. Given that $v_{MS}(\&\{\pi_i\alpha \mid \pi_i \in s(\pi)\}) = 1$ and property (5) of $\otimes$ in definition 4.1., this implies $v_{MS}(\pi^l\alpha) = 1$. Hence, $v_{MS}(\pi^l\alpha \rightarrow \pi\alpha) = v_{MS}(\pi^l\alpha) \Rightarrow v_{MS}(\pi\alpha) = 1 \Rightarrow 0$. Because of property (b) of $\Rightarrow$, this implies $v_{MS}(\pi^l\alpha \rightarrow \pi\alpha) = 0$ and thus $\not\models_{FL_h} \pi^l\alpha \rightarrow \pi\alpha$ holds. ■

Theorem 4.16. $\not\models_{FL_h} \pi_i\alpha \rightarrow \pi_i^l\alpha$

Proof. Similar to the proof of theorem 4.15. ■

4.7. A philosophical note on completeness

In [4], Lakoff conjectures the following.

> *"It seems to me unlikely that one is going to be able to get complete axiomatizations for fuzzy predicate logic containing such hedges. [...] If my guess is correct, then we will have learned something very deep and important about natural languages and how they differ from artificial languages."* [4], p. 494.

Unfortunately, it is unclear which type of hedges Lakoff has in mind. He explicitly mentions Zadeh's scalar hedges at this point in his paper, yet, given the focus of the paper on non-scalar hedges and the generality of the statement above, he also seems to refer to the latter type of hedges. However, when considering a fuzzy logic like $\mathbf{FL_h}$ developed to deal with the non-scalar hedges *technically*, *strictly speaking* and *loosely speaking*, his conjecture is wrong. $\mathbf{FL_h}$ and **FL** are evidently coextensive. Hence, if **FL** is proven to be (sound and) complete, so is $\mathbf{FL_h}$. A good example is the first-order fuzzy Gödel logic $\mathbf{G}\forall$.[5] This logic is complete with respect to [0,1]-semantics, cf. [3]. Hence, $\mathbf{G}\forall_\mathbf{h}$ is also complete.

---

[5] The proof theory of $\mathbf{G}\forall$ is obtained by adding $A \rightarrow (A \wedge A)$ to the axiom set of $\mathbf{BL}\forall$. Its semantics is obtained by letting the t-norm be the minimum function *min* (which makes the residuation operator correspond to the condition $(x \Rightarrow y) = 1$ if $x \leq y$ and otherwise $(x \Rightarrow y) = y$).

5. Conclusion

Lakoff's fuzzy semantics for the non-scalar hedges *technically*, *strictly speaking* and *loosely speaking* can be integrated in a generic fuzzy logic $\mathbf{FL_h}$. By doing this, several deficits of Lakoff's original proposal become solved. First of all, $\mathbf{FL_h}$ also incorporates interpretations for non-hedged complex predicates. Secondly, the application radius of *strictly speaking*, and *loosely speaking* becomes broadened in such a way that also predicates without a definitional core can be modified by these hedges. $\mathbf{FL_h}$ has a fully specified semantics as well as a straightforward proof theory for reasoning with these hedges. Furthermore, $\mathbf{FL_h}$ confirms intuitive logical relations between the hedged and non-hedged usage of complex predicates.

In order to make all this possible, $\mathbf{FL_h}$ makes use of a special set of selection functions. These functions collect the kind of information a reasoner can retrieve from concepts in his or her memory when interpreting a (non-)hedged complex predicate. Despite this non-standard element, $\mathbf{FL_h}$ evidently remains a conservative modification of its underlying fuzzy logic.

To end, I briefly mention two open research questions in order to illustrate the many challenges still pending. One important challenge is to adjust the formalism of $\mathbf{FL_h}$ in such a way that it can also deal with other hedges (e.g. *typically* or *regularly*). Another challenge is to adjust the system in such a way that also non-unary predicates may be hedged (as this often happens in natural language, e.g. "*Technically speaking, July is the mother of Tom.*"). Obviously, both adjustments first and foremost demand a lot of extra-logical research in line with Lakoff's original analysis.

Acknowledgements. The research for this paper was supported by the Research Fund of the Ghent University. I would like to thank both anonymous referees for their extensive and valuable comments as well as Kristof De Clercq for several inspiring discussions.

References

[1] Barsalou, L., 'Context-independent and context-dependent information in concepts', *Memory & Cognition*, 10: 82–93, 1982.

[2] Costello, F., 'An exemplar model of classification in simple and combined categories', in Gleitman, L.R., Joshi, A.K. (eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Erlbaum, 2000, pp. 95–100.

[3] Hájek, P., *Metamathematics of Fuzzy Logic*, Kluwer Academic Publishers, 1998.

[4] Lakoff, G., 'Hedges: A study in meaning criteria and the logic of fuzzy concepts', *Journal of Philosophical Logic*, 2: 458–508, 1973.

[5] Laurence, S., Margolis, E., 'Concepts and cognitive science', in Margolis, E., Laurence, S. (eds.), *Concepts: Core Readings*, The MIT Press, 1999, pp. 3–81.

[6] Rosch, E., 'On the internal structure of perceptual and semantic categories', *Unpublished paper*, Psychology Dept., University of California, Berkeley, 1971.

[7] Rosch, E., Mervis, C., 'Family resemblances: Studies in the internal structure of categories', *Cognitive Psychology*, 7: 573–605, 1975.

[8] Rosch, E., 'Principles of categorization', in Margolis, E., Laurence, S. (eds.), *Concepts: Core Readings*, The MIT Press, 1999, pp. 189–206.

[9] Wittgenstein, L., *Philosophical Investigations*, Prentice Hall, 1999.

[10] Zadeh, L.A., 'Quantitative fuzzy semantics', *Information Sciences*, 3: 159–167, 1971.

Stephan van der Waart van Gulik
Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2, 9000 Ghent, Belgium
`Stephan.vanderWaartvanGulik@UGent.be`

Piotr Łukowski

The Procedures for Belief Revision

Abstract. The idea of belief revision is strictly connected with such notions as revision and contraction given by two sets of postulates formulated by Alchourrón, Gärdenfors and Makinson in [1], [2], [5]. In the paper expansion, contraction and revision are defined probably in the most orthodox way, i.e. by Tarski's consequence operation (e.g. [11]) and Tarski-like elimination operation (see [7]). In our approach nonmonotonicity appears as a final result of alternate using of two steps of our reasoning: "step forward" and "step backward". Step forward extends the set of our beliefs and it is used when some new belief appears. Step backward reduces the set of our beliefs and it is used when we reject some previously accepted belief. A decision of adding or rejecting of some sentences is arbitrary and depends on our wish only. Thus, this decision cannot be logical and logic cannot justify it. In our approach logic is a tool for faultless and precise realization of extension or reduction of the set of our beliefs. But why some sentences should be added or refused depends on extralogical reasons.

Such understood nonmonotonicity can be considered also on logics other than classical. A reconstruction of a given logic in its deductive-reductive form is here the basis of nonmonotonicity. It means that nonmonotonic reasoning can be formalized on the base of every logic for which its deductive-reductive form can be reconstructed. Firstly our approach is tested on the ground of the classical logic. Next it is confronted with the intuitionism represented by the Heyting-Brouwer logic.

Procedures of contraction and revision are verified by using of the AGM postulates. We limit our considerations to first six conditions for contraction and revision, because of the well known relation between contraction satisfying first four conditions together with the sixth one and revision defined by this contraction and consequence operation. Satisfaction of almost every postulate is for us a good sign that our approach is reasonable. The only exception we make for the fifth postulate and for Harper identity. Analyzing the reasoning alternately extending and contracting the set of beliefs it is difficult to accept the postulate that adding previously rejected sentence we obtain again the same set of beliefs as before the contraction. This problem is deeper considered in the paper. The original AGM settles the relations between contraction and revision with, the well known Levi and Harper identities. In all cases of our approach, revision is defined on the basis of contraction which with respect to revision is prime.

Keywords: belief revision, expansion, contraction, revision; consequence operation, elimination operation; logic of truth, logic of falsehood, nonmonotonic reasoning; classical logic, Heyting-Brouwer logic, intuitionistic logic.

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

1. Introduction

Our basis in defining expansion, contraction and revision procedures will be consequence and elimination operations. Both functions are given in a dual way. Let $\mathcal{L}$ be a given language with universum L. $E, C : 2^L \to 2^L$ for any $X, Y \subseteq L$ satisfy the following conditions:

$$\begin{array}{l} E(X) \subseteq X \subseteq C(X) \\ X \subseteq Y \quad \textit{implies} \quad E(X) \subseteq E(Y) \;\; \textit{and} \;\; C(X) \subseteq C(Y) \\ E(X) \subseteq EE(X) \;\; \textit{and} \;\; CC(X) \subseteq C(X) \end{array}$$

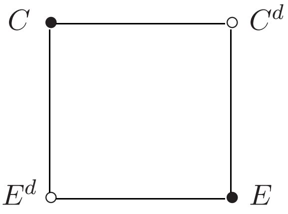
As in [7], also here a notion of logic is extended to the triple $(\mathcal{L}, C, E)$. It is evident that not any arbitrary elimination and consequence operations constitute a logic: both functions have to be somehow *linked*, i.e. they constitute a one logic. In fact there is a unique way in which they may be mutually connected (see [7]): for any $X \subseteq L$,

$$E(X) = L - C^d(L - X),$$

where C^d is a consequence operation dual in Wójcicki's sense to the given C (see [12]). More precisely, a complete reconstruction of the given logic gives two deductive-reductive (d-r) forms of this logic: the first one in the d-r form of the logic of truth $(\mathcal{L}, C, E)$, and the second one, the d-r form of the logic of falsehood $(\mathcal{L}, C^d, E^d)$. Of course, for any $X \subseteq L$,

$$E^d(X) = L - C(L - X).$$

Both triples $(\mathcal{L}, C, E)$ and $(\mathcal{L}, C^d, E^d)$ constitute a logic in the *complete d-r form*. This form of some reconstructed logic can be is represented by the square:



Let H be a class of valuations $v : L \longrightarrow \{0, 1\}$. Then,

$$\begin{array}{llll} \alpha \in C(X) & \textit{iff} & \forall v \in H & (v(X) \subseteq \{1\} \;\; \textit{implies} \;\; v(\alpha) = 1); \\ \alpha \in C^d(X) & \textit{iff} & \forall v \in H & (v(X) \subseteq \{0\} \;\; \textit{implies} \;\; v(\alpha) = 0); \\ \alpha \in E(X) & \textit{iff} & \exists v \in H & (v(L - X) \subseteq \{0\} \;\; \textit{and} \;\; v(\alpha) = 1); \\ \alpha \in E^d(X) & \textit{iff} & \exists v \in H & (v(L - X) \subseteq \{1\} \;\; \textit{and} \;\; v(\alpha) = 0). \end{array}$$

In the notation typical for relations, the complete d-r logic has a form of two triples: $(\mathcal{L}, \vdash, \dashv)$ and $(\mathcal{L}, \vdash^d, \dashv^d)$, where for any $X \subseteq L$:

$$\begin{array}{lll} X \dashv \alpha & \textit{iff} & L - X \vdash^d \alpha; \\ X \dashv^d \alpha & \textit{iff} & L - X \vdash \alpha. \end{array}$$

It is important to notice that the theory of elimination operation is qualitatively and conceptually different from the consequence theory at first sight only. Actually, one is an expansion of the other. Both theories bases on the same set X, usually finite. Indeed, an operation E works on infinite complements $L - X$ of finite sets X. But the set $L - X$ is always given by X. It means that E is *de facto* an operation based on finite sets. The same, answers on all metalogical questions, proof-theoretic and semantic problems like decidability, completeness, finite axiomatizability, which can be connected with the theory of elimination operation are fully "localized" in the set X itself, as in the case of consequence theory. Thus the elimination operation theory expanding the consequence theory sheds an additional light on the problem of acceptance of sentences constituting the finite (usually) set X.

Let us assume that $X \subseteq L$ is a set of accepted by us sentences, i.e. accepted without any logical afterthought.[1] It means that every element of the set X is in our opinion a true sentence. Then, since we have already accepted X, we should also accept $C(X)$ — all truthful consequences of X. It is an example for using of the deductive part $(\mathcal{L}, C)$ of the logic of truth $(\mathcal{L}, C, E)$. From the other hand, since we accepted only X, thus we did not accept $L - X$. This fact has also some consequences. One can say that $L - X$ is in some sense the set of rejected sentences. Since we had already rejected $L - X$, we should also reject $C^d(L - X)$ — all "falseful" consequences of X. Here, we use a deductive part $(\mathcal{L}, C^d)$ of the logic of falsehood $(\mathcal{L}, C^d, E^d)$. But what about the set of our beliefs X, i.e. the set of accepted by us sentences? Obviously, it should be reduced by all sentences from $C^d(L - X)$. It means that our new set of beliefs should be limited to the set: $L - C^d(L - X) = E(X)$. In such a way we use the reductive part $(\mathcal{L}, E)$ of the logic of truth $(\mathcal{L}, C, E)$.[2]

---

[1] I.e. without verifying of any logical consequences of accepted premises. Firstly sentences of X are accepted by us, and then operations C and E are applied.

[2] It seems that the comment above would be more convincing, when the set $C(X)$ would appear instead of X in the definition equality of E. Indeed, if one accepts a set X of sentences, then one has to accept $C(X)$, and so the expression $L - C^d(L - X)$ should be replaced by $L - C^d(L - C(X))$. However, the comment above explains the

Now, let us assume that $X \subseteq L$ is a set of sentences which are false for us. Then, all sentences from $C^d(X)$ should be also false for us, by the deductive part $(\mathcal{L}, C^d)$ of the logic of falsehood $(\mathcal{L}, C^d, E^d)$. However, simultaneously the set of false for us sentences should be limited to the set $L - C(L - X) = E^d(X)$, by the reductive part $(\mathcal{L}, E^d)$ of the logic of falsehood $(\mathcal{L}, C^d, E^d)$.

It means that a logic given in the complete form can be applied to the set of true sentences as well as to the set of false sentences. In the paper we will consider such an extension of the classical logic which can be applied to the set of sentences among which some are true and some other false.

A logic in the d-r form allows us to move freely forward and backwards in our reasoning, i.e. either to extend or to reduce our set of beliefs. Contrary to AGM the procedures considered in the sequel may work in other logics, not only in the classical one. Thanks to elimination and consequence operations all these reasoning procedures can be defined on any logic given by structural finite consequence operation and, linked to it, elimination operation, or equivalently with cofinite structural elimination operation and linked to it consequence operation. Therefore, we shall deal only with structural, finitary consequence operations and with structural, cofinitary elimination operations. Now, let us recall some useful notions and facts given in [14] and [7].

A consequence operation C is *structural*, if for any endomorphism e of the language $\mathcal{L}$ and for any $X \subseteq L$:

$$eC(X) \subseteq C(eX).$$

An elimination operation E is *structural*, if for any endomorphism e of the language $\mathcal{L}$ and for any $X \subseteq L$:

$$e(L - E(X)) \subseteq L - E(L - e(L - X)).$$

A consequence operation C is *finitary*, if for any $X \subseteq L$:

$$C(X) = \bigcup\{C(Y) : Y \subseteq X \text{ and } Y \text{ is a finite set}\}.$$

An elimination operation E is *cofinitary*, if for any $X \subseteq L$:

$$E(X) = \bigcap\{E(Y) : X \subseteq Y \text{ and } Y \text{ is a cofinite set}\}.$$

---

nature of E, and not express a philosophical motivation for it. The motivation is given by presented above semantic characterisation of four operations: C, C^d, E, E^d. So, an equality $L - C^d(L - X) = E(X)$ directly comes from the semantical relations between C, E and C^d. Moreover, the operation E is defined for every X and not for $C(X)$ only.

Sets $X = C(X)$ and $Y = E(Y)$ are *C-theory* and *E-theory* further designated by T and F, respectively. Trivial theories $T = L$ and $F = \emptyset$ are called *inconsistent* and *insufficient*, respectively.

A C-theory T is *maximal relatively to* α, if

1. $\alpha \notin T$ and
2. $\alpha \in C(T + \beta)$ for any $\beta \notin T$.

A C-theory maximal relatively to some formula is relatively maximal.

An E-theory F is *minimal relatively to* α, if

1. $\alpha \in F$ and
2. $\alpha \notin E(F - \beta)$ for any $\beta \in F$.

An E-theory minimal relatively to some formula is relatively minimal.

LINDENBAUM LEMMA. Let C be a finitary consequence operation on $\mathcal{L}$. For any consistent C-theory T and for any $\alpha \notin T$ there exists a C-theory T_0 maximal relatively to α such that $T \subseteq T_0$.

DUAL-TO-LINDENBAUM LEMMA. Let E be a cofinitary elimination operation on $\mathcal{L}$. For any sufficient E-theory F and for any $\alpha \in F$ there exists an E-theory F_0 minimal relatively to α such that $F_0 \subseteq F$.

The letter of the following notions will be useful for the further considerations:

- a maximal relatively to α C-theory T *for the set* X;

  and

- a minimal relatively to α E-theory F *for the set* X.

Assume that $\alpha \notin Y$. In order to construct a maximal relatively to α C-theory for the set X including consistent set Y, it is sufficient to modify a proof of Lindenbaum lemma, presented with all details in [13]. This construction begins from the ordering, in the form of the sequence, of all formulas of the set $L - Y$. Building a maximal relatively to α C-theory for the set X, this sequence is formed in such a way that every formula from $X - Y$ proceeds all formulas from $(L - Y) - X$. It means that all formulas from the set X will be always taken into account as first in the procedure of extending the set Y.

An analogous, small correction of the proof of the dual-to-Lindenbaum lemma results in the construction of minimal relatively E-theories for a given set X. Let $\alpha \in Y$ and $X^{\neg} = \{\neg\beta : \beta \in X\}$. The construction of the minimal relatively to α E-theory for the set X included in the sufficient set Y begins from the ordering, in the form of the sequence, of all formulas of the set Y. This sequence is formed in such a way that every formula from $Y \cap X^{\neg}$ precedes all formulas from $Y \cap (L - X)$. It means that all formulas from the set $Y \cap X^{\neg}$ will be always taken into account as first in procedure of removing from the set Y. Since, in the case of the classical logic, simultaneous removing of any formula together with its negation from some set makes this set insufficient, removing $\neg\beta$ from Y prevents the removing β from Y.

1.1. Dual to Łoś-Suszko Theorem

Syntactically, each elimination operation is defined by the so called *E-axioms*, i.e. formulas which cannot be accepted even in the whole language. So, as *C-axioms* are expressions of the shape

$$\emptyset \vdash \alpha,$$

i.e. α is a tautology of $\vdash$, E-axioms have a form of

$$L \dashv \alpha,$$

i.e. α is a counter-tautology of $\vdash$ or equivalently a tautology of $\vdash^d$. While consequence operation enables us to express a set of all tautologies, $C(\emptyset)$, elimination operation defines a set of non-counter-tautologies (i.e. all formulas which are not counter-tautologies), $E(L)$. Moreover, if C and E are linked, then a sentence $\alpha \in C(\beta)$ is equivalent to $\beta \notin E(L - \alpha)$. The case of *E-rule* is analogous. A deduction rule (*D-rule*) is of the form $\emptyset + \{\alpha_1, \ldots, \alpha_k\} \vdash \beta$, shortly $\{\alpha_1, \ldots, \alpha_k\} \vdash \beta$. The *E-rule*, the rule of elimination should be of shape $L - \{\alpha_1, \ldots, \alpha_k\} \dashv \beta$. Of course, from the syntactic point of view a definition of D-rule and E-rule is the same: it is an ordered pair of formula set and single formula. Thus, Modus Ponens deductive in character has its reductive counterpart in the form of the rule which says which formula should be removed from a given set, if some other formulas are removed from it.

In the consequence theory there is the notion of *"proof"* which plays the key role. The elimination operation is closely connected with a dual notion of *"disproof"*.

Let $\mathcal{R}$ be a set of rules of elimination operation E. A formula α is called disprovable from X by means of rules from $\mathcal{R}$, if and only if there exists in $L{-}X$ a finite sequence of formulas $\alpha_1, \ldots, \alpha_k$, such that

- $\alpha = \alpha_k$ *and*
- *for any* $i \in \{1, \ldots, k\}, \alpha_i \in L - X$ *or for some* $Y \subseteq \{\alpha_1, \ldots, \alpha_{k-1}\}$, $L - Y \dashv \alpha_i$ *is an instance of some rule from* $\mathcal{R}$.

Every sequence $\alpha_1, \ldots, \alpha_k$ *satisfying above conditions is called a disproof of* α *from* X *by means of* $\mathcal{R}$.

A formula α *is called confirmed for* X *by means of* $\mathcal{R}$ *if there exists no disproof of* α *from* X *by means of* $\mathcal{R}$.

For more details see also [7].

Directly from the definition above it follows that rules from $\mathcal{R}$ remove some of elements from X in such a way that the set $L - X$ is extended to some set $(L - X)_{\mathcal{R}} = L - E(X)$. Thus, the set $(L - X)_{\mathcal{R}}$ is closed under every rule from $\mathcal{R}$[3].

LEMMA. *Let* $E : 2^L \rightarrow 2^L$ *be any elimination operation, and* $\mathcal{R}(E)$ *be a set of all rules of* E. *Then, for any* $X \subseteq L$ *and* $\alpha \in L$:

$$\alpha \in E(X) \quad \text{implies} \quad \alpha \in E_{\mathcal{R}(E)}(X).$$

PROOF. Assume that $\alpha \notin E_{\mathcal{R}(E)}(X)$. Then, there exists a disproof $\alpha_1, \ldots, \alpha_k$ of α from X by means of $\mathcal{R}(E)$. Now, it will be shown that for any $i \in \{1, \ldots, k\}$, $\alpha_i \notin E(X)$.

For $i = 1$, $\alpha_1 \notin X$ or $L \dashv \alpha_1$ is an instance of some rule from $\mathcal{R}(E)$. In the first case, obviously, $\alpha_1 \notin E(X)$. Assume that $L \dashv \alpha_1$ is a rule from $\mathcal{R}(E)$. Then, the set $(L - X)_{\mathcal{R}}$ is closed under rule $L \dashv \alpha_1$, and so $\alpha_1 \notin E(X)$.

Now assume that $\alpha_1 \notin E(X), \ldots, \alpha_{i-1} \notin E(X)$, for $i \in \{2, \ldots, k\}$. By the above definition of disproof, $\alpha_i \notin X$ or for some $Y \subseteq \{\alpha_1, \ldots, \alpha_{i-1}\}$, $L - Y \dashv \alpha_i$ is a rule from $\mathcal{R}(E)$. Obviously, if $\alpha_i \notin X$, then $\alpha \notin E(X)$. Let for some $Y \subseteq \{\alpha_1, \ldots, \alpha_{i-1}\}$, $L - Y \dashv \alpha_i$ is a rule from $\mathcal{R}(E)$. Then, $(L - X)_{\mathcal{R}}$ is closed under rule $L - Y \dashv \alpha_i$, and since $Y \subseteq (L - X)_{\mathcal{R}}$ $(\{\alpha_1, \ldots, \alpha_{i-1}\} \subseteq (L - X)_{\mathcal{R}})$, $\alpha_i \in (L - X)_{\mathcal{R}}$. Thus, $\alpha_i \notin E(X)$. It proves that for any $i \in \{2, \ldots, k\}$, if $\alpha_1 \notin E(X), \ldots, \alpha_{i-1} \notin E(X)$, then $\alpha_i \notin E(X)$.

Since $\alpha_k = \alpha$, $\alpha \notin E(X)$. ■

DUAL TO ŁOŚ-SUSZKO THEOREM. For every cofinitary elimination operation $E : 2^L \rightarrow 2^L$ there exists $\mathcal{R}$, a set of E-rules such that

$$E = E_{\mathcal{R}}.$$

---

[3] In our approach, the set $(L - X)_{\mathcal{R}}$ is a C^d-theory, i.e. a theory of the logic dual to C in Wójcicki's sense.

Proof. Assume that $\alpha \in E(X)$. Then, by the Lemma above, $\alpha \in E_{\mathcal{R}(E)}(X)$. Now assume that $\alpha \notin E(X)$. Since E is a cofinitary elimination operation, $\alpha \notin E(L - Y)$, for some finite $Y \subseteq L - X$. Then, $\{\alpha_1, \ldots, \alpha_k\} \dashv \alpha$ is a rule from $\mathcal{R}(E)$, where $\{\alpha_1, \ldots, \alpha_k\} = Y$. Of course, $\alpha_1, \ldots, \alpha_k, \alpha$ is a disproof of α from X by means of $\mathcal{R}(E)$. Thus, $\alpha \notin E_{\mathcal{R}(E)}(X)$. ■

2. Nonmonotonicity on classical base

Classical logic will serve here as the first illustration. Albeit many other logics could be also considered, the main aim of this section is to check how our approach realizes AGM postulates (Alchourrón-Gärdenfors-Makinson postulates) for contraction and revision on typical for AGM ground of the classical propositional logic.

The classical propositional logic of truth in the deductive-reductive form is a triple:

$$(\mathcal{L}_{CL}, C_{CL}, E_{CL}),$$

where

$$\mathcal{L}_{CL} = (L_{CL}, \neg, \wedge, \vee, \rightarrow).$$

C_{CL} is the very well known deductive part and the reductive part E_{CL} of the classical propositional logic of truth is given by the following axiom set:

1_E $L_{CL} \dashv_{CL} \neg(\neg(\alpha \rightarrow \beta) \rightarrow \alpha)$
2_E $L_{CL} \dashv_{CL} \neg(\neg(\neg(\gamma \rightarrow \alpha) \rightarrow \neg(\beta \rightarrow \alpha)) \rightarrow \neg(\neg(\gamma \rightarrow \beta) \rightarrow \alpha))$
3_E $L_{CL} \dashv_{CL} \neg((\alpha \wedge \beta) \rightarrow \alpha)$
4_E $L_{CL} \dashv_{CL} \neg((\alpha \wedge \beta) \rightarrow \beta)$
5_E $L_{CL} \dashv_{CL} \neg(\neg(\neg(\gamma \rightarrow (\alpha \wedge \beta)) \rightarrow \neg(\gamma \rightarrow \beta)) \rightarrow \neg(\gamma \rightarrow \alpha))$
6_E $L_{CL} \dashv_{CL} \neg(\alpha \rightarrow (\alpha \vee \beta))$
7_E $L_{CL} \dashv_{CL} \neg(\beta \rightarrow (\alpha \vee \beta))$
8_E $L_{CL} \dashv_{CL} \neg(\neg(\neg((\alpha \vee \beta) \rightarrow \gamma) \rightarrow \neg(\beta \rightarrow \gamma)) \rightarrow \neg(\alpha \rightarrow \gamma))$
9_E $L_{CL} \dashv_{CL} \neg(\neg(\beta \rightarrow \neg\alpha) \rightarrow \neg(\alpha \rightarrow \neg\beta))$
10_E $L_{CL} \dashv_{CL} \neg(\beta \rightarrow (\alpha \rightarrow \alpha))$
11_E $L_{CL} \dashv_{CL} \alpha \wedge \neg\alpha$
MT_E $L_{CL} - \{\beta, \neg(\alpha \rightarrow \beta)\} \dashv_{CL} \alpha$

A meaning of MT_E is the following: removing two sentences $\neg(\alpha \rightarrow \beta)$ and β from X removes α from X.

In order to obtain a syntax for classical logic of falsehood ${C^d_{CL}}^4$, a logic dual in Wójcicki's sense to C_{CL} (see [12]), it suffices to replace "$L_{CL} \dashv_{CL}$" by "$\emptyset \vdash_{CL^d}$" in 1_E-11_E, and MT_E by MT_C: $\{\beta, \neg(\alpha \rightarrow \beta)\} \vdash_{CL^d} \alpha$. It means that one can formulate an equality expressing the relation between E_{CL} and C^d_{CL}. For $X \subseteq L$:

$$E_{CL}(X) = L - C^d_{CL}(L - X).$$

Further, i.e. throughout the "classical" section of the paper only, for simplicity, let us use "C", "E" and "L" instead of "C_{CL}", "E_{CL}" and "L_{CL}", respectively.

It is useful to formulate narrower notions of the relativeness of theories, which shall be of special use in the sequel.

A C-theory T will be called *maximal relatively to the given formula set* $\{\alpha_1, \ldots, \alpha_k\}$, if T is maximal relatively to $\alpha_1 \vee \ldots \vee \alpha_k$.

An E-theory F will be called *minimal relatively to the given formula set* $\{\alpha_1, \ldots, \alpha_k\}$, if F is minimal relatively to $\alpha_1 \wedge \ldots \wedge \alpha_k$.

Moreover,

1. a C-theory T is *C-prime*, if for any α, β: $\alpha \vee \beta \in T$ iff ($\alpha \in T$ or $\beta \in T$); and
2. an E-theory F is *E-prime*, if for any α, β: $\alpha \wedge \beta \in F$ iff ($\alpha \in F$ and $\beta \in F$).

Obviously, every relatively maximal C-theory is C-prime and every relatively minimal E-theory is E-prime.

2.1. Expansion

The simplest case possible is the procedure which coincides with the consequence. Treating the consequence operation as extending the belief set is not exceptional but naturally adopted by AGM (see [1], [2] and [5]). Belief set is closed under logical consequences set of sentences, further called C-set of beliefs. Thus, in our notation we obtain

EXPANSION:

$$T \oplus \varphi = C(T + \varphi) \tag{I}$$

In the form of expansion we code a reasoning informing us which sentences we have to accept, when we have already accepted some other sentences.

---

[4] Of course, $C^d_{CL} = C_{CL^d}$.

2.2. Contraction

On the basis of deductive reasoning, contraction is a typical step backward. And, if we decide to reject given sentence from the C-set of our beliefs, we employ a contraction, i.e., such a procedure, which leads from one C-set of beliefs with a given sentence to a new, obviously smaller C-set of beliefs being a C-theory but already without this sentence.

Removing some sentence from C-theory results in a C-theory. It means that with this sentence we have to remove other sentences from which it follows. It is easy to see that the elimination operation allows for removing exactly those sentences. However, usually more than one solution is possible. Let us assume that we intend to remove a sentence φ from the C-set T. Obviously, in $E(L-\varphi)$ (also in $E(T-\varphi)$) there are no sentences from which φ follows. Unfortunately, neither $E(L-\varphi)$ nor $E(T-\varphi)$ is a C-theory. Let us suppose that $\varphi \in C(\alpha_1, \alpha_2, \alpha_3)$. Of course, $\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \notin E(L-\varphi)$. However, the condition "$\alpha \wedge \beta \notin X$ iff $\alpha \notin X$ or $\beta \notin X$" holds for E-prime E-theories only. It means that $E(L-\varphi)$ as a non-relatively minimal E-theory can contain all sentences: α_1, α_2, α_3, and then $\varphi \in CE(L-\varphi)$.

An assumption $\varphi \in C(\alpha_1, \alpha_2, \alpha_3)$ means that it suffices to remove only one of these sentences to remove φ — as well. It is not difficult to imagine a situation in which we do not wish to lose, for example, α_2. Then, we may fix it by replacing $E(L-\varphi)$ with $E_{\alpha_2}(L-\varphi)$ which is an E-theory minimal relatively to α_2. For this reason the notion expressed by $K\dot{-}A$[5] and appearing in AGM postulates is not considered here: speaking precisely $K\dot{-}A$ is a name of a class of contractions and not of only one concrete contraction. Of course, so called informational economy gives some explanation how $K\dot{-}A$ can be only one contraction procedure. But it seems that such approach makes problem too poor. Indeed, let us consider a simple example: In our world a sentence *"when it rains, the streets are wet"* ($\alpha{\to}\beta$) is true, because great majority of streets are roofless. Hence when it rains we know that streets are wet. Now suppose, that we are in bed and hear some delicate sound. We conjecture that *"it rains"* (α), so we think that *"our street is wet"* (β). Later after a short nap, we wake up and look through the window and realize that the street is dry. In result from the C-set of our beliefs have to reject a sentence β, i.e. *"my street is wet"*. Of course, this sentence was a conclusion of some other beliefs we had, namely $\alpha \to \beta$ and α. Probably, nobody supposes that the first of them should be removed from the C-set of beliefs. In contrary, it would be natural to leave this sentence,

---

[5] $K\dot{-}A$ is an original for AGM notation, where K is a set of beliefs (of course, C-set) and A — a sentence rejected from K (cf [5]).

which means that the only reasonable solution is to take $E_{\alpha\to\beta}(L-\beta)$, an E-theory minimal relatively to $\alpha \to \beta$. Obviously, α does not belong to this theory.

Evidently, it is possible to find an opposite example where the only reasonable solution would be to keep α and to remove $\alpha \to \beta$. One way or another, it is absolutely natural and necessary to take into account the possibility of keeping the desired sentences.

Unfortunately, a name "E-theory minimal relatively to some sentence set" also designates a class of contractions instead of only one contraction procedure. Moreover, applying this notion one can cancel all sentences which are not connected with removed sentence. Indeed, assume that our C-set T contains φ, following from $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in T$, and ψ which has no links with the mentioned sentences. Building an E-theory minimal relatively, for example, to α_1, among others we have to verify two sentences ψ and $\neg(\varphi \to \psi)$. In this situation, everything depends on the order which sentence will be checked as the first one. If at first we verify $\neg(\varphi \to \psi)$, this sentence will be removed from the E-theory minimal relatively to α_1, while ψ will belong to this theory. An opposite situation occurs, when we check sentence ψ as first. Then, this sentence will be removed from our minimal relatively E-theory, and $\neg(\varphi \to \psi)$ will belong to this theory. Obviously, contraction should work only on these sentences which are connected with the removed sentence. All other sentences are preserved. Thus, in order to avoid such an undesired result like removing additional sentences, for the construction of contraction procedure, instead of the E-theory minimal relatively to some sentence, we should use a minimal relatively to some sentence E-theory for some sentence set. It is visible, that our minimal relatively to α_1 E-theory for the set T contains ψ and of course, $\neg(\varphi \to \psi)$ does not belong to such theory.

A relatively minimal E-theory for some set satisfies a condition which probably can be called … a postulate of informational economy. Since relatively minimal E-theory is "for the set T", all sentences which do not have to be removed from the belief set will be saved. Returning to our example, because of $\varphi \in C(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ (assume that φ does not follow from any proper subset of this four-sentences set), a minimal relatively to α_1 E-theory for T has to contain two other sentences from $\{\alpha_2, \alpha_3, \alpha_4\}$. Unfortunately it is not predeterminated which sentences are to be saved. It means that also a "relatively minimal E-theory for some set" names a class of possible contractions, and not only one procedure. Of course a scope of such possible contractions is strongly limited — in the case of our example there are only three different minimal relatively to α_1 E-theories for T.

Now, it is clear that the intersection of all minimal relatively to some sentences E-theories for a given set is the one and only one set of sentences. In our example, an intersection of all minimal relatively to α_1 E-theories for T contains among others as many sentences from T as possible, i.e. all these sentences belonging to T from which φ does not follow. Simultaneously, in the case of the set $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ only one sentence α_1 belongs to the intersection in question. Sentences α_2, α_3 and α_4 are removed from the intersection. One can say that the procedure based on the intersection of relatively minimal E-theories cancels all sentences from which a removed sentence follows, with the only exception of chosen sentences to which E-theories are relatively minimal. Let us extend our example and suppose that the sentence φ follows not only from $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ but also from the set $\{\beta_1, \beta_2, \beta_3\}$. Obviously, no sentence from the set of betas belongs to the intersection of all minimal relatively to α_1 E-theories for T.

Let us recall that every relatively minimal E-theory is prime and so is a C-theory. Thus, an intersection of any family of relatively minimal E-theories is a C-theory.

Finally, let us formulate

CONTRACTION:

$$(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T \cap \bigcap E^T_{\alpha_1,\ldots,\alpha_k}(L - \varphi) \tag{II}$$

with $E^T_{\alpha_1,\ldots,\alpha_k}(L - \varphi)$ an E-theory for T, minimal relatively to the set $\{\alpha_1, \ldots, \alpha_k\} \subseteq T$. According to these definition our C-set of beliefs T is decreased by φ but the desired sentences $\alpha_1, \ldots, \alpha_k$ are still our beliefs. Obviously, the final belief C-set is a C-theory (not necessarily prime). Moreover, if φ follows from the set $\{\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_n\}$ and φ does not follow from any proper its subset, $\{\alpha_{k+1}, \ldots, \alpha_n\} \cap (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = \emptyset$.

Now let us see how the so called AGM postulates are satisfied by this procedure. In our notation, first six postulates for contraction are of the following form:

Con_1 $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$ *is a C-theory*
Con_2 $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} \subseteq T$
Con_3 *if* $\varphi \notin T$, *then* $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T$
Con_4 *if* $\varphi \notin C(\emptyset)$, *then* $\varphi \notin (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$
Con_5 *if* $\varphi \in T$, *then* $T \subseteq (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} \oplus \varphi$
Con_6 *if* $C(\varphi) = C(\psi)$, *then* $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = (T \ominus \psi)_{\alpha_1,\ldots,\alpha_k}$

Of course, in all conditions T is a C-theory.

A satisfaction of Con_1 and Con_2 is evident and directly follows from the equality (2). Let us check the rest of the above postulates.

Assume that $\varphi \notin T$. Suppose, moreover, that $\beta \in T$ and $\beta \notin E(L - \varphi)$, for some $\beta \in L$. By reductive theorem for classical logic (see [7]) $\neg(\beta \to \varphi) \notin E(L)$. So, $\neg(\beta \to \varphi)$ is a countertautology, and the same $\beta \to \varphi \in T$ as a tautology. It means that by assumption $\varphi \in T$ — a contradiction. Thus, $T \subseteq E(L-\varphi)$. Then, directly from the construction of the relatively minimal E-theory for the set T, it follows that every relatively minimal E-theory for the set T is disjoint with $T^{\neg}$. It means that T is included in every relatively minimal E-theory for T, and so Con_3 is satisfied.

Obviously, $\varphi \notin E(L - \varphi)$, for any $\varphi \in L$. Of course, if $\varphi \in C(\emptyset)$, then $E(L - \varphi) = \emptyset$, and so $\varphi \notin E(L - \varphi)$. Thus, the condition Con_4 holds. Our verification shows that the fourth AGM postulate can be here replaced by its stronger version:

Con_4^* $\quad \varphi \notin (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$

An assumption that $\varphi \leftrightarrow \psi$ is a classical tautology implies that $\neg(\varphi \to \psi)$ as well as $\neg(\psi \to \varphi)$ are countertautology. Thus, $C(\varphi) = C(\psi)$ implies $E(L - \varphi) = E(L - \psi)$[6]. It means that Con_6 is satisfied (see [7]).

Assume that $\varphi, \psi \in T$. Let moreover, $\varphi \in C(\psi)$ and $\psi \notin C(\varphi)$. Then, by deductive theorem $\psi \to \varphi \in C(\emptyset)$. Thus, $\neg(\psi \to \varphi) \notin E(L)$, and by reductive theorem, $\psi \notin E(L - \varphi)$. It means that $\psi \notin (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$. Moreover, since $\psi \notin C(\varphi)$, thus $\psi \notin (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} \oplus \varphi$. It falsifies the condition Con_5.

From the intuitive point of view, it is difficult to accept the fifth AGM condition. In fact, adding to the C-set of beliefs some sentence removed from it earlier by contraction, not always brings back the sentences from which it follows although they formerly belonged there. It is sufficient to recall our example with rain and a wet street. Sentences α and $\alpha \to \beta$ will be again in T only if contraction removing β from T will remove $\alpha \to \beta$ as well. But definition of contraction limited to such cases although correct accordingly to informational economy idea is rather unreasonable even in such simple cases like our example. A contraction giving us the full scope of possible solutions fails to obey a postulate in AGM form. Probably, the following version of Con_5 seems to be more intuitive:

Con_5^* if $\varphi \notin T$, *then* $T \subset (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} \oplus \varphi$;
if $\varphi \in T$, *then* $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} \oplus \varphi \subset T$.

---

[6] An opposite implication also holds.

Checking the first sentence of Con_5^*, suppose that $\varphi \notin T$. Then by Con_3, $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T$, and so $T \subseteq C((T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} + \varphi)$. Given our assumption, the inclusion in question is proper. In the light of the falsification above and Con_2, the condition for $\varphi \in T$ is evident.

An examination of the first six AGM postulates for contraction shows that, they are satisfied with only one exception of the fifth one. The fourth condition is fulfilled even in a more general case.

2.3. Revision

In the already quoted here publications dealing with belief revision it is underlined that contraction and revision are independent as far as their introduction is concerned but it is possible to find between them some mutual relations formalized by the so called Levi and Harper identities. Levi identity provides a method of defining revision by contraction, while the second one defines contraction by revision. In our case, contraction is a prime function with respect to the revision. It means that contraction is the first step in revision. Thus, if we plan to add a sentence contradictory to some other our belief, we have to resign from the old belief at first. Only then we can consistently accept the new sentence. Thus, contraction is a proper part of revision and obviously, it is impossible to make revision without contraction. This idea is expressed by Levi identity, and the next definitions follow it.

REVISION:

$$(T \odot \varphi)_{\alpha_1,\ldots,\alpha_k} = (T \ominus \neg\varphi)_{\alpha_1,\ldots,\alpha_k} \oplus \varphi \tag{III}$$

As in the case of contraction also our notion of revision takes into account a fact that for a given belief C-set T and sentence φ there exists not only one revision. Thus, accordingly to the above definition, a revision of C-set T by sentence φ depends on sentences $\alpha_1, \ldots, \alpha_k$ which are still hold in the C-set.

It is shown in [1], [2] and [5] that for $\ominus$ satisfying Con_1-Con_4, Con_6 and $\oplus$ being a consequence operation the function $\odot$ defined by (5) satisfies the following six conditions being the well known AGM postulates for revision:

Rev_1 $(T \odot \varphi)_{\alpha_1,\ldots,\alpha_k}$ *is a C-theory*
Rev_2 $\varphi \in (T \odot \varphi)_{\alpha_1,\ldots,\alpha_k}$
Rev_3 $(T \odot \varphi)_{\alpha_1,\ldots,\alpha_k} \subseteq T \oplus \varphi$
Rev_4 *if* $\neg\varphi \notin T$, *then* $T \oplus \varphi \subseteq (T \odot \varphi)_{\alpha_1,\ldots,\alpha_k}$
Rev_5 *if* $\neg\varphi \in C(\emptyset)$, *then* $(T \odot \varphi)_{\alpha_1,\ldots,\alpha_k} = L$
Rev_6 *if* $C(\varphi) = C(\psi)$, *then* $(T \odot \varphi)_{\alpha_1,\ldots,\alpha_k} = (T \odot \psi)_{\alpha_1,\ldots,\alpha_k}$

The already mentioned Harper identity enables us to reconstruct AGM contraction for a given AGM revision. In our notation its form should be the following

$(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T \cap (T \odot \neg\varphi)_{\alpha_1,\ldots,\alpha_k}$ i.e.
$(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T \cap C((T \ominus \neg\neg\varphi)_{\alpha_1,\ldots,\alpha_k} + \neg\varphi)$.

Contrary to Levi identity this one does not work in general but only for $\varphi \notin T$, due to Con_3 and Con_6, $(T \ominus \neg\neg\varphi)_{\alpha_1,\ldots,\alpha_k} = (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = T$, and so Harper identity is satisfied. Assume, however, that φ belongs to the consistent C-set T. Let, moreover, $\psi = \gamma \wedge (\gamma \to \varphi)$ as well as $\delta = \neg\varphi \to (\gamma \wedge (\gamma \to \varphi))$ also are in T. Obviously, $\neg\varphi \notin T$, $\psi \notin (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$ but $\delta \in (T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k} = (T \ominus \neg\neg\varphi)_{\alpha_1,\ldots,\alpha_k}$. Evidently $C((T \ominus \neg\neg\varphi)_{\alpha_1,\ldots,\alpha_k} + \neg\varphi) = L$ but neither ψ nor φ belong to $(T \ominus \varphi)_{\alpha_1,\ldots,\alpha_k}$. Thus, Harper identity is falsified.

3. Nonmonotonicity on intuitionistic base

The reconstruction of the reductive part for the intuitionistic logic is not easy. The problem is strictly connected with the axiomatization of the logic dual in Wójcicki's sense to the well known intuitionistic logic. As it was shown (see [3], [4], [10]), there is no finite axiomatization of this logic. Since,

$$E_{INT}(X) = L - C^d_{INT}(L - X),$$

it means that there is also no finite axiomatization of the reductive part of the intuitionistic logic in the deductive-reductive form. By the same, there is a real problem with successful defining of intuitionistic contraction and revision. Fortunately, it is possible to extend intuitionistic logic to the Heyting-Brouwer logic, which can be a base for easy defining of both belief revision functions.

Heyting-Brouwer logic (*HB*) as well as HB^d, the logic dual to *HB* was define in seventies by C. Rauszer (see [8], [9]). Rauszer constructs both logics on the classical language extended by two connectives: co-implication ($\leftharpoondown$) and weak negation ($\sim$);

$$\mathcal{L}_{HB} = (L_{HB}, \neg, \sim, \wedge, \vee, \to, \leftharpoondown)$$

The syntax of *HB* consists of axiom set for the intuitionistic logic and the following axioms defining both new connectives:

$\emptyset \vdash_{HB} \alpha \to (\beta \vee (\alpha \leftharpoondown \beta))$
$\emptyset \vdash_{HB} (\alpha \leftharpoondown \beta) \to \ \sim(\alpha \to \beta)$

$\emptyset \vdash_{HB} ((\alpha \leftharpoonup \beta) \leftharpoonup \gamma) \to (\alpha \leftharpoonup (\beta \vee \gamma))$
$\emptyset \vdash_{HB} \neg(\alpha \leftharpoonup \beta) \to (\alpha \to \beta)$
$\emptyset \vdash_{HB} (\alpha \to (\beta \leftharpoonup \beta)) \to \neg\alpha$
$\emptyset \vdash_{HB} \neg\alpha \to (\alpha \to (\beta \leftharpoonup \beta))$
$\emptyset \vdash_{HB} ((\beta \to \beta) \leftharpoonup \alpha) \to \ {\sim}\alpha$
$\emptyset \vdash_{HB} {\sim}\alpha \to ((\beta \to \beta) \leftharpoonup \alpha)$

for $\alpha, \beta, \gamma \in L_{HB}$; and two rules of inference: Modus Ponens and $\alpha \vdash_{HB} \neg{\sim}\alpha$.

In [9], there is defined HB^{d} by the following axioms:

$\emptyset \vdash_{HB^d} ((\beta \leftharpoonup \gamma) \leftharpoonup (\alpha \leftharpoonup \gamma)) \leftharpoonup ((\beta \leftharpoonup \alpha) \leftharpoonup \gamma)$
$\emptyset \vdash_{HB^d} ((\gamma \leftharpoonup \alpha) \leftharpoonup (\gamma \leftharpoonup \beta)) \leftharpoonup (\beta \leftharpoonup \alpha)$
$\emptyset \vdash_{HB^d} (\alpha \wedge \beta) \leftharpoonup \alpha$
$\emptyset \vdash_{HB^d} (\alpha \wedge \beta) \leftharpoonup \beta$
$\emptyset \vdash_{HB^d} ((\gamma \leftharpoonup (\alpha \wedge \beta)) \leftharpoonup (\gamma \leftharpoonup \beta)) \leftharpoonup (\gamma \leftharpoonup \alpha)$
$\emptyset \vdash_{HB^d} \alpha \leftharpoonup (\alpha \vee \beta)$
$\emptyset \vdash_{HB^d} \beta \leftharpoonup (\alpha \vee \beta)$
$\emptyset \vdash_{HB^d} (((\alpha \vee \beta) \leftharpoonup \gamma) \leftharpoonup (\beta \leftharpoonup \gamma)) \leftharpoonup (\alpha \leftharpoonup \gamma)$
$\emptyset \vdash_{HB^d} ((\gamma \leftharpoonup (\alpha \vee \beta)) \leftharpoonup ((\gamma \leftharpoonup \beta) \leftharpoonup \alpha)$
$\emptyset \vdash_{HB^d} ((\gamma \leftharpoonup \beta) \leftharpoonup \alpha) \leftharpoonup (\gamma \leftharpoonup (\alpha \vee \beta))$
$\emptyset \vdash_{HB^d} ((\beta \to \alpha) \wedge \beta) \leftharpoonup \alpha$
$\emptyset \vdash_{HB^d} ({\sim}\alpha \leftharpoonup {\sim}\beta) \leftharpoonup (\beta \leftharpoonup \alpha)$
$\emptyset \vdash_{HB^d} ((\beta \wedge \gamma) \to \alpha) \leftharpoonup (\gamma \to (\beta \to \alpha))$
$\emptyset \vdash_{HB^d} \neg(\beta \leftharpoonup \alpha) \leftharpoonup (\beta \to \alpha)$
$\emptyset \vdash_{HB^d} (\alpha \leftharpoonup \beta) \leftharpoonup \ {\sim}(\alpha \to \beta)$
$\emptyset \vdash_{HB^d} {\sim}\alpha \leftharpoonup ((\beta \to \beta) \leftharpoonup \alpha)$
$\emptyset \vdash_{HB^d} ((\beta \to \beta) \leftharpoonup \alpha) \leftharpoonup \ {\sim}\alpha$
$\emptyset \vdash_{HB^d} \neg\alpha \leftharpoonup (\alpha \to (\beta \leftharpoonup \beta))$
$\emptyset \vdash_{HB^d} (\alpha \to (\beta \leftharpoonup \beta)) \leftharpoonup \neg\alpha$

HB^d has two primitive inference rules: $\alpha \leftharpoonup \beta, \beta \vdash_{HB^d} \alpha$ and ${\sim}\neg\alpha \vdash_{HB^d} \alpha$. For the semantical characterization of *HB* and HB^d let us consider a structure

$$\mathcal{M} = (\mathcal{A}, \{D_s : s \in S\}),$$

where

$$\mathcal{A} = (A, \neg, {\sim}, \cap, \cup, \to, \leftharpoonup)$$

is an algebra similar to the *HB*-language, a non-empty set S is partially ordered by $\leq$, and for every $s \in S$, $D_s \subseteq A$.

$\mathcal{M}$ is a model for *HB* if for any $a, b \in A$, $s \in S$:

$$\begin{array}{rcl}
a \in D_s & \textit{implies} & \textit{for any} \quad t \geq s, \quad a \in D_t; \\
\neg a \in D_s & \textit{iff} & \textit{for any} \quad t \geq s, \quad a \notin D_t; \\
{\sim} a \in D_s & \textit{iff} & \textit{for some}\ t \leq s, \quad a \notin D_t; \\
a \cap b \in D_s & \textit{iff} & a \in D_s \ \textit{and} \ b \in D_s; \\
a \cup b \in D_s & \textit{iff} & a \in D_s \ \textit{or} \ b \in D_s; \\
a \rightarrow b \in D_s & \textit{iff} & \textit{for any} \quad t \geq s, \quad a \notin D_t \ \textit{or} \ b \in D_t; \\
a \leftharpoondown b \in D_s & \textit{iff} & \textit{for some} \quad t \leq s, \quad a \in D_t \ \textit{and} \ b \notin D_t.
\end{array}$$

If $\mathcal{M} = (\mathcal{A}, \{D_s : s \in S\})$ is a model for *HB*, then $\mathcal{M} = (\mathcal{A}, \{A - D_s : s \in S\})$ is a model for HB^d. Thus, $\mathcal{M}$ is a model for HB^d if for any $a, b \in A$, $s \in S$:

$$\begin{array}{rcl}
a \in D_s & \textit{implies} & \textit{for any} \quad t \geq s, \quad a \in D_t; \\
\neg a \in D_s & \textit{iff} & \textit{for some} \quad t \leq s, \quad a \notin D_t; \\
{\sim} a \in D_s & \textit{iff} & \textit{for any} \quad t \geq s, \quad a \notin D_t; \\
a \cap b \in D_s & \textit{iff} & a \in D_s \ \textit{or} \ b \in D_s; \\
a \cup b \in D_s & \textit{iff} & a \in D_s \ \textit{and} \ b \in D_s; \\
a \rightarrow b \in D_s & \textit{iff} & \textit{for some} \quad t \leq s, \quad a \notin D_t \ \textit{and} \ b \in D_t; \\
a \leftharpoondown b \in D_s & \textit{iff} & \textit{for any} \quad t \geq s, \quad a \in \mathrm{D}_t \ \textit{or} \ b \notin D_t.
\end{array}$$

In [9] Rauszer formulates an Important Observation (I.O.)[7]:

$$\alpha \rightarrow \beta \textit{ is a tautology of } HB \qquad \textit{iff} \qquad \alpha \leftharpoondown \beta \textit{ is a tautology of } HB^d.$$

This observation can be easily semantically verified. Of course, a tautology of HB^d is a countertautology of *HB*.

Axioms for the reductive part of the Heyting-Brouwer logic are axioms for HB^d with "$\emptyset \vdash_{HB^d}$" replaced by "$L_{HB} \dashv_{HB}$". Primitive rules for the reductive part of *HB* have the following shape:

$$L_{HB} - \{\beta, \alpha \leftharpoondown \beta\} \dashv_{HB} \alpha \qquad \textit{and} \qquad L_{HB} - \{{\sim}\neg\alpha\} \dashv_{HB} \alpha.$$

Let T and F be a C_{HB}-theory and E_{HB}-theory, respectively. Belief revision functions on the base of Heyting-Brouwer logic have an expected form:

EXPANSION:

$$T \oplus_{HB} \varphi = C_{HB}(T + \varphi).$$

CONTRACTION:

$$(T \ominus_{HB} \varphi)_{\alpha_1,\ldots,\alpha_k} = T \cap \bigcap E^T_{HB\alpha_1,\ldots,\alpha_k}(L_{HB} - \varphi).$$

REVISION:

$$(T \odot_{HB} \varphi)_{\alpha_1,\ldots,\alpha_k} = (T \ominus_{HB} \neg\varphi)_{\alpha_1,\ldots,\alpha_k} \oplus_{HB} \varphi.$$

---

[7] "Important Observation" is an original name given by Rauszer.

The verification of Con_1, the first postulate for contraction, uses two facts. The first of them is proved by Rauszer in [9]: let $X \subset L_{HB}$ then,

1. X is a relatively maximal C-theory of HB if and only if $L_{HB} - X$ *is a relatively maximal C-theory of* HB^d*.*

The second fact is a direct conclusion of the equation:

$$E_{HB}(X) = L_{HB} - C^d_{HB}(L_{HB} - X);$$

and says that:

2. The complement of the relatively maximal C-theory of HB^d *is a relatively minimal E-theory of HB as well as the complement of the relatively minimal E-theory of HB is a relatively maximal C-theory of* HB^d*.*

Thus, from facts 1 and 2 it directly follows that:

Conclusion. X is a relatively maximal C-theory of HB if and only if X is a relatively minimal E-theory of HB.

Using the above Conclusion, it is easy to prove that $\bigcap E^T_{HB\alpha_1,\dots,\alpha_k}(L_{HB} - \varphi)$ is a C-theory of *HB*. Thus, $(T \ominus_{HB} \varphi)_{\alpha_1,\dots,\alpha_k}$ also is a C-theory of *HB* and Con_1 is satisfied. The verification of the rest of postulates uses I.O. and is analogous to the verification in the case of classical logic. It means that *HB*-contraction satisfies Con_1-Con_3, Con^*_4, Con^*_5, Con_6, and so *HB*-revision satisfies Rev_1-Rev_6.

4. Generalization

Deductive-reductive form of logic can be especially useful for defining of the belief change procedures. It does not matter if the logic is classical or not. Everything depends on the reconstruction of the reductive part for a given logic. If the reconstruction is possible, the defining of contraction as well as revision is simple. Thus, our proposal opens the belief change area for non-classical logics.

Another benefit of the approach is that the work with formula sets becomes possible. The functions we have presented in the paper have some special form: all of them are defined for a given C-set T and for some sentence φ. However, using the elimination and consequence operations it is possible to redefine all considered reasonings in more general terms, i.e. with a set of sentences $\{\varphi_1, \dots, \varphi_n\}$ instead of a separate sentence φ. Thus, repeating all definitions in a new form we obtain

for Expansion:

$$T \oplus \{\varphi_1, \ldots, \varphi_n\} = \\ C(T \cup \{\varphi_1, \ldots, \varphi_n\}),$$

for Contraction:

$$(T \ominus \{\varphi_1, \ldots, \varphi_n\})_{\alpha_{1_{\varphi_1}}, \ldots, \alpha_{k_{\varphi_1}}, \ldots, \alpha_{1_{\varphi_n}}, \ldots, \alpha_{k_{\varphi_n}}} = \\ T \cap \bigcap E^T_{\alpha_{1_{\varphi_1}}, \ldots, \alpha_{k_{\varphi_1}}, \ldots, \alpha_{1_{\varphi_n}}, \ldots, \alpha_{k_{\varphi_n}}}(L - \{\varphi_1, \ldots, \varphi_n\}),$$

for Revision:

$$(T \odot \{\varphi_1, \ldots, \varphi_n\})_{\alpha_{1_{\varphi_1}}, \ldots, \alpha_{k_{\varphi_1}}, \ldots, \alpha_{1_{\varphi_n}}, \ldots, \alpha_{k_{\varphi_n}}} = \\ (T \ominus \{\neg\varphi_1, \ldots, \neg\varphi_n\})_{\alpha_{1_{\varphi_1}}, \ldots, \alpha_{k_{\varphi_1}}, \ldots, \alpha_{1_{\varphi_n}}, \ldots, \alpha_{k_{\varphi_n}}} \oplus \{\varphi_1, \ldots, \varphi_n\}.$$

Probably the most important side of our approach is that the belief change can take concrete form unlimited to general postulates.

References

[1] Alchourrón, C., Makinson, D., 'On the logic of theory change: contraction functions and their associated revision functions', *Theoria*, 48: 14–37, 1982.

[2] Alchourrón, C., Gärdenfors, P., Makinson, D., 'On the logic of theory change: Partial meet contraction revision functions', *Journal of Symbolic Logic*, 50: 510–530, 1985.

[3] Bryll, G., *Metody Odrzucania Wyrażeń*, Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1996.

[4] Dutkiewicz, R., 'The method of axiomatic rejection for the intuitionistic propositional logic', *Studia Logica*, 48: 449–460, 1989.

[5] Gärdenfors, P., 'Rules for rational changes of belief', in Pauli, T. (ed.), *Philosophical Essays Dedicated to Lennart Aqvist on His Fiftieth Birthday*, University of Uppsala Philosophical Studies 34, 1982, pp. 88–101.

[6] Mc Kinsey, J.C.C., Tarski, A., 'On closed elements in closure algebras', *Ann. of Math.*, 47: 122–162, 1946.

[7] Łukowski, P., 'A deductive-reductive form of logic: general theory and intuitionistic case', *Logic and Logical Philosophy*, 10: 59–78, 2002.

[8] Rauszer, C., 'Semi-Boolean algebras and their applications to intuitionistic logic with dual operations', *Fundamenta Mathematicae*, LXXXIII: 219–249, 1974.

[9] Rauszer, C., *An algebraic and Kripke-style approach to a certain extension of intuitionistic logic*, Dissertationes Mathematicae, vol. CLXVII, PWN, Warszawa, 1980.

[10] Skura, T., 'A complete syntactical characterization of the intuitionistic logic', *Reports on the Mathematical Logic*, 23: 75–80, 1990.

[11] Tarski, A., 'Über einige fundamentale Begriffe der Metamathematik', *Compt. Rend. Séances Soc. Sci. Lett. Varsovie*, cl.III, 23, pp. 22–29.

[12] Wójcicki, R., 'Dual counterparts of consequence operation', *Bulletin of the Section of Logic*, 2(1): 54–57, 1973.

[13] Wójcicki, R., *Lectures on Propositional Calculi*, The Polish Academy of Sciences, Institute of Philosophy and Sociology, Ossolineum, 1984.

[14] Wójcicki, R., *Theory of Logical Calculi: Basic Theory of Consequence Operations*, Synthese Library vol. 199, Kluwer Academic Publishers, 1988.

Piotr Łukowski
Łódź University
Department of Logic and Methodology of Science
ul. Kopcińskiego 16/18
90–232 Łódź, Poland
`lukowski@filozof.uni.lodz.pl`

Hans Rott

Shifting Priorities: Simple Representations for Twenty-seven Iterated Theory Change Operators

Abstract. Prioritized bases, i.e., weakly ordered sets of sentences, have been used for specifying an agent's 'basic' or 'explicit' beliefs, or alternatively for compactly encoding an agent's belief state without the claim that the elements of a base are in any sense basic. This paper focuses on the second interpretation and shows how a shifting of priorities in prioritized bases can be used for a simple, constructive and intuitive way of representing a large variety of methods for the change of belief states — methods that have usually been characterized semantically by a system-of-spheres modeling. Among the methods represented are 'radical', 'conservative' and 'moderate' revision, 'revision by comparison' in its raising and lowering variants, as well as various constructions for belief expansion and contraction. Importantly, none of these methods makes any use of numbers.

Keywords: theory change, belief bases, belief revision, prioritization, iteration, two-dimensional revision operators.

1. Introduction

"All necessary reasoning without exception is diagrammatic," said Charles Sanders Peirce (1903, p. 212). According to Peirce, the only way of understanding logical and mathematical propositions is by perceiving generalities in diagrams. The history of belief revision seems to confirm this thesis. By far the most intuitive representation of what is involved in various operations of belief change uses a modelling by means of systems of spheres (briefly, *SOS*) in the style of Lewis (1973) and Grove (1988). The SOS picture, however, is not without disadvantages. First, while it is excellently suited for the representation of the *changes* of belief states, it does not make for an easy grasp of the *contents* of the belief states in question. Second, SOS's are sets of sets of large cardinalities. A more constructive approach would seem to be welcome in order to turn our semantic intuitions into something more manageable. Third, it is not evident at all where the systems of spheres of possible worlds come from.

Prioritized (or 'stratified') bases, on the other hand, have been used (i) for the representation of an agent's explicit beliefs (e.g., in Rescher 1964,

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*, Trends in Logic, 28

Nebel 1992, Rott 1992, Dubois, Lang and Prade 1994, Williams 1995) as well as (ii) for the compact encoding of belief states (e.g., in Rott 1991b). The motivating ideas are quite different in the two cases. In interpretation (i), it makes an essential difference whether one has p and q separately or conjoined into $p \wedge q$ in the belief base, in interpretation (ii) these are just notational variants without a difference in "meaning". Still the most elaborate account of the first interpretation of belief bases (without prioritization) is due to Hansson (1999). In this paper we are only interested in the second interpretation. Prioritized bases have been used to represent *single* belief states. In this paper, I will explain how they can be used in what appears to me a very elegant way of representing a large variety of *changes of* belief states.

Once one has a syntactic representation that corresponds to the semantic SOS modelling of single belief states, it is natural to ask whether there are operations on these syntactic representations that correspond to reasonable transformations of SOS's. This is the topic of this paper.

2. Representing doxastic states: Prioritized belief bases, entrenchment, systems of spheres

A *prioritized belief base* is a sequence of sets of sentences $\overrightarrow{H} = \langle H_1, \dots, H_n \rangle$. For $i < j$, the elements in H_j are supposed to be more "certain" or "reliable" or more "important" than the elements in H_i, while the elements within each H_i are tied. We presume that there are no incomparabilities. We shall also frequently use the alternative notation

$$\overrightarrow{H} \;=\; H_1 \prec \cdots \prec H_n$$

This generates, in an obvious way, a transitive and complete ordering $\preceq$ between the H_i's, and also between the elements of the H_i's.

If $\overrightarrow{H}$ were intended to be a belief base representing the *explicit beliefs* of an agent, then the syntactical structure of the elements in each H_i would be important. In this paper, however, I am only interested in prioritized belief bases as compact and convenient *representations* of doxastic states (in structured *axiomatizations* as it were). Let us assume in the following for the sake of simplicity that not only the number of H_i's but also each of the individual sets H_i is finite. Under the interpretation as compact representations, then, there is no obstacle to conjoining the elements in each base layer H_i into a single sentence $h_i = \bigwedge H_i$. For the constructions we shall discuss in this paper, no change will result by such a maneuver. Rather than $\overrightarrow{H}$, we can then equivalently use the string

$$\overrightarrow{h} \;=\; h_1 \prec \cdots \prec h_n$$

It will be assumed throughout this paper that the beliefs of the highest priority, H_n or h_n, are consistent. Contradictions may arise only at lower levels.

Let $\top$ (*verum*) and $\bot$ (*falsum*) be the sentential constants that are always true and false, respectively. If we liked, we could extend strings by putting "$\bot \prec$" in front or attaching "$\prec \top$" to the end of $\overrightarrow{h}$. But the former is not necessary because, as we shall soon see, inconsistent "up-sets" are irrelevant anyway.[1] And the latter is not desirable as a general requirement on prioritized belief bases, because we want to allow revision methods that push up contingent sentences to the level of tautologies. As a consequence, the AGM postulate of 'consistency preservation', according to which only an inconsistent input can lead into an inconsistent belief set, is not validated by such methods studied in this paper.

We introduce some notation and abbreviations. Unless otherwise noted, i ranges from 1 to n:

$$H = H_1 \cup \cdots \cup H_n$$

$$H_{\geq i} = H_i \cup \cdots \cup H_n$$

$$H_{>i} = H_{i+1} \cup \cdots \cup H_n \;\text{ for } 0 \leq i \leq n-1$$

$$h := h_1 \wedge \cdots \wedge h_n$$

$$h_{\geq i} := h_i \wedge \cdots \wedge h_n$$

$$h_{>i} := h_{i+1} \wedge \cdots \wedge h_n$$

$$\overrightarrow{H_{\geq i}} \;:=\; \langle H_i, \ldots, H_n \rangle \;=\; H_i \prec \cdots \prec H_n$$

$$\overrightarrow{H_{\leq i}} \;:=\; \langle H_1, \ldots, H_i \rangle \;=\; H_1 \prec \cdots \prec H_i$$

$$\overrightarrow{h_{\geq i}} \;:=\; h_i \prec \cdots \prec h_n \;,\quad \overrightarrow{h_{\leq i}} \;:=\; h_1 \prec \cdots \prec h_i$$

$$\overrightarrow{h \wedge \alpha} := h_1 \wedge \alpha \prec \cdots \prec h_n \wedge \alpha$$

$$\overrightarrow{h \vee \alpha} := h_1 \vee \alpha \prec \cdots \prec h_n \vee \alpha$$

$$\overrightarrow{h \overset{+}{\vee} \alpha} := h_1 \prec h_1 \vee \alpha \prec h_2 \prec h_2 \vee \alpha \prec \cdots \prec h_n \prec h_n \vee \alpha$$

---

[1] This statement has to be qualified. If the dynamics of belief are driven by syntactical manipulations on prioritized belief bases, it does matter whether there are lower levels that make the base inconsistent as a whole. Statically equivalent bases may be dynamically different. I neglect this point in the present paper.

And, for example

$$\overrightarrow{h_{\geq i} \wedge \alpha} := h_i \wedge \alpha \prec \cdots \prec h_n \wedge \alpha$$

$$\overrightarrow{h_{<i} \vee \alpha} := h_1 \vee \alpha \prec \cdots \prec h_{i-1} \vee \alpha \quad \text{for } 2 \leq i \leq n+1$$

$$\overrightarrow{h_{>i} \overset{+}{\vee} \alpha} := h_{i+1} \prec h_{i+1} \vee \alpha \prec \cdots \prec h_n \prec h_n \vee \alpha \quad \text{for } 0 \leq i \leq n-1$$

For $\vec{h} = h_1 \prec \cdots \prec h_n$ and $\vec{g} = g_1 \prec \cdots \prec g_m$ we define the concatenations $\vec{h} \prec. \alpha = h_1 \prec \cdots \prec h_n \prec \alpha$ and $\vec{h} \prec. \vec{g} = h_1 \prec \cdots \prec h_n \prec g_1 \prec \cdots \prec g_m$. (The dot next to a $\prec$ symbol indicates that at least one of the relata is not a set of sentences or a single sentence, but an ordered sequence itself.)

The most important sets definable by prioritized bases are the up-sets $H_{\geq i}$ and the sentences $h_{\geq i}$. They serve as standards of consistency and inconsistency in a way to be explained soon.

Instead of numbers, we can also use *sentences* in order to define the relevant up-sets. If H implies α, let in the following definitions i be the greatest number such that $H_{\geq i}$ implies α (so $H_{>i}$ does not imply α). Then we define $H_{\geq\alpha} = H_{\geq i}$ and $H_{>\alpha} = H_{>i}$ and $H_{=\alpha} = H_i$. If H does not imply α, we set $H_{\geq\alpha} = H_{>\alpha} = H$. Notice that $H_n = H_{=\top}$, but not necessarily $H_1 = H_{=h}$ (but this does hold for purified bases, see below). In the same fashion, we define $h_{\geq\alpha} = h_{\geq i}$, $h_{>\alpha} = h_{>i}$ and $h_{=\alpha} = h_i$, where i is the greatest number such that $h_{\geq i}$ implies α. Notational devices mixing sentences and numbers like $h_{>\alpha+1}$ or $h_{=h-1}$ should be understood in the obvious way.

The *belief set $\mathcal{B}$ supported by a prioritized base* $\vec{H}$ is defined as $Bel(\vec{H}) = Cn\,(H_{>\perp})$. Here and throughout this paper, we use Cn to indicate a consequence operation governing the language that is Tarskian, includes classical propositional logic and satisfies the deduction theorem.[2] Notice that belief sets so conceived are always consistent (except perhaps in the limiting case when H_n is itself inconsistent).

Beliefs in $Bel(\vec{H})$ can be ranked according to their certainty, reliability or importance. We employ a *Weakest Link Principle* according to which a chain is just as strong as its weakest link. Less metaphorically, a set of premises is just as strong as its weakest element. In accordance with this idea, it would be possible to define $rank_H(\alpha)$ to be the largest integer i

---

[2] A logic Cn is Tarskian iff it is reflexive ($H \subseteq Cn\,(H)$), monotonic (if $H \subseteq H'$, then $Cn(H) \subseteq Cn(H')$), idempotent ($Cn(Cn(H)) \subseteq Cn(H)$) and compact (if $\alpha \in Cn(H)$, then $\alpha \in Cn(H')$ for some finite $H' \subseteq H$). The deduction theorem says that $\alpha \to \beta \in Cn(H)$ if and only if $\beta \in Cn(H \cup \{\alpha\})$. We write $H \vdash \alpha$ for $\alpha \in Cn\,(H)$.

such that $H_{\geq i}$ implies α. But here is an important warning: *Numbers don't really mean anything in our framework* – never apply arithmetic operations (addition, subtraction, multiplication) to any such ranks! So let us work with a relation instead:

(Def $\leq$ from $\preceq$) $\alpha \leq \beta$ iff for every i, if $H_{\geq i}$ implies α then it also implies β

Such relations are often called relations of *epistemic entrenchment*. The idea of (Def $\leq$ from $\preceq$) has become folklore in the belief revision literature and was put to use, for instance by Rott (1991b) and Williams (1995). Entrenchment relations were first introduced and axiomatized by Gärdenfors and Makinson (1988). Notice, however, that the Gärdenfors-Makinson 'maximality condition' that says that only logical truths are maximally entrenched is not a necessary property of the entrenchment relations used in this paper.

An alternative and, as we said, more vivid representation of the significance of prioritized bases is in terms of possible worlds or more exactly, in terms of models of the underlying language. A prioritized base may be thought of as structuring the space of all models of the underlying language into a system \$ of nested spheres (à la Lewis 1973 and Grove 1988):

(Def \$ from $\prec$): *The system of spheres* \$ *generated by a prioritized belief base* $\overrightarrow{H}$ is the set of sets S_i of models such that for each i, S_i is the set of models of $H_{\geq i}$, in symbols:

$$\$ = \{\, \mathrm{mod}(H_{\geq i}) : i = 1, \ldots, n\}$$

The idea is that the models of $H = H_{\geq 1}$ are the most plausible worlds, the models of $H_{\geq 2}$ that are not models of $H_{\geq 1}$ are the second most plausible worlds, etc., and the models of H_n that are not models of $H_{\geq n-1}$ are the least plausible worlds — except for those models that do not even satisfy H_n and may be regarded as completely 'inaccessible' to the agent's mind. The set H_n characterizes the agent's 'certainties' or 'commitment set' the elements of which he or she is extremely reluctant to give up.[3]

Now let us introduce an important operation on prioritized belief bases. Prioritized bases can be simplified or 'purified' without affecting the generated positions of the beliefs or worlds. The *purification* of a base $\overrightarrow{H}$ deletes,

---

[3] Though not absolutely reluctant, see for instance the models of moderate and very radical expansion below. Segerberg (1998) unofficially calls what is characterized by H_n 'knowledge'. It seems to me, however, that it is more adequate (though not fully adequate) to identify knowledge with belief that is indefeasible by true inputs, and this does not require maximal entrenchment. For discussions of this concept of knowledge in game-theoretic and epistemological contexts, see Stalnaker (1996) and Rott (2004), respectively.

for every i, all sentences α in H_i which are entailed by $H_{>i}$ (where $i < n$). If after these deletions a set H_i turns out to be empty, it is deleted as a coordinate from $\overrightarrow{H}$. Similarly, the purification of a base $\overrightarrow{h}$ deletes, for every i, every sentence h_i which is entailed by $h_{>i}$ (where again $i < n$). If some H_i or h_i is deleted, so is the symbol '$\prec$' to its right. Purification makes prioritized bases less misleading. If $h_i \prec h_j$ in a purified base, then it is guaranteed that also $h_i < h_j$ in the generated epistemic entrenchment ordering.[4] If $\overrightarrow{H} = H_1 \prec \cdots \prec H_n$ is purified, then the number of spheres in the generated system of spheres \$ is n, and the number of equivalence classes in the generated entrenchment relation $\leq$ is $n+1$ if $\overrightarrow{H}$ is consistent, and n if $\overrightarrow{H}$ is inconsistent. It is easy to see that the entrenchment relation or system of spheres generated by a purified base is identical with that of the unpurified one. For this reason we regard an original base and its purified forms as equivalent. We may always purify a prioritized base, but we are not forced to, when performing any of the belief change operations that follow.

The aim of this paper is to show that the most common qualitative approaches to iterated revision can be represented in a smooth, perspicuous and computationally efficient way as operations on prioritized bases. The operations have a much more constructive flavour than the equivalent operations on entrenchment relations or systems of models. In contrast to the latter, they are syntactic rather than semantic in nature.

Doxastic states $\mathcal{S}$ can be represented, e.g., by systems of spheres of possible models \$, by entrenchment relations $\leq$ or by prioritized belief bases $\overrightarrow{H}$. Doxastic states define belief sets, e.g., using the equations $K = Bel(\$) = \{\alpha : \alpha \text{ is true in all models that are contained in every non-empty } S \in \$\}$, $K = Bel(\leq) = \{\alpha : \bot < \alpha\}$ and $K = Bel(\overrightarrow{H}) = Cn\,(H_{>\bot})$, respectively.[5] In the following, we indeed assume that K is derived from some state $\mathcal{S}$, i.e., from some system of spheres \$, some entrenchment relation $\leq$ or some prioritized belief base $\overrightarrow{H}$. The traditional AGM notation $K * \alpha$ denoting the revised belief set is then to be read as an abbreviation for $Bel(\mathcal{S} * \alpha)$, i.e., $Bel(\$ * \alpha)$, $Bel(\leq * \alpha)$ or $Bel(\overrightarrow{H} * \alpha)$, respectively.

We now take revision operations to operate on doxastic states. The revisions below will be presented as revisions of $\overrightarrow{h}$ rather than $\overrightarrow{H}$, but the

---

[4] Compare the 'Entailment condition' as a test for determining whether an 'E-base' is really an E-base *for its generated entrenchment relation* in Rott (1991b, p. 146).

[5] These definitions guarantee that the belief set is consistent, except in extreme limiting cases in which \$ contains only the empty set, $\top \leq \bot$, or H_n is inconsistent, respectively.

generalization to the latter case will be obvious. Just do 'the same' that is being done to h_i to all the members of H_i individually, and keep them together at their common level. Limitations of space not only make explicit proofs impossible, but also prevent me from dealing with the limiting cases in due detail. The reader is asked to read all coming claims and conditions for revisions by α as restricted to the case in which α is considered possible by the agent, i.e., in which h_n is consistent with α, or equivalently, in which some sphere in $ contains some α-models, or again equivalently, in which $\neg\alpha$ is less entrenched than $\top$.

3. Variants of expansion

In traditional AGM theory, the expansion of a set of plain beliefs consists in simply adding a sentence to a given stock of beliefs and closing under deduction. This is a clear method offering no possibilities of choice. Its disadvantage becomes evident, however, when the input sentence is inconsistent with the prior beliefs. There is room for choice and different methods, however, if we consider expansions of belief states (like prioritized bases, entrenchment relations, system of spheres) rather than just belief sets. We think of expansions as applying sensibly to the paradigm case where the input is consistent with the prior beliefs, that is, with h. No claim is made that the expansion methods must make sense in the belief-contravening case. But it is instructive to study the number of possibilities of expansions (see Figs. 1–5) that mirror quite nicely the corresponding revision and contraction operations. No such analogy between (trivial) expansions and (non-trivial) revisions of belief sets is present in AGM theory.

$$\begin{array}{lccl}
\text{Conservative expansion by } \alpha\text{:} & \overrightarrow{h} & \mapsto & \alpha \prec\cdot\; \overrightarrow{h} \\
\text{Plain expansion by } \alpha\text{:} & \overrightarrow{h} & \mapsto & h_1 \wedge \alpha \prec\cdot\; \overrightarrow{h_{>1}} \\
\text{Moderate expansion by } \alpha\text{:} & \overrightarrow{h} & \mapsto & \overrightarrow{h} \prec\cdot\; \alpha \prec\cdot\; \overrightarrow{h \vee \alpha} \\
\text{Radical expansion by } \alpha\text{:} & \overrightarrow{h} & \mapsto & \overrightarrow{h_{<n}} \prec\cdot\; h_n \wedge \alpha \\
\text{Very radical expansion by } \alpha\text{:} & \overrightarrow{h} & \mapsto & \overrightarrow{h} \prec\cdot\; \alpha
\end{array}$$

One can see immediately the symmetry between plain and radical and between conservative and very radical expansion. The differences are only those between inserting the input sentence *at* the lowest or highest level vs. inserting it in a *newly created* lowest or highest level of priority. It is also very evident now why the moderate method is called 'moderate': the input sentence gets assigned a middle rank. Let us have a look what happens to the number of different levels after purification of the revised prioritized base. Assume the principal case for expansion in which h implies neither α nor $\neg\alpha$. Then plain expansion leaves the number of levels at n, while conservative expansion raises it to $n+1$. Radical expansion give at most n levels, very radical expansion at most $n+1$, and finally, moderate expansion gives at least $n+1$ and at most $2n+1$ levels. (Very) radical expansion tends to coarsen, while moderate expansion tends to refine it.

Very radical expansion accepts α as more certain than all the previous beliefs, thereby making some previously inaccessible α-models accessible. That is, some previously maximally entrenched sentences lose their status as certainties.[6] This prevents very radical expansion from being commutative. In radical (but not very radical) expansion, the new information gets as highly entrenched as the maximal prior information, but not higher than that. This operation is commutative.

4. Radical revision

We take as paradigmatic for revision the case where the new information is incompatible with the original belief set (the *belief-contravening* case). We continue to assume that the agent is bound to accept the input sentence α and denote the posterior entrenchment relation by $\leq'$.

All methods for iterated revision to be discussed in this paper have essentially AGM revision as a limiting case for the case of a one-step revision. In terms of systems of spheres, this means that the innermost sphere of the revised SOS is exactly the intersection of the set of models of α with the smallest sphere in the original SOS that contains any models of α. In terms of entrenchments, a sentence β is more entrenched than $\perp$ with respect to the revised entrenchment relation $\leq'$ if and only if the conditional $\alpha \rightarrow \beta$ is more entrenched than $\neg\alpha$ with respect to the original entrenchment relation $\leq$.[7]

The SOS representations of each of the belief change operations to came are given in the Appendix. In the following main text, we list the corresponding operations on prioritized bases, the entrenchment representations, and finally the characterizations in terms of iterated belief changes.

The first method that we discuss is Segerberg's (1998) 'irrevocable revision' (see Fig. 6), which I like to call 'radical revision'. Fermé (2000) studied the same operation in terms of epistemic entrenchment.

---

[6] The same happens in 'moderate' expansion. Cf. footnote 10 below.

[7] In the limiting case in which no sphere in the SOS contains α-models, or in which $\neg\alpha$ is maximally entrenched, we can decide that the revised SOS and the revised entrenchment relation are identical with the original ones.

Here is a representation of the radical revision of $\overrightarrow{h}$ by the input α.

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<n}} \prec\!. \; h_n \wedge \alpha$$

The new base may then be purified. Equivalently, one could use the representation $\overrightarrow{h \wedge \alpha}$ plus purification.

An even more radical strategy is the recipe of *very radical revision* (see Fig. 7):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h} \prec\!. \; \alpha$$

Here the same comments apply as in the case of very radical expansion.

The revised entrenchment relation generated by the radical revision of a prioritized base is defined by

$$\gamma \leq' \delta \quad \text{iff} \quad \alpha \rightarrow \gamma \leq \alpha \rightarrow \delta$$

the revised entrenchment relation generated by the very radical revision of a prioritized base is defined by

$$\gamma \leq' \delta \quad \text{iff} \quad \alpha \rightarrow \gamma \leq \alpha \rightarrow \delta \,, \;\; \text{and} \;\; \alpha \nvdash \gamma \text{ or } \alpha \vdash \delta$$

The recipe for radical revision corresponds to the rule (RER) of Rott (1991a, p. 171; 2003, p. 130).[8] This operation revises an arbitrary prior entrenchment ordering $\leq$ without assuming that it was generated from a prioritized belief base.

Against the background of the AGM axioms for one-step revisions, radical revision can be characterized in terms of an iterated revision postulate as follows

$$(K * \alpha) * \beta = K * (\alpha \wedge \beta)$$

and very radical revision is similarly characterized by

$$(K * \alpha) * \beta = \begin{cases} K * (\alpha \wedge \beta) & \text{if } \; K * (\alpha \wedge \beta) \nvdash \bot \\ Cn\,(\beta) & \text{if } \; \{\alpha, \beta\} \vdash \bot \\ Cn\,(\alpha, \beta) & \text{otherwise} \end{cases}$$

5. Conservative revision

Conservative revision (see Fig. 8), originally called ‘natural revision’, was advocated and studied by Boutilier (1993, 1996) and Rott (2003).

---

[8] Actually, (RER) in the later paper has an extra clause ‘and $\nvdash \gamma$ or $\vdash \delta$’ that guarantees that $\alpha <' \top$ for non-tautological α. As already mentioned, I do not want to require AGM’s maximality condition in the present paper.

Here is a representation of the conservative revision of $\overrightarrow{h}$ by the input α.

$$\overrightarrow{h} \quad \mapsto \quad \alpha \prec . \overrightarrow{h_{\leq\neg\alpha} \vee \alpha} \prec . \overrightarrow{h_{>\neg\alpha}}$$

If h is purified and does not imply α, no posterior purification is necessary, and the posterior base has $n+1$ levels. If h is purified and does imply α, then the term '$\alpha \prec$' will be dropped in purification, and the posterior base is identical with the n-level prior base $\overrightarrow{h}$.

The revised entrenchment relation $\leq'$ generated by the conservative revision of a prioritized base by α is defined by

$$\gamma \leq' \delta \quad \text{iff} \quad \alpha \rightarrow \gamma \leq \neg\alpha \ , \ \text{ or } \ \gamma \leq \delta \text{ and } \neg\alpha < \alpha \rightarrow \delta$$

This is the condition (CER) for 'conservative entrenchment revision' of Rott (2003, p. 122).

Conservative revision can be characterized in terms of an iterated revision postulate as follows

$$(K * \alpha) * \beta = \begin{cases} K * (\alpha \wedge \beta) & \text{if } \beta \text{ is consistent with } K * \alpha^{9} \\ K * \beta & \text{otherwise} \end{cases}$$

Mirroring the difference between conservative and plain (AGM) expansion, we can define a variant of conservative revision which is obtained by an AGM contraction (see Section 8 below) with respect to $\neg\alpha$, followed by a plain (AGM) expansion by α, i.e., by a version of the well-known Levi identity:

Here is a representation of what might be called the *plain revision* of $\overrightarrow{h}$ by the input α.

$$\overrightarrow{h} \quad \mapsto \quad \alpha \prec . \overrightarrow{h_{>1,\leq\neg\alpha} \vee \alpha} \prec . \overrightarrow{h_{>\neg\alpha}}$$

This operation forgets as it were about the lowest ranked elements of the prioritized belief base, or correspondingly, the innermost ring of the prior system of spheres. Since it does not seem to be a very natural revision operation, I refrain from giving alternative representations of it.

6. Moderate revision

Moderate revision is my name for what is often called 'lexicographic revision' (see Fig. 9). It has been advocated and studied by Nayak and his collabo-

---

[9] In a more general context without 'dispositional coherence', we should put $Cn\,((K * \alpha) \cup \{\beta\})$ in this case rather than $K * (\alpha \wedge \beta)$, see Rott (2003). But given the dispositional coherence encoded in AGM's 7th and 8th axioms, this comes down to the same thing.

rators (1994, 2003), but also by many other researchers. It has become part of the folklore of belief revision research, but here does not seem to be a standard reference paper for it. We present a formulation here that does not presume consistency preservation for revision functions (or the maximality condition for entrenchments).

Here is a representation of the moderate revision of $\overrightarrow{h}$ by the input α.

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h} \prec\!\!. \ \alpha \prec\!\!. \ \overrightarrow{h \vee \alpha}$$

As always, the new base may be purified.

The revised entrenchment relation $\leq'$ generated by the moderate revision of a prioritized base by α is defined by

$$\gamma \leq' \delta \ \text{ iff } \begin{cases} \alpha \rightarrow \gamma \leq \alpha \rightarrow \delta \ \text{ and } \ \alpha \nvdash \gamma & \textit{or} \\ \gamma \leq \delta \ \text{ and } \ \alpha \vdash \delta \end{cases}$$

This is similar to, but not exactly the same as condition (MER) (for ‘moderate entrenchment revision’) of Rott (2003, p. 131). The slight modification suggested here is correct also when the revision function does not satisfy the fifth AGM postulate (‘consistency preservation’) or, equivalently, when the entrenchment relation has not only tautologies as maximal elements (what was excluded in Rott 2003).

Moderate revision can be characterized in terms of an iterated revision postulate as follows[10]

$$(K * \alpha) * \beta = \begin{cases} K * (\alpha \wedge \beta) & \text{if } K * (\alpha \wedge \beta) \text{ is consistent} \\ K * \beta & \text{if } \alpha \vdash \neg\beta \\ Cn\,(\alpha \wedge \beta) & \text{otherwise} \end{cases}$$

7. Restrained revision

Recently, Booth and Meyer (2006) advocated the interesting operation of restrained revision (see Fig. 10), which can be seen as composition of a refinement by α (see Section 9) followed by a conservative revision by α.

---

[10] A somewhat *more moderate revision* could be defined thus:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h} \prec\!\!. \ \alpha \prec\!\!. \ \overrightarrow{h_{<n} \vee \alpha} \prec\!\!. \ h_n$$

In terms of SOS’s, the more moderate revision never turns previously inaccessible worlds into accessible ones (what moderate revision usually does). In terms of entrenchments, more moderate revision replaces the clauses ‘$\alpha \nvdash \gamma$’ and ‘$\alpha \vdash \delta$’ by ‘$\alpha \rightarrow \gamma < \top$’ and ‘$\top \leq \alpha \rightarrow \delta$’, respectively. In terms of iterated revision, more moderate revision replaces the last two lines of moderate revision by ‘$(K*\alpha)*\beta = K*\beta$ otherwise’. Also cf. footnote 15.

Here is a representation of the restrained revision of $\overrightarrow{h}$ by the input α.

$$\overrightarrow{h} \quad \mapsto \quad \alpha \prec.\, \overrightarrow{h_{\leq\neg\alpha} \vee \alpha} \prec.\, \overrightarrow{h_{>\neg\alpha} \overset{+}{\vee} \alpha}$$

plus purification.

The revised entrenchment relation $\leq'$ generated by the restrained revision of a prioritized base by α is defined by

$$\gamma \leq' \delta \text{ iff } \alpha \to \gamma \leq \neg\alpha, \text{ or } \gamma \leq \delta \text{ and } \begin{cases} \alpha \to \gamma \leq \gamma & \text{or} \\ \neg\alpha < \alpha \to \delta \text{ and } \gamma < \alpha \to \delta & \end{cases}$$

While its entrenchment representation is somewhat difficult to comprehend, restrained revision can be characterized elegantly in terms of an iterated revision postulate (Booth and Meyer 2006):

$$(K * \alpha) * \beta = \begin{cases} K * (\alpha \wedge \beta) & \text{if } K * \alpha \nvdash \neg\beta \text{ or } K * \beta \nvdash \neg\alpha \\ K * \beta & \text{otherwise} \end{cases}$$

8. Variants of contraction

The simplest way of getting rid of a belief α is a method that has been called ‘Rott contraction’ by Fermé and Rodriguez (1998), ‘severe withdrawal’ by Pagnucco and Rott (1999) and ‘mild contraction’ by Levi (2004) (see Fig. 11). The method was extended to iterated belief change in Rott (2006).

Here is a representation of the *severe withdrawal* of α from $\overrightarrow{h}$

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{>\alpha}}$$

The revised entrenchment relation corresponding to the severe withdrawal operation with respect to α is this:

$$\gamma \leq' \delta \quad \text{iff} \quad \gamma \leq \alpha \text{ or } \gamma \leq \delta$$

The so-called Levi identity recommends to construct a revision by α through applying an operation of expansion by α after a preparatory contraction by $\neg\alpha$. Accordingly, we can define different concepts of *severe revision* by applying different expansion operations after a severe withdrawal. Let us distinguish three versions of severe revision by α.

This is severe withdrawal combined with conservative expansion (see Fig. 12):

$$\overrightarrow{h} \quad \mapsto \quad \alpha \prec.\, \overrightarrow{h_{>\neg\alpha}}$$

Here is severe withdrawal combined with plain expansion (see Fig. 13):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{=\neg\alpha+1} \wedge \alpha} \prec\cdot \overrightarrow{h_{>\neg\alpha+1}}$$

And here is severe withdrawal combined with moderate expansion (see Fig. 14):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{>\neg\alpha}} \prec\cdot \alpha \prec\cdot \overrightarrow{h_{>\neg\alpha} \vee \alpha}$$

The most faithful extrapolation of one-step *AGM contraction* of belief sets to the revision of belief states seems to be the *conservative contraction* with respect to α (see Fig. 15):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{\leq\alpha} \vee \neg\alpha} \prec\cdot \overrightarrow{h_{>\alpha}}$$

The revised entrenchment relation corresponding to the conservative contraction operation with respect to α is this:

$$\gamma \leq' \delta \quad \text{iff} \quad \gamma \leq \bot \ \text{ or } \ \alpha \vee \gamma \leq \alpha \ \text{ or } \ (\alpha < \alpha \vee \delta \text{ and } \gamma \leq \delta)$$

Notice that $\gamma \leq' \delta$ according to conservative contraction implies $\gamma \leq' \delta$ according to severe withdrawal.

Finally, here is a representation of the *moderate contraction* (see Fig. 16) of $\overrightarrow{h}$ with respect to α.

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{>\alpha}} \prec\cdot \overrightarrow{h \vee \neg\alpha}$$

Nayak, Goebel and Orgun (2007) propose an operation of *lexicographic contraction* which corresponds to Nayak and others' operation of lexicographic revision. This interesting proposal is, however, too complex to receive a treatment in the present paper.

9. Refinement: Neither revision nor contraction

Papini (2001) introduced an interesting belief change operation that is neither a revision nor a contraction operation. It is a kind of *reverse lexicographic belief change*, that I like to call *refinement.* In the system of spheres modelling, each level of the prior system is kept in place, but split in such a way that the α-models of a certain level are after the change made more plausible than the $\neg\alpha$-models of the same level (see Fig. 17).

Refinement of $\overrightarrow{h}$ by input α.

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\neg\alpha}} \prec\cdot \overrightarrow{h_{\geq\neg\alpha} \overset{+}{\vee} \alpha}$$

plus purification.

The revised entrenchment relation $\leq'$ generated by the reverse lexicographic change of a prioritized base by α is defined by

$$\gamma \leq' \delta \quad \text{iff} \quad \gamma \leq \delta \;\; \text{and} \;\; \begin{cases} \alpha \rightarrow \gamma \leq \gamma & \text{or} \\ \gamma < \alpha \rightarrow \delta & \end{cases}$$

Using the notation K/α for the belief set resulting from the refinement of $\overrightarrow{H}$ or $\leq$ or \$ by α, we note that the operation / is not always successful in the way revision operations $*$ are supposed to be successful. More precisely, we have $\alpha \in K/\alpha = Cn\,(K \cup \{\alpha\})$ if and only if α is consistent with K; otherwise $\neg\alpha \in K/\alpha = K$.

There is no characterization of reverse lexicographic belief change ('refinement') in terms of iterated 'revision' postulates, perhaps simply because refinement is no revision operation.[11] Refinement need not have any effects on the belief set level, but may be confined to worlds in outer systems of spheres or to sentences higher up in the entrenchment ranking. We have the property (which is too weak to characterize refinement)

$$K/\alpha/\beta = \begin{cases} K/(\alpha \wedge \beta) = K + (\alpha \wedge \beta) & \text{if } \neg(\beta \wedge \alpha) \notin K \\ K/\alpha = K + \alpha & \text{if } \neg\alpha \notin K, \neg(\beta \wedge \alpha) \in K \\ K/\beta = K + \beta & \text{if } \neg\alpha \in K, \neg\beta \notin K \\ K/\alpha = K/\beta = K/(\alpha \wedge \beta) = K & \text{if } \neg\alpha, \neg\beta \in K \end{cases}$$

10. Two-dimensional operators: Revision by comparison

The idea of two-dimensional belief change operators is that a belief state is transformed in such a way that a sentence α (the 'input') gets accepted with the certainty of a sentence β (the 'reference sentence'). The input is something like '$\beta \leq \alpha$'. The operation of revision by comparison (see Fig. 18) was studied by Cantwell (1997), who called it 'raising', and by Fermé and Rott (2004), who used the notation $\circ_\beta\alpha$. The principal case is when β is more entrenched than α (which we may think of not being accepted in the prior belief state); some interesting limiting cases will be addressed presently.

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\beta}} \prec .\; h_{=\beta} \wedge \alpha \prec .\; \overrightarrow{h_{>\beta}}$$

plus purification.

Fermé and Rott (2004, p. 13) give the following definition of revision by comparison in terms of epistemic entrenchment. Let $\leq$ be a prior entrenchment ordering (usually not thought of arising from an e-base). Assuming

---

[11] A characterization should be possible using a postulate for $K/\alpha * \beta$, where $*$ is an AGM revision function.

again that the agent is to accept α (the input sentence) at least as certainly as β (the reference sentence), the posterior entrenchment relation $\leq' = \leq^*_{\beta\leq\alpha}$ is defined by

$$\gamma \leq' \delta \ \text{ iff } \begin{cases} \beta \wedge (\alpha \to \gamma) \leq (\alpha \to \delta) \ \text{ and } \ \gamma \leq \beta & \textit{or} \\ \gamma \leq \delta \ \text{ and } \ \beta < \gamma \end{cases}$$

It is surprising that the extremely simple operation on prioritized bases indeed captures the operation of revision by comparison which was characterized and studied in rather laborious ways by Fermé and Rott.

There are a number of interesting *unary special cases of revision by comparison.* The special case $\circ_\alpha \bot$ with input sentence $\bot$ and reference sentence α reduces to a severe withdrawal of α (cf. Section 8). The special case $\circ_\top \alpha$ with input sentence α and reference sentence $\top$ reduces to an irrevocable or radical revision by α (cf. Section 4). Another operation worth mentioning is that of irrefutable revision obtained by fixing a reference sentence ε and defining $K * \alpha := K \circ_\varepsilon \alpha$. Though similar with irrevocable revision, especially if a highly entrenched reference sentence ε is chosen, there are some interesting differences (cf. Rott 2006). The change of the prioritized knowledge base cannot further be reduced (see Fig. 19).

11. Two-dimensional operators: Cantwell's lowering

Cantwell (1997) argued that there are two ways of dealing with the situation when we have the prior relation $\alpha < \beta$ and when the input is something like '$\beta \leq \alpha$'. What 'revision by comparison' in the sense of Fermé and Rott does in some intuitive way is to promote α to the rank of β. Although it is problematic to make cross-relational comparisons,[12] the above representation with prioritized belief bases illustrates this: α is simply inserted into the rank of β. But Cantwell saw that there is also a dual operation. One can also obtain the intended effect, in the same principal situation, by demoting β to the rank of α (see Fig. 20).[13] This is the relevant operation on prioritized bases:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\alpha}} \prec . \ h_{=\alpha,\leq\beta} \wedge \beta \prec . \ \overrightarrow{h_{>\alpha,\leq\beta} \vee \neg\beta} \prec . \ \overrightarrow{h_{>\beta}}$$

plus purification.

---

[12] Compare Fermé and Rott (2004, pp. 25–26).

[13] Something like that can happen in 'revision by comparison' as well, see the case of severe withdrawal above. However, the paradigm case for the application of revision by comparison is $\alpha \leq \bot < \beta$, while the paradigm case of lowering is $\bot < \alpha < \beta$.

Assuming that the agent is to accept α (the input sentence) at least as certainly as β (the reference sentence) and that $\beta < \top$, the revised entrenchment relation $\leq' = \leq^*_{\beta_{\leq\alpha}}$ as generated by the lowering of β to the degree of α is defined by the following recipe:[14]

$$\begin{aligned}&\gamma \leq' \delta \ \text{ iff}\\ &(\gamma \leq \delta \text{ and } \gamma \leq \alpha) \text{ or } (\gamma \leq \delta \text{ and } \beta < \beta \vee \delta) \text{ or } (\alpha \leq \delta \text{ and } \beta \vee \gamma \leq \beta)\end{aligned}$$

As far as I know, this condition is new, but it is similar in spirit to Cantwell's axiomatization of lowering. It looks more complicated than it is. Roughly, the explanation for it is this: The old ordering $\leq$ remains undisturbed below α, and indeed the relationship $\gamma \leq \delta$ does not change as long as δ is not lowered (which happens when $\beta < \beta \vee \delta$). If not $\gamma \leq \delta$, we can get a new relationship $\gamma \leq' \delta$ if γ is lowered (which happens when $\beta \vee \gamma \leq \beta$) and δ is at least as entrenched as α.

The above condition does not give us the lowering operation if $\top \leq \beta$, for in this case it reduces to $\gamma \leq' \delta$ iff $\alpha \leq \delta$ or $\gamma \leq \delta$. This operation is a kind of dual to severe withdrawal (which rules that $\gamma \leq' \delta$ iff $\gamma \leq \alpha$ or $\gamma \leq \delta$). While, roughly speaking, severe withdrawal collapses the levels below α into one, this operation collapses the levels above α into one, and in fact into the highest possible one (see Fig. 21).

Given our general assumption that $\alpha < \beta$ in the prior belief state, there are two *unary special cases of the lowering operation.* First, fix $\beta = \top$. Then the lowering operation on bases gives

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\alpha}} \ \prec.\ h_{=\alpha} \wedge \top \ \prec.\ \overrightarrow{h_{>\alpha} \vee \neg\top}$$

which produces no change at all. Tautologies simply cannot be lowered.

Second, fix $\alpha = \bot$. In this case the lowering operation reduces to

$$\overrightarrow{h} \quad \mapsto \quad h_{=\bot} \wedge \beta \ \prec.\ \overrightarrow{h_{>\bot,\leq\beta} \vee \neg\beta} \ \prec.\ \overrightarrow{h_{>\beta}}$$

which results in a conservative contraction (= AGM contraction) with respect to β.

The recipe for lowering is quite similar to the recipe for conservative revision. The obvious input that might show that conservative revision by α is in fact a special case of lowering, namely an extreme lowering of $\neg\alpha$, would

---

[14] In the case $\top \leq \beta$, the revised entrenchment relation $\leq' = \leq^*_{\beta_{\leq\alpha}}$ generated by lowering is defined as

$$\gamma \leq' \delta \quad \text{iff} \quad (\gamma \leq \delta \text{ and } \gamma \leq \alpha) \ \text{ or } \ (\gamma \leq \delta \text{ and } \vdash \beta \vee \delta) \ \text{ or } \ (\alpha \leq \delta \text{ and } \nvdash \beta \vee \gamma)$$

If even $\vdash \beta$, this reduces to $\gamma \leq' \delta$ iff $\gamma \leq \delta$.

be '$\neg\alpha \leq \bot$'. But this doesn't quite give us a revision, it only amounts to conservative contraction with respect to $\neg\alpha$. It eliminates $\neg\alpha$, but it does not promote α to the rank of a belief, i.e., above $\bot$. For revision, we need another kind of lowering operation. Inputs in the form of strict inequalities will help us to solve the problem (see Section 13 below).

12. Gentle raising and lowering

Revision by comparison (raising) is different from lowering even when α and β are 'neighbours' in the sense that in the prior entrenchment ordering, there is no sentence strictly between α and β. One can see this in the operations of 'gentle promotion' and 'gentle demotion', in which the rank of α is raised and lowered by one, respectively (see Figures 22 and 23).

Gentle promotion of α:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{\leq\alpha}} \prec\!. \ h_{=\alpha+1} \wedge \alpha \prec\!. \ \overrightarrow{h_{>\alpha+1}}$$

Gentle demotion of α:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\alpha-1}} \prec\!. \ h_{=\alpha-1} \wedge \alpha \prec\!. \ h_{=\alpha} \vee \neg\alpha \prec\!. \ \overrightarrow{h_{>\alpha}}$$

The reader is invited to compare this with the related operation advocated by Darwiche and Pearl (1997, p. 15).

13. Two-dimensional operators: Raising and lowering by strict comparisons

Now suppose the initial situation is that $\alpha \leq \beta$. Can an inequality $\beta < \alpha$ as input be processed in just the same way as the equality $\beta \leq \alpha$? First we have to be clear that not any input of the form $\beta < \alpha$ is admissible. If α implies β, then α cannot be more entrenched than β. So let us assume that α does not imply β, i.e., that $\alpha \rightarrow \beta$ is not a logical truth, and have a look at the official definitions.

Raising with input $\beta < \alpha$ would seem to be simply this:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{\leq\beta}} \prec\!. \ h_{=\beta+1} \wedge \alpha \prec\!. \ \overrightarrow{h_{>\beta+1}}$$

But there is a precondition here if the operation is to be successful: $h_{>\beta} \wedge \alpha$ must not imply β. If it does, one has to put α somewhere further up, and exactly to the lowest level i such that $h_{\geq i} \wedge \alpha$ does not imply β. So the right idea to meet the constraint specified by the input is this (see Fig. 24):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{\leq(\alpha\rightarrow\beta)}} \prec\!. \ h_{=(\alpha\rightarrow\beta)+1} \wedge \alpha \prec\!. \ \overrightarrow{h_{>(\alpha\rightarrow\beta)+1}}$$

It is clear that the new prioritized base generates the relation $\beta < \alpha$. The two sentences occupy neighbouring layers of entrenchment separated at the left occurrence of '$\prec\!.$'.

Prima facie, *lowering* with input $\beta < \alpha$ would seem to be this:

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\alpha-1}} \prec\cdot\; h_{=\alpha-1} \wedge \beta \prec\cdot\; \overrightarrow{h_{\geq\alpha,\leq\beta} \vee \neg\beta} \prec\cdot\; \overrightarrow{h_{>\beta}}$$

Due to the disjuncts '$\neg\beta$', there is no danger that β is implied by higher levels. However, here we have a problem complementary to the one before. It is no longer guaranteed that the levels higher than β after the change still imply α. The solution is similar. Again we have to replace β by $\alpha \to \beta$. The right recipe turns out to be, after a little simplification (see Fig. 25):

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{<\alpha-1}} \prec\cdot\; h_{=\alpha-1} \wedge \beta \prec\cdot\; \overrightarrow{h_{\geq\alpha,\leq(\alpha\to\beta)} \vee (\alpha \wedge \neg\beta)} \prec\cdot\; \overrightarrow{h_{>(\alpha\to\beta)}}$$

It is clear that the new prioritized base generates the relation $\beta < \alpha$. Again the two sentences occupy neighbouring layers of entrenchment separated at the middle occurrence of '$\prec\cdot$'.

Conservative revision is a special case of lowering with strict inputs. The input is simply $\bot < \alpha$. It turns out that shifting contradictions below the level of α is nothing but conservatively accepting α.

14. Two-dimensional operators: Bounded revision

Conservative revision was soon recognized as being too conservative: Only *very few* α-models are made more plausible than the $\neg\alpha$-models. On the other hand, moderate revision is still fairly radical: *All* α-models are treated as more plausible than *all* the $\neg\alpha$-models. It seems a good idea to employ a two-dimensional operator to steer a middle course. Revision by comparison (raising) and lowering, however, are not the right solutions to this problem, since they are not "between" the one-dimensional operators of conservative and moderate revision, and they do not satisfy the Darwiche-Pearl postulates. *Bounded revision* is a two-dimensional operator that is in a precise sense between conservative and moderate revision. It is motivated and explored in Rott (2007). It seems to dispel Spohn's (1988, pp. 112–113) early complaints about the disadvantages of both conservative and moderate revision.

14.1. Bounded revision, strict version.

The idea of this operation is to accept an input sentence α as long as β holds along with α (see Fig. 26). The reference sentence β functions here as a measure of how much of α the agent should consider most plausible after the change. This idea is similar to that of revision by comparison, but not quite the same. The operation to be considered in this subsection does not move α to an entrenchment level exceeding that of β. (It is only the variant that we are going to consider in subsection 14.2 that achieves this.)

Here is a representation of the strict bounded revision of $\overrightarrow{h}$ by input α as long as β (along with α).

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h} \prec\!\cdot\; \alpha \prec\!\cdot\; \overrightarrow{h_{<(\alpha\to\beta)}} \vee \alpha \prec\!\cdot\; \overrightarrow{h_{\geq(\alpha\to\beta)}}$$

or equivalently, modulo purification,

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{>\neg\alpha,<(\alpha\to\beta)}} \prec\!\cdot\; \alpha \prec\!\cdot\; \overrightarrow{h_{<(\alpha\to\beta)}} \vee \alpha \prec\!\cdot\; \overrightarrow{h_{\geq(\alpha\to\beta)}}$$

Now let us look at the definition of bounded revision in terms of epistemic entrenchment. If $\leq$ is the prior entrenchment ordering, then the posterior entrenchment relation $\leq' = \leq^*_{\alpha;\beta}$ is given by

$$\gamma \leq' \delta \;\text{ iff }\; \begin{cases} \alpha\to\gamma \leq \alpha\to\delta & \text{, if } \alpha\to(\gamma\wedge\delta) < \alpha\to\beta \\ \gamma \leq \delta & \text{, otherwise} \end{cases}$$

In the following equation for iterated revisions, read $K * \alpha := K *_{;\varepsilon} \alpha$ and $K * \beta := K *_{;\varepsilon'} \beta$ for some ε and ε'.

$$(K*\alpha)*\beta \;=\; \begin{cases} K*(\alpha\wedge\beta) & \text{if } \neg(\alpha\wedge\beta) < \alpha\to\varepsilon \\ K*\beta & \text{otherwise} \end{cases}$$

Notation: Here $\neg(\alpha\wedge\beta) < \alpha\to\varepsilon$ is short for the condition that ε is in, but $\neg\beta$ is not in $K*(\alpha\wedge(\neg\beta\vee\varepsilon))$. This abbreviation is in accordance with usual entrenchment theories.

The strict version of bounded revision reduces to *moderate revision* if one takes a logical truth like $\top$ as the reference sentence, except for a limiting case.[15]

14.2. Bounded revision, non-strict version.

The idea of this operation is to accept an input sentence α as long as β holds along with α, and even just a little more (see Fig. 27). The operation of this subsection moves α to an entrenchment level just above that of β. In this respect it is quite close to revision by comparison.

Here is a representation of the non-strict bounded revision of $\overrightarrow{h}$ by input α as long as β (along with α).

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h} \prec\!\cdot\; \alpha \prec\!\cdot\; \overrightarrow{h_{\leq(\alpha\to\beta)}} \vee \alpha \prec\!\cdot\; \overrightarrow{h_{>(\alpha\to\beta)}}$$

---

[15] The difference in the limiting case is precisely that between moderate and more moderate revision. See footnote 10.

or equivalently, modulo purification,

$$\overrightarrow{h} \quad \mapsto \quad \overrightarrow{h_{\geq\neg\alpha,\leq(\alpha\to\beta)}} \prec\cdot\ \alpha \prec\cdot\ \overrightarrow{h_{\leq(\alpha\to\beta)} \vee \alpha} \prec\cdot\ \overrightarrow{h_{>(\alpha\to\beta)}}$$

We again look at the definition of this version of bounded revision in terms of epistemic entrenchment. Let $\leq$ be a prior entrenchment ordering. Then the posterior entrenchment relation $\leq' = \leq^*_{\alpha,\beta}$ is given by

$$\gamma \leq' \delta \;\text{ iff }\; \begin{cases} \alpha \to \gamma \leq \alpha \to \delta, & \text{if } \alpha \to (\gamma \wedge \delta) \leq (\alpha \to \beta) \\ \gamma \leq \delta, & \text{otherwise.} \end{cases}$$

In the following equation, read $K * \alpha := K *_{,\varepsilon} \alpha$ and $K * \beta := K *_{,\varepsilon'} \beta$ for some ε and ε'.

Iterated revision postulate

$$(K * \alpha) * \beta \;=\; \begin{cases} K * (\alpha \wedge \beta), & \text{if } \neg(\alpha \wedge \beta) \leq \alpha \to \varepsilon \\ K * \beta, & \text{otherwise.} \end{cases}$$

Notation: Here $\neg(\alpha \wedge \beta) \leq \alpha \to \varepsilon$ is short for the condition that either $\neg\beta$ is not in or ε is in $K * (\alpha \wedge (\neg\beta \vee \varepsilon))$. This abbreviation is in accordance with usual entrenchment theories.

The non-strict version of bounded revision reduces to *conservative revision* if one takes $\neg\alpha$ as the reference sentence, except for a limiting case.[16]

15. Conclusion

A prioritized belief base represents an agent's belief state. The set of her beliefs as well as her ranking of beliefs in terms of entrenchment can easily be obtained from a prioritized base. The prioritized base representation has, I believe, a number of significant advantages over the more established models. It is compact, constructive and convenient. While the semantics of spheres of possible worlds helps us understand the changes of belief states very well, the syntax of prioritized bases helps us to read off at a glance much of the contents and ranks of a base. We have used bases as compact and convenient tools for representing belief states, without implying that the elements of such a base themselves carry any epistemological weight as "basic" or "explicit" beliefs. Bases are finite and typically have only a comparatively small number of layers and a small number of sentences within

---

[16] In terms of iterated revision, for instance, the difference is as follows. Non-strictly bounded revision with $\varepsilon = \neg\alpha$ gives the inconsistent set $K * \alpha * \beta = K * (\alpha \wedge \beta)$ if $K * \alpha$ is inconsistent, while our official definition of conservative revision gives $K * \alpha * \beta = K * \beta$ in this case. This difference could easily be adapted, if we liked.

each layer. In contrast, other representations of doxastic states typically involve large numbers of possible worlds, or of beliefs to be ordered by some preference relation.

We have presented a fairly wide, though certainly not exhaustive, variety of methods for belief revision by way of manipulations of prioritized bases. These manipulations display quite clearly where in an existing priority ordering the new input is being placed: at the bottom (conservative revision, severe revision), at the top (radical revision) or somewhere in the middle (moderate revision, raising and lowering). There is a surprising multiplicity of revision methods that can be captured in this way. We have collected sphere models of 27 change functions in the Appendix.

A main point of this paper has been to show that prioritized bases are a very good way of representing not only belief states at a certain time, but also changes of belief states. Besides the calculation of implications, the operations to be performed on prioritized bases are: Copying some list of sentences, cutting some such list, applying booleans ($\neg$, $\vee$ and $\wedge$) to the elements of a list, and concatenating lists. Prioritized belief base engineering is a little like DNA engineering. It probably is not realistic psychologically, but it should have nice computational properties. All operations are simple, transparent and give the user an immediate feeling of the status that a new piece of input is assigned in the posterior belief state.

We have gathered considerable inductive evidence that the revisions of belief states systematically defined via SOSs can all be captured by fairly simple syntactical means (prioritized belief base engineering). It does not seem that this can be proved, however, given the vagueness of the terms "systematically defined" and "fairly simple".

After this paper was conceived, I was alerted to the fact that ideas very similar to prioritized base changes as presented here have already been explored in the framework of possibilistic logic in a series of papers, e.g. in Benferhat, Dubois and Prade (2001). The work of Meyer, Ghose and Chopra (2001) is also relevant. The research of both groups was done in the more general (and more interesting) area of belief merging. I recommend the reader to closely study these works and also consult the references mentioned therein. The present paper complements these earlier works in the following respects. The presentation as given here is somewhat simpler; I survey a larger number of methods of iterated belief change that can all lay claim to being regarded as rational; and finally, I make it fully clear that no numbers are needed for any of the belief change methods considered.

All the methods considered are purely qualitative, in the sense that there are no meaningful numbers involved. The numbers used in the representation of prioritized belief bases, as well as the numbers appearing in the sphere pictures only encode orderings. In view of the abundance of qualitative methods at our disposal, we are not likely to subscribe to the view of proponents of numerical methods, according to which purely qualitative methods will always remain too poor to model a reasonable evolution of our beliefs. The problem is rather the reverse: We are facing an embarrassment of riches. What we urgently need is some substantive metatheory that tells us which method to apply in what situations. Unfortunately, we do not have anything like such a methodology yet.

Acknowledgements. This paper is a corrected and improved version of a paper with the same title that appeared in Henrik Lagerlund, Sten Lindström and Rysiek Sliwinski (eds.), *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, Uppsala Philosophical Studies, Vol. 53, University of Uppsala 2006. First versions of this paper were presented at Schloss Dagstuhl and the Certosa di Pontignano, Siena, in 2005. I am grateful to the organizers Jim Delgrande, Jérôme Lang, Cristiano Castelfranchi and Fabio Paglieri for the opportunity to give the paper there, and to the audiences for excellent discussions. I thank Didier Dubois and Henri Prade for drawing my attention to the existence of related work in possibilistic logic, Eduardo Fermé, Brian Hill, Franz Huber, and two anonymous referees for helpful comments on earlier versions of this paper, and Ralf Busse for making me aware that Peirce's beautiful quote on diagrammatic reasoning is relevant for the theory of belief revision. Last but not least I want to thank David Makinson, Jacek Malinowski and Heinrich Wansing for encouraging me to submit this paper for inclusion in the present volume.

References

Alchourrón, Carlos, Gärdenfors, Peter, Makinson, David (1985), 'On the logic of theory change: Partial meet contraction functions and their associated revision functions', *Journal of Symbolic Logic*, 50: 510–530,

Areces, Carlos, Becher, Veronica (2001), 'Iterable AGM functions', in Williams, Rott (eds.), pp. 261–277.

Benferhat, Salem, Dubois, Didier, Prade, Henri (2001), 'A computational model for belief change and fusing ordered belief bases', in Williams, Rott (eds.), pp. 109–134.

Booth, Richard, Meyer, Thomas (2006), 'Admissible and restrained revision', *Journal of Artificial Intelligence Research*, 26: 127–151.

Boutilier, Craig, (1993), 'Revision sequences and nested conditionals', in Bajcsy, R. (ed.), *IJCAI-93—Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pp. 519–525.

Boutilier, Craig (1996), 'Iterated revision and minimal change of conditional beliefs', *Journal of Philosophical Logic*, 25: 263–305.

Cantwell, John (1997), 'On the logic of small changes in hypertheories', *Theoria*, 63: 54–89.

Darwiche, Adnan, Pearl, Judea (1997), 'On the logic of iterated belief revision', *Artificial intelligence*, 89: 1–29.

Dubois, Didier, Lang, Jérôme, Prade, Henri, 'Possibilistic logic', in: Gabbay, D.M., Hogger, C.J., Robinson, J.A. (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3, Nonmonotonic Reasoning and Uncertain Reasoning*, Clarendon Press, Oxford, 1994, pp. 439–513.

Fermé, Eduardo (2000), 'Irrevocable belief revision and epistemic entrenchment', *Logic Journal of the IGPL*, 8: 645–652.

Fermé, Eduardo, Rodriguez, Ricardo (1998), 'A brief note about Rott contraction', *Logic Journal of the IGPL*, 6: 835–842.

Fermé, Eduardo, Rott, Hans (2004), 'Revision by comparison', *Artificial Intelligence*, 157: 5–47.

Gärdenfors, Peter, Makinson, David (1988), 'Revisions of knowledge systems using epistemic entrenchment', in Vardi, M. (ed.), *Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann, Los Altos, CA, 1988, pp. 83–95.

Grove, Adam (1988), 'Two modellings for theory change', *Journal of Philosophical Logic*, 17: 157–170.

Hansson, Sven O. (1999), *A Textbook of Belief Dynamics. Theory Change and Database Updating*, Kluwer, Dordrecht.

Levi, Isaac (2004), *Mild Contraction: Evaluating Loss of Information due to Loss of Belief*, Oxford University Press, Oxford.

Lewis, David (1973), *Counterfactuals*, Blackwell, Oxford.

Meyer, Thomas, Ghose, Aditya, Chopra, Samir (2001), 'Syntactic representations of semantic merging operations', in *Proceedings of the IJCAI-2001 Workshop on Inconsistency in Data and Knowledge*, Seattle, USA, August 2001, pp. 36–42.

Nayak, Abhaya C. (1994), 'Iterated belief change based on epistemic entrenchment', *Erkenntnis*, 41: 353–390.

Nayak, Abhaya C., Pagnucco, Maurice, Peppas, Pavlos (2003), 'Dynamic belief revision operators', *Artificial Intelligence*, 146: 193–228.

Nayak, Abhaya, Goebel, Randy, Orgun, Mehmet (2007), 'Iterated belief contraction from first principles', *International Joint Conference on Artificial Intelligence (IJCAI'07)*, pp. 2568–2573.

Nebel, Bernhard (1992), 'Syntax-based approaches to belief revision', in: Gärdenfors, Peter (ed.), *Belief Revision*, Cambridge University Press, Cambridge, pp. 52–88.

Pagnucco, Maurice, Rott, Hans (1999), 'Severe withdrawal—and recovery', *Journal of Philosophical Logic*, 28: 501–547. (Full corrected reprint in the JPL issue of February 2000.)

Papini, Odile (2001), 'Iterated revision operations stemming from the history of an agent's observations', in Williams, Rott (eds.), pp. 279–301.

PEIRCE, CHARLES S. (1903), 'The nature of meaning', Harvard Lecture delivered on 7 May 1903, published in *The Essential Peirce*, vol. 2 (1803–1913), ed. by the Peirce Edition Project, Indiana University Press, Bloomington, 1998, pp. 208–225.

RESCHER, NICHOLAS (1964), *Hypothetical Reasoning*, North-Holland, Amsterdam.

ROTT, HANS (1991a), 'Two methods of constructing contractions and revisions of knowledge systems', *Journal of Philosophical Logic*, 20: 149–173.

ROTT, HANS (1991b), 'A non-monotonic conditional logic for belief revision I', in Fuhrmann, A., Morreau, M., *The Logic of Theory Change*, LNCS vol. 465, Springer, Berlin, pp. 135–181.

ROTT, HANS (1992), 'Modellings for belief change: Prioritization and entrenchment', *Theoria*, 58: 21–57.

ROTT, HANS (2000), ' "Just because": Taking belief bases- seriously', in Buss, Samuel R., Hájek, Petr, Pudlák, Pavel (eds.), *Logic Colloquium '98—Proceedings of the 1998 ASL European Summer Meeting*, Lecture Notes in Logic, vol. 13, Urbana, Ill. Association for Symbolic Logic, pp. 387–408.

ROTT, HANS (2001), *Change, Choice and Inference*, Oxford University Press, Oxford.

ROTT, HANS (2003), 'Coherence and conservatism in the dynamics of belief. Part II: Iterated belief change without dispositional coherence', *Journal of Logic and Computation*, 13: 111–145.

ROTT, HANS (2004), 'Stability, strength and sensitivity: converting belief into knowledge', in Brendel, Elke, Jäger, Christoph (eds.), *Contextualisms in Epistemology*, special issue of *Erkenntnis* 61: 469–493.

ROTT, HANS (2006), 'Revision by comparison as a unifying framework: Severe withdrawal, irrevocable revision and irrefutable revision', *Theoretical Computer Science*, 355: 228–242.

ROTT, HANS (2007), 'Bounded revision: Two-dimensional belief change between conservatism and moderation', in Rønnow-Rasmussen, Toni et al. (eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*, internet publication, http://www.fil.lu.se/hommageawlodek/site/abstra.htm.

SEGERBERG, KRISTER (1998), 'Irrevocable belief revision in dynamic doxastic logic', *Notre Dame Journal of Formal Logic*, 39: 287–306.

SPOHN, WOLFGANG (1988), 'Ordinal conditional functions', in Harper, W.L., Skyrms, B. (eds.), *Causation in Decision, Belief Change, and Statistics, vol. II*, Reidel, Dordrecht, pp. 105–134.

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

WILLIAMS, MARY-ANNE (1994), On the logic of theory base change', in MacNish, C., Pearce, D., Pereira, L.M. (eds.), *Logics in Artificial Intelligence*, LNCS, vol. 838, Springer, Berlin, pp. 86–105.

WILLIAMS, MARY-ANNE (1995), 'Iterated theory base change: A computational model', in *IJCAI'95—Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Mateo, pp. 1541–1550.

WILLIAMS, MARY-ANNE, ROTT, HANS (eds.), (2001), *Frontiers in Belief Revision*, Kluwer, Dordrecht.

Appendix: Sphere pictures

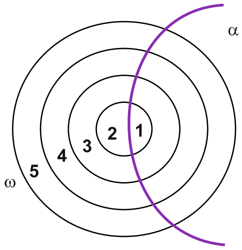


Figure 1: Conservative expansion

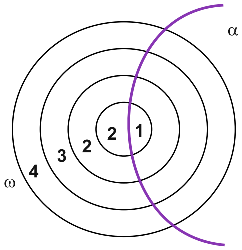


Figure 2: Plain expansion

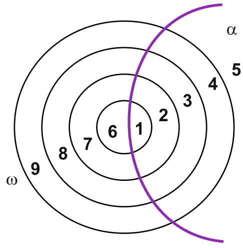


Figure 3: Moderate expansion

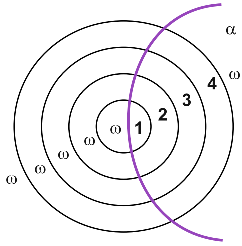


Figure 4: Radical expansion

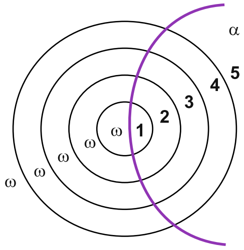


Figure 5: Very radical expansion

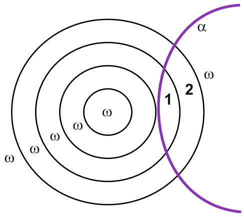


Figure 6: Radical revision (= 'irrevocable revision')

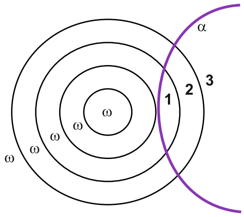


Figure 7: Very radical revision

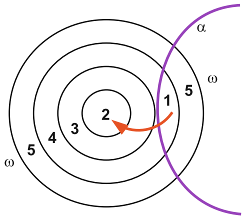


Figure 8: Conservative revision (= 'natural revision')

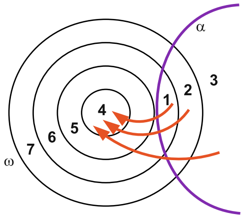


Figure 9: Moderate revision (= 'lexicographic revision')

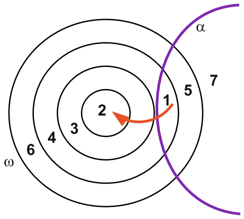


Figure 10: Restrained revision

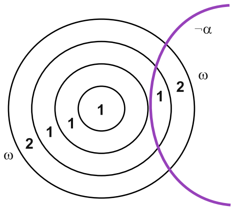


Figure 11: Severe withdrawal (= 'mild contraction')

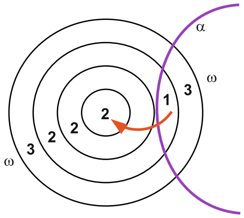


Figure 12: Severe revision

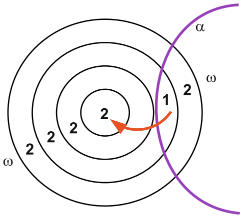


Figure 13: Plain severe revision

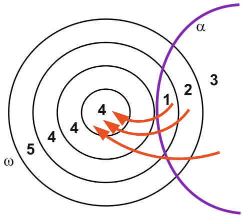


Figure 14: Moderate severe revision

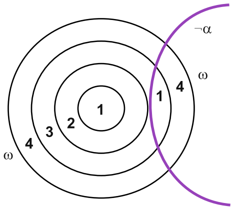


Figure 15: Conservative contraction (≈ AGM contraction)

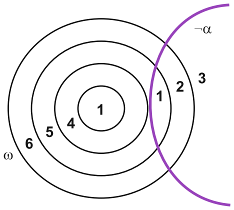


Figure 16: Moderate contraction

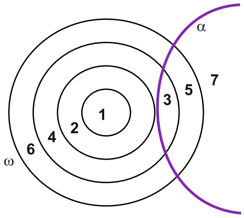


Figure 17: Refining (= 'Reverse lexicographic belief change')

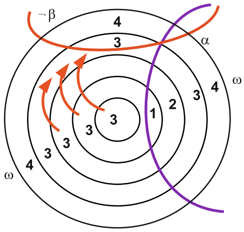


Figure 18: Revision by comparison (= 'raising')

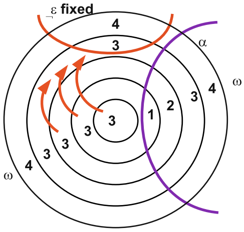


Figure 19: Irrefutable revision (with fixed ε)

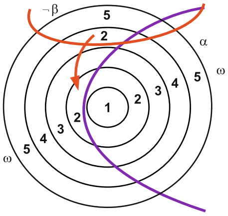


Figure 20: Lowering

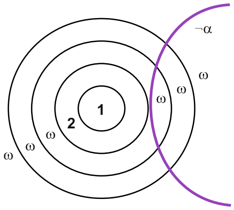


Figure 21: Dual to severe withdrawal

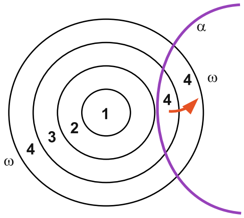


Figure 22: Gentle raising

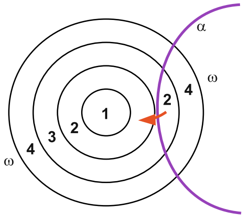


Figure 23: Gentle lowering

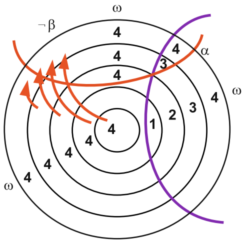


Figure 24: Raising by strict comparison

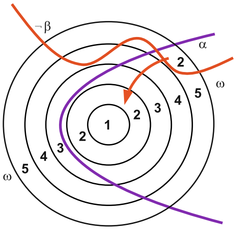


Figure 25: Lowering by strict comparison

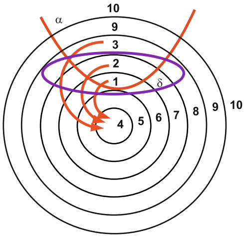


Figure 27: Bounded revision (non-strict version)

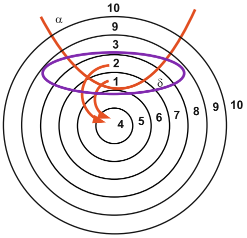


Figure 26: Bounded revision (strict version)

Hans Rott
Institut für Philosophie
Universität Regensburg
93040 Regensburg, Germany
`Hans.Rott@psk.uni-regensburg.de`

Jason Jingshi Li
Rex Bing Hung Kwok
Norman Foo

The Coherence of Theories — Dependencies and Weights

Abstract. One way to evaluate and compare rival but potentially incompatible theories that account for the same set of observations is coherence. In this paper we take the quantitative notion of theory coherence as proposed by [Kwok et al., 1998] and broaden its foundations. The generalisation will give a measure of the efficacy of a sub-theory as against single theory components. This also gives rise to notions of dependencies and couplings to account for how theory components interact with each other. Secondly we wish to capture the fact that not all components within a theory are of equal importance. To do this we assign weights to theory components. This framework is applied to game theory and the performance of a coherentist player is investigated within the *iterated Prisoner's Dilemma*.

Keywords: coherence, philosophy of science, theory evaluation, game theory.

1. Introduction

The core of scientific theories are laws. These laws often make use of theoretical terms, linguistic entities which do not directly refer to observables. There is therefore no direct way of determining which theoretical assertions are true. This suggests that multiple theories may exist which are incompatible with one another but compatible with all possible observations. Since such theories make the same empirical claims, empirical tests cannot be used to differentiate or rank such theories. Hawking very nicely summarised this positivist approach in the philosophy of science: "A scientific theory is a mathematical model that describes and codifies the observations we make. A good theory would describe a large range of phenomena on the basis of a few postulates, and make definite predictions that can be tested" [Hawking, 2001]. One property that has been suggested for evaluating rival theories is *coherence*. This was investigated qualitatively in the philosophy of science (see, e.g, summaries in [van Fraassen, 1980] and [Nagel, 1961]) until [Kwok et al., 1998] introduced a coherence measure based on the average use of formulas in accounting for observations. Prior to this measure, the qualitative approaches considered properties of theories typified by informal notions

---

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

like "tightness of coupling" of the axioms, "brevity", "predictive scope", etc. Kwok et.al. (op.cit.) took these as guides for their quantification. The idea was to identify highly coherent theories as those whose formulas are tightly coupled in accounting for observations, while low coherence theories contain many disjointed and isolated statements. It proved to be quite fruitful; for instance this provided a rebuttal to Craig's method [Craig, 1953] for the elimination of theoretical terms by showing that the method yields theories with very low coherence.

Later work [Kwok et al., 2003], [Kwok et al., 2007] by the same authors generalised the approach to better mirror scientific practice. For instance, a standard way to use a theory is to design experiments with varying input and output sets. However, another way is to regard observations as inputs and explanations as outputs. The generalisation accommodates both views, and in fact permits other interpretations of input-output relations to test theories for coherence. It is also able to explain notions like theory modularisation.

It is fair to say that this approach to reifying coherence is in effect a combinatorial grounding that relies on the widely understood concept of support sets that plays an important role in artificial intelligence logic in areas as diverse as diagnoses, logic program semantics and abduction. One may question whether the hitherto qualitative notion of coherence is appropriately captured by our quantitative measure. Our response is that we propose a *plausible* way to fix the interpretation of coherence that can be tested by its efficacy in explicating some well-known examples, with the awareness that other plausible methods may emerge in future that capture variant qualitative interpretations.

In the current paper we take the above as starting points and widen the foundations of coherence as defined through support sets. Two ideas are broached, based on intuitions from scientific practice that were not considered in [Kwok et al., 1998] and [Kwok et al., 2003]. The first widening derives from the observation that coherence should also measure how well pairs, triples, etc. of formulas jointly account for observations or outputs. This gives rise to the quantitative notion of dependency in coherence. The second widening mirrors the practical fact that not all formulas may be considered to be equal in importance. This is already acknowledged in the works on belief revision, primarily the AGM approach [Gardenfors, 1988], where varying commitments to particular beliefs goes by the name of entrenchment. The possibilistic logicians' fuzzy measures aimed at capturing the same intuition have been shown to be equivalent to entrenchment. In our paper we use weights on formulas to do this. This enhanced definition of coherence reduces to the previous version when dependencies are among singletons and all weights are equal.

Numerous formal examples will illustrate the efficacy of the new definition, but we also apply it to a domain not traditionally considered in the philosophy of science which initially motivated our work. The domain is game theory, specifically forms of the (in)-famous Prisoner's Dilemma [Axelrod, 1981], where the one-time game is classically represented as a matrix that displays the payoffs for each of the two players depending on their choice of action (called strategy). Game theorists then assume rational decisions by each player and analyse the action choices that must be entailed. Iterating the one-time game was then studied by a number of researchers (see, e.g. Axelrod [Axelrod, 1981]). It is this iterated version to which we will apply the notion of coherence. We will model a player's reasoning (using its beliefs, desires and intentions) as formulas, and the player's adaptations during the game is seen as attempts to maintain high coherence among these doxastic qualities. Computer simulations of this approach are also described and analysed.

Finally we discuss future directions that this work may profitably take. It is plausible that traditional norms of "rationality" in the evaluation of scientific theories as well as economic and social behaviour may be modulated by current discomfort with the policies that result from them. Wider notions of what it means for these theories to be coherent can contribute to modifications of the existing norms.

2. Internalist Coherence

This section reviews the previous contribution by [Kwok et al., 2003], and suggests innovations in areas that were not addressed up to this date, such as the utility of a set of formulas, and the relationship between sets of formulas: how one may dominate over another, and how tightly they have coupled to account for observations. It can also be seen the other way round, as how closely they have been associated when supported by empirical evidence. For the time being, we call these nominated properties "Internalist Coherence".

2.1. Support Sets

The building blocks of coherence are support sets. They describe how a theory accounts for an observation from specific inputs. In this framework, a *theory*, an *input* set and an *output* set are all sets of formulas from a first-order language $\mathcal{L}$. It is appropriate to motivate the setting assumed by the next definition. We conceive of logical theories as formal models of selected aspects of the world that interest us. In science the theories of a domain such as chemistry are often painfully constructed over the course of time, and subject to much testing and revision. We do not address the revision issue here, but as we shall see the testing is implicit. A theory T can be used in many ways. It may help to visualise T as a blackbox into which the "input" set I formulas is fed, and an "output" set of formulas O is produced. The "directionality" suggested by these terms should not be taken literally. The interpretation of I and O depends on how T is intended to be used. For instance, O could be a set of observed outcomes of an experiment, in which case I could describe the initial conditions of that experiment. If given certain hypotheses, we interpret O as desired conclusions of T, I could be such a set of hypotheses. Moreover, T itself can have atoms which say that we are only interested in models of T that satisfy those atoms. It is a matter of modelling to decide which atoms ("facts") to place in I, O or T, and different choices will yield different coherence measures. To see that this flexibility is an advantage, consider the following. Suppose someone proposes a theory T that purports to account for some phenomena. If we wish to test T only in settings where conditions C hold, one way to do that is to consider instead the theory $T \cup \{C\}$. But if we already have a set O of observations, and we wish to find conditions C under which T can account for O, then C is part of I.

For brevity in the sequel we sometimes use the term *axiom* for an element of T.

Definition 1 (Support Sets [Kwok et al., 2003]). Given input set I, output set O, a subset of the theory T be Γ. Γ is an I-relative support set of O if

1. $\Gamma \wedge I \models O$ and
2. Γ is minimal (wrt set inclusion).

Let $S(T, I, O)$ denote the family of all I-relative support sets for O. As explained above different choices of input set I will result in different support sets. This approach is designed to be "independent of any commitment to causality or particular use of laws" [Kwok et al., 1998]. This definition is intended to capture the idea that Γ alone cannot account for O but it can do that with the help of I; moreover we want I to be as small as possible, viz. no redundancy.

Example 1 (Socrates is Mortal). Given the input I:

$$I : \{man(Socrates)\} - Socrates\ is\ a\ man$$

output O:

$$O : \{mortal(Socrates)\} - Socrates\ is\ mortal$$

the theory T:

$$T = \{\alpha_1 : \forall(x)\ man(x) \rightarrow mortal(x), \alpha_2 : \forall(x)\ deity(x) \rightarrow \neg mortal(x)\}$$
$$-all\ men\ are\ mortal, -all\ deities\ are\ not\ mortal$$

$\{\alpha_1\}$ constitutes a support set for I and O, since it explains how O is derived from I, whereas $\{\alpha_2\}$ does not constitute a support for I and O.

EXAMPLE 2. Let T be the theory that geniuses would only pass if they are not intoxicated; and if one is not a genius, then one would only pass after study:

$\neg genius(x) \wedge \neg study(x) \rightarrow \neg pass(x)$
$\neg genius(x) \wedge study(x) \rightarrow pass(x)$
$genius(x) \wedge intoxicated(x) \rightarrow \neg pass(x)$
$genius(x) \wedge \neg intoxicated(x) \rightarrow pass(x)$

Suppose we wish to explain an output set $O = \{\neg pass(john)\}$. Possible input sets are:

$I_1 = \{genius(john), intoxicated(john)\}$ and
$I_2 = \{\neg genius(john), \neg study(john)\}$.

Observe that we may re-interpret O as a prediction given the input information I_1 or I_2. For this O the second and fourth formulas in T are not used. However, should O be changed to $\{\neg pass(john), pass(verana)\}$ it can be seen that all the formulas in T will be used to compute the input support sets.

2.2. Utility of a set of formulas

Recall the informal properties of coherence, such as "tightness of coupling" and "work together", that we wish to encapsulate in our formal quantitative framework. One element missing from the previous approach [Kwok, et.al. 03] was the notion of measuring the usefulness of a group of formulas, or a sub-theory. This is an important concern as the utility of the sub-theory would reflect both the utility of the components of the sub-theory, and the tightness of the coupling between the components, and thus capture some of the desired properties of coherence in our representation.

We wish to measure the contribution of not only one formula, but several formulas in how they *together* have contributed to support observations. Building on the [Kwok et al., 2003] definition, we now examine how a set of formulas "work together". For instance, in a theory T consisting of elements α, β and γ; we may wish to consider not only the individual utilities of elements α and β, but their synergistic qualities of working together, e.g. the utility of the set $\Theta = \{\alpha, \beta\}$.

The next definition formalises this intuition. A higher level of utility for a set means that its formulas occur together often in support of observations.

DEFINITION 2 (Utility of a Set of Formulas). Given a theory T and a non-empty set of formulas $A \subseteq T$, its utility is:

$$U(A, T, I, O) = \frac{\mid \{\Gamma : A \subseteq \Gamma \text{ and } \Gamma \in S(T, I, O)\} \mid}{\mid S(T, I, O) \mid} \quad \text{if} \quad S(T, I, O) \neq \emptyset$$

This formal definition provides a measure of how well all formulas in the set "work together" in supporting observations. It sees the formulas as equal, and does not discriminate one over another. Informally the idea is as follows. To measure the utility of the set A we do this: first count how many times it appears within the support sets for the given I and O; we then express this as a fraction of the total number of those support sets — hence the more frequently A so appears the higher its utility. If one formula does not work with the group, the utility for the group will be rendered as zero. The connection between the utility of individual formulas (singleton set) and the utility of sets of which it is a member is addressed in Lemma 1 below.

LEMMA 1 (Joint Utility). *Let a set of formulas A consist of two proper subsets B and Δ, i.e. $A = B \cup \Delta$. The following properties hold:*

(i) $$U(A, T, I, O) = U(B \cup \Delta, T, I, O)$$

(ii) *if $S(T, I, O) \neq \emptyset$, then*

$$U(A, T, I, O) = \frac{\mid \{\Gamma : B \subseteq \Gamma \ \wedge \ \Delta \subseteq \Gamma \ \wedge \ \Gamma \in S(T, I, O)\} \mid}{\mid S(T, I, O) \mid}$$

since

$$\{\Gamma : B \cup \Delta \subseteq \Gamma\} = \{\Gamma : B \subseteq \Gamma \wedge \Delta \subseteq \Gamma\}$$

This will be useful in subsequent proofs where sets of axioms appear together.

2.3. Dependencies between formulas

The "tightness of coupling" between elements of a theory can be reflected in two ways. We shall elaborate the two different senses of "tightness" over the next two sections. First, this property can be exhibited in the reliance of one set of formulas upon another. For example, to account for the observation "Socrates is mortal", the axiom "Socrates is a man" would not make sense without the other axiom "all men are mortal". However, if there are two independent explanations of Socrates' mortal nature based on he is a man, then the axiom "Socrates is a man" would be less dependent on each of the set of formulas that amounts to the respective explanations.

Formally, we wish to see how dependent a specific set of formulas is upon another. It may be that this set in isolation is not a support set, but that in combination with another set it is one; then informally the first set can be regarded as dependent on the second. More precisely, if set Φ is contained in most of the support sets that contain another set Θ, then Θ would have a high dependency on Φ. This dependency is generally asymmetric.

Definition 3 (Dependency Coefficient).

$$D(\Theta, \Phi, T, I, O) = \frac{\mid \{\Gamma : \Gamma \in S(T, I, O) \text{ and } \Theta \subseteq \Gamma \text{ and } \Phi \subseteq \Gamma\} \mid}{\mid \{\Gamma : \Gamma \in S(T, I, O) \text{ and } \Theta \subseteq \Gamma\} \mid}$$

This defines the dependency of Θ on Φ.

The dependency above also reflects the importance of the set Φ. Consider a formula α in T that not only occurs in most support sets, but where other formulas are dependent on it to make a support set, this then makes α important in T. This can be captured as the *weight* of a formula which we discuss later. Section 2.3 discusses the use of dependencies.

Dependency is related to utility. Given two sub-theories Θ and Φ, the dependency of Θ to Φ measures the proportion of support sets that contain both Θ and Φ against those that contain Θ. The higher the dependency, the more support sets that contain Θ also contain Φ.

Corollary 1 (Dependency-Utility Connection).

$$D(\Theta, \Phi, T, I, O) = \frac{U(\Theta \ \cup \ \Phi, T, I, O)}{U(\Theta, T, I, O)}$$

Proof. Recall:

$$D(\Theta, \Phi, T, I, O) = \frac{\mid \{\Gamma_1 : \Theta \subseteq \Gamma_1 \text{ and } \Phi \subseteq \Gamma_1 \text{ and } \Gamma_1 \in S(T, I, O)\} \mid}{\mid \{\Gamma_2 : \Theta \subseteq \Gamma_2 \text{ and } \Gamma_2 \in S(T, I, O)\} \mid}$$

Divide numerator and denominator by $\mid S(T, I, O) \mid$:

$$D(\Theta, \Phi, T, I, O) = \frac{\dfrac{\mid \{\Gamma_1 : \Theta \subseteq \Gamma_1 \text{ and } \Phi \subseteq \Gamma_1 \text{ and } \Gamma_1 \in S(T, I, O)\} \mid}{\mid S(T, I, O) \mid}}{\dfrac{\mid \{\Gamma_2 : \Theta \subseteq \Gamma_2 \text{ and } \Gamma_2 \in S(T, I, O)\} \mid}{\mid S(T, I, O) \mid}}$$

Then from Lemma 1, translate the numerator and denominator back to utility:

$$D(\Theta, \Phi, T, I, O) = \frac{U(\Theta \ \cup \ \Phi, T, I, O)}{U(\Theta, T, I, O)} \qquad \blacksquare$$

2.4. Coupling of formulas

The second way to encapsulate the "tightness of coupling" property is to see how elements of a theory mutually need each other. That is, how much they "work together" in proportion to the total amount of work they do in forming I-relative support sets. The greater the ratio, the "tighter" the elements coupled together. This is different to the previous definition of dependency, as this looks at how much both sub-theories take part in accounting for observations.

We wish to formalise a notion of mutual dependency between two sub-theories. Intuitively, this will measure the degree to which the sub-theories need each other in accounting for observations. The following symmetric definition formalises this intuition.

Definition 4 (Coupling Coefficient).

$$CP(\Theta, \Phi, T, I, O) = \frac{\mid \{\Gamma_1 : \Theta \subseteq \Gamma_1 \text{ and } \Phi \subseteq \Gamma_1 \text{ and } \Gamma_1 \in S(T, I, O)\} \mid}{\mid \{\Gamma_2 : (\Theta \subseteq \Gamma_2 \text{ or } \Phi \subseteq \Gamma_2) \text{ and } \Gamma_2 \in S(T, I, O)\} \mid}$$

This coupling coefficient represents how two sub-theories mutually need each other. The higher the coupling, the more they work together, reflecting the properties of coherence as stated from the informal definition proposed by [Kwok et al., 1998].

2.5. Example — Socrates is wise

Example 3 (Socrates is wise). Consider the proposal that Socrates is wise because he had a wise student named Plato. Plato, apart from being wise, was also a prolific writer in philosophy. Therefore, we may have two possible ways of accounting for the fact that Socrates is wise, being either "The teacher of a wise man is also wise", or "The teacher of a prolific writer is wise". The theory can be formalised as:

I = $\emptyset$
T = {α_1: ∀x∀y teacher(y, x) ∧ wise(y) → wise(x),
α_2: teacher(Plato, Socrates),
α_3: wise(Plato),
α_4: prolificWriter(Plato),
α_5: ∀x∀y teacher(y, x) ∧ prolificWriter(y) ∧ philosopher(y) → wise(x),
α_6: philosopher(Plato) }
O = {wise(Socrates)}

As the theory itself is sufficient to account for the observations, we therefore do not require inputs in this example. However, we still consider support sets to be I-relative as we still consider the input set together with the theory to account for observations, and in this case the input set just happens to be empty. This formalised theory enables us to investigate the utility of formulas and sub-theories, the dependencies of one component of the theory to another, and the coupling between the components. Hence we find a measure for "usefulness" of components of the theory and how they are "tightly coupled".

2.5.1. Utility of sets

The two I-relative support sets for O are: $\{\alpha_1, \alpha_2, \alpha_3\}$, and $\{\alpha_2, \alpha_4, \alpha_5, \alpha_6\}$.

Hence the utility of formulas $\{\alpha_5, \alpha_6\}$ as a set would be $\frac{1}{2}$, since they appear together in only one of the two possible support sets; the set $\{\alpha_1, \alpha_4\}$ have the utility value of 0 since they do not work together at all; and the utility of $\{\alpha_2\}$ is 1 due to the fact that it appeared in all support sets.

2.5.2. Dependencies

Case 1: High Dependency
A formula α would have high dependency on a set Γ if $\{\alpha\} \cup \Gamma$ occurs in most support sets that contain α. So in the support sets and the theory illustrated above, the formula α_1 has a high dependency on both α_2 and α_3. Because without either formula, α_1 would not be able to account for the observation. In a theory where only one explanation is possible, the dependencies of all formulas in the support set relative to each other would be 1.

Case 2: Moderate/Low Dependency
A formula α would have a moderate/low dependency on a set Γ if α occurs in multiple support sets. This way the formulas in Γ may not always occur in support sets containing α. In the example above, α_2 has a moderate dependency on other formulas. This is because α_2 is contained in two support sets, and no other formula in T also occurs in the same two support sets. However, occurring in multiple support sets does not necessarily guarantee a moderate/low dependency to other formulas, for there could be another formula δ which occurs in the same support sets, thus having a high coupling coefficient. This is examined later in the section on couplings.

Case 3: Zero Dependency
Formulas will have zero dependency if they have nothing to do with each other. In the current theoretical context it means that they do not share

any support sets. Here axioms $\alpha_{1\ldots3}$ are totally disjointed from axioms $\alpha_{4\ldots6}$, thus any pairs selected with one from each set would yield zero dependence to each other.

2.5.3. Couplings

Case 1: High Coupling
High coupling occurs when two formulas (sets) often appear in the same support sets. In our example, α_5 and α_6 are required in the same support sets, since Plato needed to be both a prolific writer and a philosopher. Together, they have a coupling value of 1. However, this is different to dependency. If there were another formula α that also occurs across both support sets, then α and α_2 would have a high coupling value of 1 despite being spread across more than one support set.

Case 2: Moderate/low Coupling
Moderate/low coupling happens when two formulas (sets) appear in some support sets together, but in other support sets only one formula (set) is required. With our example, α_1 and α_2 have a coupling value of $\frac{1}{2}$. This value reflects the fact that both α_1 and α_2 appear in one support set, but only α_2 appears in the other support set. The coupling value between sets A and B is greater than 0 as long as they appear together in one support set. Formally:

$$CP(A, B, T, I, O) > 0 \text{ if and only if, for some } \Gamma \in S(T, I, O), A \cup B \subseteq \Gamma$$

Case 3: Zero Coupling
Like dependency, two formulas (sets) have zero coupling when they have nothing to do with each other; they do not ever work together to account for an observation. In our example, α_1 and α_5 have zero coupling.

2.6. Formulas with weights

Within a theory T, some axioms may be considered more important than others. This quality is described in the AGM framework [Gardenfors, 1988]. The importance of an axiom can either be innate, judgemental or could be determined from its usage in accounting for observations (its occurrence in support sets). Although some axioms are not frequently used, they may still be essential to the integrity of the theory. The measure of utility will be generalised to take into account an innate or judgemental weighing of axioms. In AGM entrenchment a logically weaker statement entailed by a stronger one will have an entrenchment at least as high as the latter. The analog of this for utility is the following: a weaker statement would account for at least as many input-output sets as a stronger one. This property is preserved by the definitions below of observational and natural weights. However, if weights are just subjective judgements then the analog of AGM entrenchment may not hold.

Definition 5 (Weight of a Formula). Let T be a finite theory $\{\alpha_1, \ldots, \alpha_n\}$, the weighing coefficient $W : T \mapsto \mathbf{R}$ is the subjective distribution of weights in T. $W(\alpha_i)$ reflects the innate weight of formula α_i.

However, it is possible to abuse the weighing process by arbitrarily adding weight to make the theory carry a high degree of coherence. Socrates may say: "My theory has half the coherence of your theory, so I just give each formula three times the weight, then mine would be more coherent!" To avoid this, and to make different theories comparable, weighing should be normalised in order to reflect the proportionate importance of formula α_i to the theory T.

Definition 6 (Normalisation Criterion). Let T be a finite theory $\{\alpha_1, \ldots, \alpha_n\}$, the normalisation criterion states that:

$$\sum_{i=1}^{n} W(\alpha_i) = n$$

Ontologically it does not make sense to give any formula a negative weight, for at worst it plays no part in support sets. Hence we assume:

Assumption 1 (Positivity Assumption). Let T be a finite theory $\{\alpha_1, \ldots, \alpha_n\}$

$$W(\alpha_i) > 0 \text{ for every } i : 1 \leq i \leq n$$

2.6.1. Observational Weights

Thus far, utility has been defined relative to an individual input set I and output set O. The pair (I, O) can be thought of as a single experiment or application of theory T. However, a theory is typically applicable and testable under many situations. It is therefore natural to consider what utility might mean across a vector of experiments or applications. Consider vectors (or sequences) of input and output sets, $\mathbf{I} = (I_1, I_2, \ldots, I_m)$ and $\mathbf{O} = (O_1, O_2, \ldots, O_m)$. One may interpret this vector as a sequence of experiments, e.g., a pair (I_k, O_k) being the k-th experiment with I_k being the initial conditions and O_k being the observation that results; other interpretations are of course possible, including O_k being an observation and I_k being the explanation. However, some observations may be considered more important than others, e.g., as in "crucial" experiments that may undermine a theory. To reflect this, experiments can be associated with a rank or weight that represents its judged significance. (In subsection 2.6.2 we propose a rather more objective assignment of weights.) The weight can then be "shared" by formulas that support this observation.

First we define a notion of *support weight* (SW). For each O_j in $\mathbf{O} = (O_1, \ldots, O_m)$, we associate a payoff $P(O_j)$. Then the support weight $SW(\alpha_i, T, I_j, O_j)$ can be the "share" of the payoff for α_i.

Definition 7 (Support Weight).

$$SW(\alpha_i, T, I_j, O_j) = \frac{P(O_j)}{\mid S(T, I_j, O_j) \mid} \sum_{\Gamma \in S(T, I_j, O_j) \text{ and } \alpha_i \in \Gamma} \frac{1}{\mid \Gamma \mid}$$

Hence from the support weight we define the observational weight that eventually reflects the importance of a formula.

Definition 8 (Observational Weight). For a theory T $= \{\alpha_1, \ldots, \alpha_n\}$, with input $\mathbf{I} : (I_1, \ldots, I_m)$ and output $\mathbf{O} : (O_1, \ldots, O_m)$, the Weight Share (WS) of axiom α_i in T is:

$$WS(\alpha_i) = \frac{1}{m} \sum_{j=1}^{m} SW(\alpha_i, T, I_j, O_j)$$

and the observational weight (OW) is:

$$OW(\alpha_i) = n \frac{WS(\alpha_i)}{\sum_{j=1}^{n} WS(\alpha_j)}$$

2.6.2. Natural Weights

By the original definition, weighing is a subjective measure of "importance" of formulas in a theory. However, it is possible to define a scheme of weighing from the dependency coefficient as defined before, since intuitively, if a formula is more needed by others, then it is more important.

For every formula α_i in T, we can define a *dependency weight* from how each formula in T depends on α_i. This represents an implicit weight of the specific formula.

Definition 9 (Dependency Weight). For an axiom α_i in a theory $T : \{\alpha_1, \ldots, \alpha_n\}$ with input $\mathbf{I} : (I_1, \ldots, I_m)$ and output $\mathbf{O} : (O_1, \ldots, O_m)$

$$DW(\alpha_i, T, \mathbf{I}, \mathbf{O}) = \sum_{j=1}^{n} \sum_{k=1}^{m} D(\alpha_j, \alpha_i, T, I_k, O_k)$$

Just as before, we could derive a measure of Natural Weight (NW) from the building block of dependency weights from the axioms.

Definition 10 (Natural Weight). So for a theory T $= \{\alpha_1, \dots, \alpha_n\}$ with input $\mathbf{I} : (I_1, \dots, I_m)$ and output $\mathbf{O} : (O_1, \dots, O_m)$

$$NW(\alpha_i) = n \frac{DW(\alpha_i, T, \mathbf{I}, \mathbf{O})}{\sum_{j=1}^{n} DW(\alpha_j, T, \mathbf{I}, \mathbf{O})}$$

In this way the ranking of axioms is accomplished by how other components of the theory depend on this component, and thus its weight is proportional to its importance in the theory. The advantage of this approach is that the weight of an axiom is no longer a subjective distribution given by the user, either by entrenchment or weights of observations. The natural weighing takes advantage of the natural properties of dependency, and thus weighing becomes an automated process.

2.7. Weighted Utility and Coherence

From this framework of weighted axioms, we can adopt a new and feature-rich definition of weighted utility. This makes utility useful in its own right, for we are able not only to compare between different theoretical systems, but components within a theoretical system. One possible application of this newly found role is in game theory, which shall be further investigated in this paper.

Definition 11 (Weighted Utility of a Formula). The *Weighted Utility* of a formula α in a theory T with respect to an input set I and an output set O, and a weight function W is:

$$WU(\alpha, T, I, O) = U(\alpha, T, I, O)W(\alpha)$$

Notice that we have introduced two weight functions: observational weight and natural weight. Observational weight is based on a subjective value placed on input/output (I, O) pairings while natural weight is based on dependency calculations. Both weighing functions are valid instances of W in the above definition.

The generalisation of coherence to weighted formulas will follow the intuition from [Kwok et al., 1998], as the average of weighted utilities.

Definition 12 (Coherence of a Weighted Theory).

$$C(T, \mathbf{I}, \mathbf{O}) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} WU(\alpha_i, T, I_j, O_j)$$

This culminating definition of coherence allow rival and possibly incompatible theories with weighted axioms to be evaluated and compared in a quantitative fashion. The evaluation is based on the brevity of the theory and the weighted utility of each of the theory components. It provides a perspective of how a theory can be judged based on inputs and observations, while taking into account the varying weights of different axioms in the theory.

2.8. Examples — Socrates is Wise 2

EXAMPLE 4 (Socrates is wise).

T = {α_1: ∀x∀y teacher(y, x) ∧ wise(y) → wise(x),
α_2: teacher(Plato, Socrates),
α_3: wise(Plato),
α_4: prolificWriter(Plato),
α_5: ∀x∀y teacher(y, x) ∧ prolificWriter(y) ∧ philosopher(y) → wise(x),
α_6: philosopher(Plato) }
O = {wise(Socrates)}

Recall the above "Socrates is Wise" example. It contains two support sets for the same observation. They are $\{\alpha_1, \alpha_2, \alpha_3\}$, and $\{\alpha_2, \alpha_4, \alpha_5, \alpha_6\}$. The first support contain three axioms, where the second contained four. In this example we denote $\{\alpha_1, \alpha_2, \alpha_3\}$ as the first support set and $\{\alpha_2, \alpha_4, \alpha_5, \alpha_6\}$ as the second support set. The common element is α_2, which is featured in both support sets. The other axioms would be called exclusive members of their support sets.

2.8.1. Example — Observational Weight

Suppose we assign the payoff of 100 points to the observation $wise(Socrates)$. Since both support sets adequately explain the observation, they deserve an equal share of the payoff, i.e., each support set will be apportioned 50 points. Each axiom that belong strictly to the first support set (of size 3) such as α_1 would receive an equal share of the payoff given to that support set, i.e., $\frac{1}{3} \times 50 = 16\frac{2}{3}$. Axiom α_4, belonging strictly to the second larger support set of size 4, would get a lesser share at $\frac{1}{4} \times 50 = 12\frac{1}{2}$.

Axiom α_2, contained in both support sets will have the greatest support weight at $\frac{1}{3} \times 50 + \frac{1}{4} \times 50 = 29\frac{1}{6}$. Hence its observational weight would be $\frac{6}{100} \times 29\frac{1}{6} = 1\frac{3}{4}$, which is also its weighted utility, since it appears in all support sets of O. The utility of the other formulas would be half of their observational weight, since there is only one observation and they belong strictly to one of the two support sets. It would be $\frac{1}{2}$ for the exclusive members of the first support set and $\frac{3}{8}$ for exclusive members of the second support set. Hence the weighted coherence value would be $\frac{31}{48}$. This value reflects the degree of coherence of the given theory with respect to a set of observations with weights.

2.8.2. Example — Natural Weight

For the given support set, the dependency weight of the exclusive member of support sets would be $2\frac{1}{2}$ and $3\frac{1}{2}$ respectively. Since this definition values the support from other axioms, the members of the larger support set would receive more weight. The common element α_2 would receive a dependency weight of 6.

Therefore the natural weight of α_2 would be $\frac{72}{43}$. The exclusive members of the first support set would receive a natural weight of $\frac{30}{43}$, and the second support set $\frac{42}{43}$. This is also their utility value since there is only one observation. The weighted coherence value is $\frac{55}{86}$. This value reflects the degree of coherence with respect to the internal structure of the theory, thus the value is different to that derived from observational weights. We consider both to be valid, but different measures of coherence. The user would make the choice in selecting which measure to use depending on its applications.

Further examples of this new weighted system of coherence, particularly in observational weights, are illustrated in the following section with an application in Game Theory.

3. Application to Game Theory

3.1. Concept

Coherence, once quantified, can be used as a comparator between any two theoretical systems. In typical agent interactions, all of an agent's beliefs, desires and intentions (BDI) can be represented in formal semantics [Rao and Georgeff, 1991]. These enable us to assess a systemic coherence in one's belief, and the process of interaction can be seen as an effort by each agent to modify its own system in order to achieve a satisfactory outcome with respect to the other agents while maintaining a high level of its internal coherence. The intuition is that the agent will choose an action that is most coherent with its set of beliefs, desires and intentions.

3.2. Prisoner's Dilemma Simulation

3.2.1. Background

The Prisoner's Dilemma was originally formulated by mathematician Albert W. Tucker. The iterated version of the game was proposed in [Axelrod, 1981]. It has since become the classic example of a "non-zero sum" game in economics, political science, evolutionary biology, and of course game theory. So that the exposition below may be independently understood, we briefly recount the set-up. In the game, two prisoners are interrogated separately in different cells. The two prisoners can either choose to cooperate (keep silent) or defect (blame the other). If they both cooperate, they receive a sentence of 2 years in prison. If one cooperates but the other betrays, the first gets 10 years in prison, and the second gets 1 year. If both betray, each will get 4 years. The payoff (years in prison) of an action is dependent on the action of the other player. It is therefore in the interests of a player to minimise this payoff. The way the payoff is set out means that whatever a player chooses to do, the other player can reduce its payoff by defecting, so in a one-time game both players will defect, resulting in 4 years for each. A better result will be for both to cooperate, suffering a sentence of only 2 years each; but they cannot communicate to negotiate, and even if they can, lack of trust may enter the picture. This "bad" solution of both defecting can intuitively be ameliorated if the game is played repeatedly, whence each player understands that if it defects now the other player can retaliate in the next iteration. Thus, in the iterated version, the players repeatedly play the game and have a memory of their previous encounters. We set out to test the application of our coherence calculations in this scenario, and how it behaves in an iterated game with evolution of populations.

3.2.2. The Coherentist Agent

In coherence-based evaluative simulations in game theory, we set out to play the game repeatedly, and the histories of past games are recorded by each player. This history then forms the Belief in what had happened in the past, which can be seen as the theory T in the calculation. The player's Desire (D) is to maximise its payoff (or minimise it if interpreted as a penalty). This desire can be seen as a mode of evaluating payoff as weights of each outcome. The beliefs (B), together with the criterion of selection (D), will lead to the calculation of the utility of each of the actions that the player may take. A selection of the action according to its utility will lead the player to formulate the intention to act upon this decision.

More specifically, the inputs (I) describes the rules of what the player knows about the nature of the game. This includes actions of the player, consequences of these actions, states (in a finite-state game) and payoffs associated with a particular state. The observations (O), whether a result, consequence or state, is the corresponding payoff. The desire of the player will be driven by the ranking of these payoffs. Hence the Support Set consists of the list of axioms, which together with the given input, will make a particular observation true.

In the iterated game of Prisoner's Dilemma, the prisoner evaluates the history played against the respective player to reach a rational decision. Each history element consists of the player's move at that iteration, and the returned value / payoff from that particular move. The returned value can be seen as the weight of that observation, and hence the support weight for that particular action.

The coherentist agent uses the paradigm of the observational weight as discussed in Section 2.6.1. This way the weight of a formula is reflected by the observations that it supports. The weight of an action can be evaluated from the history of payoffs for a given opponent.

The Input (I) for a particular iteration are the rules of the game, and the move of other players. Below is a summary of the rules, expressed logically. The propositions *Betray* and *otherBetray* mean respectively that a player betrays and the other also betrays; *Cooperate* and *otherCooperate* have corresponding meanings. The numbers are the payoffs for a player, depending on the move of the other player; recall that these are the years in prison, and hence a penalty to be minimised.

$$I = \left\{ \begin{array}{l} \mathit{Betray} \wedge \mathit{otherBetray} \rightarrow 4, \\ \mathit{Betray} \wedge \mathit{otherCooperate} \rightarrow 1, \\ \mathit{Cooperate} \wedge \mathit{otherBetray} \rightarrow 10, \\ \mathit{Cooperate} \wedge \mathit{otherCooperate} \rightarrow 2 \end{array} \right\}$$

This input set will remain fixed for each game of the Prisoner's Dilemma.

For each move of a player the other player has the choices *otherBetray* or *otherCooperate*. The theory, to be evaluated, are the rival options the player could adopt. viz., *Betray* or *Cooperate*, bearing in mind that in any iteration the moves of both players are to be made *simultaneously*.

$$T = \{\mathit{Betray}, \mathit{Cooperate}\}$$

For instance, consider a history (Action, Penalty) of (Cooperate, 2), (Betray, 1), (Betray, 4), (Betray, 4). This implies that at the same time the other player had made the corresponding choices of *otherCooperate*, *otherCooperate*, *otherBetray* and *otherBetray*. Hence the sequence of output observation set is:

$$\begin{array}{rcl} \mathbf{O} & = & (O_1 : \{\mathit{otherCooperate} \wedge 2\}, \\ & & O_2 : \{\mathit{otherCooperate} \wedge 1\}, \end{array}$$

$$O_3 : \{otherBetray \wedge 4\},$$
$$O_4 : \{otherBetray \wedge 4\})$$

In section 2.7 we associated a payoff to each output set. This payoff was a measure of the importance of the output set. For our application to the Prisoner's Dilemma, we wish to measure how advantageous each output is to an agent. This would be inversely proportional to the prison sentence. In our simulation studies, we simply used the length of the prison sentence as the payoff and chose the option with the *smaller* Observational Weight. With the above example, the Weight Share of Betray is $(1+4+4\ )\div 3 = 3$, whereas the Weight Share of Cooperate is $2 \div 1 = 2$. Hence the Observational Weight is evaluated at $\frac{3}{4}$ for Betray, and $\frac{2}{4}$ for Cooperate. Therefore, in a system where lower weight (penalty) is favoured, Cooperate is the preferred strategy. This was the approach adopted in the experiments. However, for the analogous approach where the system favours higher weights, the payoff can be taken as the inverse of the penalty. Therefore the agent's choice of an axiom of higher weighted utility reflects its pursuit of a higher degree of coherence.

The problem of course is that it is difficult to predict what the other player will do at any iteration. In the tournament organised by Axelrod [Axelrod, 1981] the system pitted many players together and simulated the iterations, looking for the best performing players. In the simulations we ran, we investigated how coherentist players performed against other kinds of players, including the best performing player in Axelrod's tournaments.

3.2.3. Simulation

We define five types of agents in the simulation. They are *reckless*, *cooperative*, *tit-for-tat*, *suspicious* and *trusting*. The last two types are the same coherentist agent with different initial conditions. A reckless player is one who always defects, whereas the cooperative player is one who always cooperates (does not defect). The "tit-for-tat" strategy was traditionally regarded as the best deterministic strategy developed by Anatol Rapoport, which cooperates in the first turn, and subsequently plays the opposing player's previous move. The coherentist agents are divided into two groups, one being suspicious, for its members would betray at the initial phase, whereas the other group, the trusting agents, would cooperate.

To initialise simulation, the user specifies how many of each type of agent there are in the game. The user also specifies how many iterations are to be simulated. In each iteration a player will play a round-robin tournament, playing once with every other player in the simulation. When two players meet, they have the option to either betray or cooperate. The move and the payoff will be recorded, and the agent can review this as a history element when playing this opponent in the next iteration.

After a specified number of iterations, the old players will die and a new generation of players will replace them. They will be free of the history from previous players. However, their proportions, according to agent type, will be inversely proportional to the average time the particular type of agent spent in jail. The result is then normalised to maintain the population size. Although rounding error is allowed, the overall population size will only decrease due to rounding, and increases are prohibited.

3.2.4. Trends and Behaviours

Initially, we set 20 iterations per generation with an equal proportion of each player type. It turns out that the coherentist player performs well compared to other agent types. As predicted the cooperative agents perish rather quickly in the simulation. In the end it was the "suspicious" coherentist agents that took over the population, while others struggled to hang on. (Figure 1)

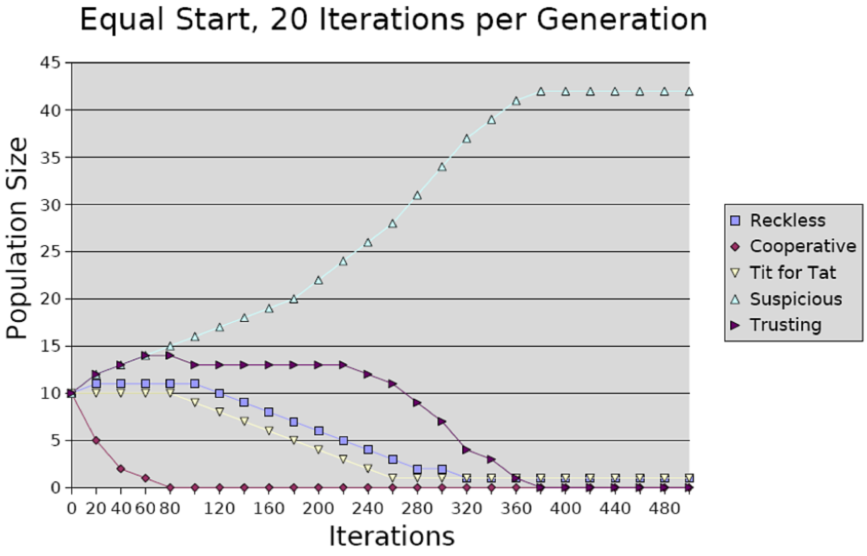


Figure 1. All five players with equal initial population, 20 iterations per generation

The suspicious and trusting agents only differ in their initial response when they have no previous history of playing against the other player. Yet the impact is significant as the suspicious player takes over the population after a brief initial period when both coherentist agents perform well. Both the coherentist agents gain an edge over tit-for-tat, as they exploit the cooperative agents while tit-for-tat is only nice to them.

When the number of iterations per generation is raised from 20 to 50, the results are slightly different. One feature is that the reckless agents performed much more badly, while the tit-for-tat agents played better, though not as well as the coherentist agents. (Figure 2)

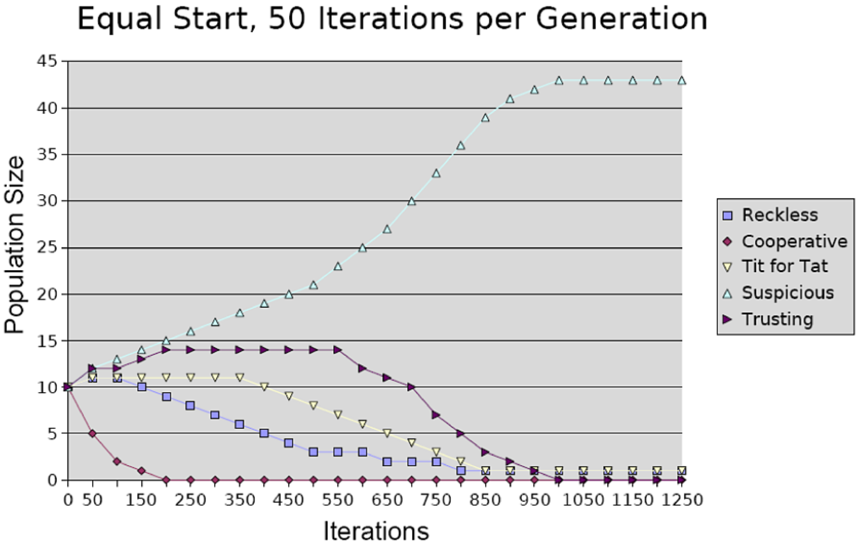


Figure 2. All five players with equal initial population, 50 iterations per generation

In both simulations the coherentist agents came out on top. This may be associated with the coherentist agent's flexible approach of punish reckless behaviour, cooperate with rational, nice agents, and exploit the overly nice and vulnerable agents. In particular the latter characteristic is absent in the behaviour of tit-for-tat agents. However, this positive outcome may not necessarily be associated with simply a coherentist behaviour. Instead, it may be the case that the macro-environment of the game in this situation enabled the coherentist agents to be the fittest. For a different environment with different rules, coherentist agents may not perform as well as agents of the "simple faith", such as the reckless or tit-for-tat agents.

What emerges from these results is that coherence alone as a property of agents is an aid to their performance, but external factors and initial conditions (such as the first move) also matter. A way to think of the role of coherence is that it constrains agent choices in such a way that its use of its theory aligns those choices well with its observations.

4. Summary and Discussion

We aim to establish basic principles governing the coherence of laws within theoretical systems. Such principles provide a means for evaluating and comparing different systems. By defining a measure of how a sub-theory contributes to a theory, in terms of *Group Utility*, *Dependency* and *Coupling*, the formalism captures a number of important properties of coherence. Specifically, the formalism provides a rendering of informal characteristics of coherence, *viz.* how axioms "work together" and are "coupled tighter". The framework has also been significantly enhanced by the introduction of weights to axioms and observations. By relativising one axiom's weight, either in terms of the weight of observations or the dependency to other axioms, we derived an account of the importance and rank of axioms in a theory.

Our proposed framework of coherence serves as a useful treatment of an old problem in the philosophy of science, namely the evaluation of rival, but possibly incompatible theories. It also provides a perspective on the development of scientific theories, where anomalies found in observations contribute to the degree of incoherence of a theory, and scientific developments to account for these anomalies can be viewed as the pursuit of a greater coherence.

This measure of coherence is not only useful for the domain of the philosophy of science, it is also useful for describing reasoning, deliberation and interaction in agents. The example of Prisoner's Dilemma illustrated how coherence can be used in game theory. When an agent chooses the option that is most coherent with its beliefs, the agent has a rational basis for reasoning and acting.

Acknowledgements. We would like to thank Anika Schumann, Anthony J. Flynn and the three anonymous reviewers for their proof-reading and their comments on the paper, and National ICT Australia for supporting this research under its knowledge representation and reasoning program.

References

Axelrod, R. (1981), 'The evolution of cooperation', *Science*, 211(4489): 1390–1396.

Craig, W. (1953), 'On axiomatizability within a system', *The Journal of Symbolic Logic*, 18: 30–32.

van Fraassen, B. (1980), *The Scientific Image*, Clarendon Press, Oxford, pp. 14–19.

Gardenfors, P. (1988), *Knowledge In Flux*, MIT Press, Cambridge, MA.

Hawking, S.W. (2001), *The Universe in a Nutshell*, Bantam Books, New York.

Kwok, R.B.H., Nayak, A.C., Foo, N. (1998), 'Coherence measure based on average use of formulas', in *Proceedings of the Fifth Pacific Rim Conference on Artificial Intelligence*, LNCS, vol. 1531, Springer Verlag, pp. 553–564.

Kwok, R.B.H., Foo, N., Nayak, A.C. (2003), 'The coherence of theories', in *Proceedings of the 18th Joint International Conference on Artificial Intelligence*, IJCAI03, Acapulco, Mexico, August 2003.

Kwok, R.B.H., Foo, N., Nayak, A.C. (2007), 'Coherence of laws', *UNSW Computer Science and Engineering Technical Report*, Number: UNSW-CSE-TR-0719, October 2007. ftp://ftp.cse.unsw.edu.au/pub/doc/papers/UNSW/0719.pdf.

Nagel, E. (1961), *Structure of Science*, Harcourt.

Rao, A.S., Georgeff, M.P. (1991), 'Modeling rational agents within a BDI-architecture', in *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pp. 473–484.

Jason Jingshi Li
Research School of Information Science and Engineering
Australian National University
Canberra, ACT0200, Australia
`jason.li@anu.edu.au`

Rex Bing Hung Kwok
School of Computer Science and Engineering
University of New South Wales
Sydney, NSW2032, Australia
`rkwok@cse.unsw.edu.au`

Norman Y. Foo
School of Computer Science and Engineering
University of New South Wales
Sydney, NSW2032, Australia
`norman@cse.unsw.edu.au`

Urszula Wybraniec-Skardowska

On Meta-knowledge and Truth

Abstract. The paper deals with the problem of logical adequacy of language knowledge with cognition of reality. A logical explication of the concept of language knowledge conceived of as a kind of codified knowledge is taken into account in the paper. Formal considerations regarding the notions of meta-knowledge (logical knowledge about language knowledge) and truth are developed in the spirit of some ideas presented in the author's earlier papers (1991, 1998, 2001a,b, 2007a,b,c) treating about the notions of *meaning*, *denotation* and *truthfulness* of well-formed expressions (*wfes*) of any given categorial language. Three aspects connected with knowledge codified in language are considered, including: 1) syntax and two kinds of semantics: *intensional* and *extensional*, 2) three kinds of non-standard language models and 3) three notions of truthfulness of *wfes*. Adequacy of language knowledge to cognitive objects is understood as an agreement of truthfulness of sentences in these three models.

Keywords: Meta-knowledge, categorial syntax, meaning, denotation, categorial semantics, non-standard models, truthfulness, language knowledge adequacy.

Introduction

It is commonly realized that the term 'knowledge' is ambiguous. Speaking about knowledge, we disregard psychological knowledge offered through unit cognition, although it is from knowledge of that sort that verbal knowledge codified by means of language arose. Knowledge will be understood as an inter-subjective knowledge preserved in language, where it is formed and transferred to others in cognitive-communicative acts. Representation of this knowledge is regarded as language knowledge.

For our purposes, in this paper we will consider three aspects of language knowledge: one syntactic and two semantic ones: *intensional* and *extensional*. The main aim of the paper is to answer the following well-known, classical philosophical problem:

`When is our language knowledge in agreement with our cognition of reality?`

D. Makinson, J. Malinowski, H. Wansing (eds.), *Towards Mathematical Philosophy*,
Trends in Logic, 28

In this paper, the problem is considered from a logical and mathematical perspective and is called: *the problem of logical adequacy of language knowledge.* We will consider it as:

1) an adequacy of syntax and two kinds of semantics,
2) concord between syntactic forms of language expressions and their two correlates: meanings and denotations, and
3) an agreement of three notions of truth: one syntactic and two semantic ones.

The main ideas of our approach to meta-knowledge (logical knowledge about language knowledge) and truthfulness of sentences in which knowledge is encoded will be outlined in Section 1. In Section 2 we will give the main assumptions of a formal-logical theory of syntax and semantics which are the basis for theoretical considerations, and in Section 3 we will define three notions of truthfulness of sentences. The paper ends with Section 4 containing a formulation of a general condition for adequacy of language knowledge with regard to these notions.

⋆⋆⋆

The paper is a result of many years of research conducted by the author and a summary of results obtained earlier [47–58]. The synthetic character of the article provides a strong motivation for the conceptual apparatus introduced further. The apparatus employs some formal-logical and mathematical tools. The synthesis being produced does not always allow detailed, verbal descriptions of particular formal fragments of the paper; nor can it allow for development of some formal parts. The author does, however, believe that the principal ideas and considerations in the paper will be clear to the reader.

1. Ideas

The notion of meta-knowledge is connected with the relationships defined by the triad: language-cognition-reality (see *Figure 1*).

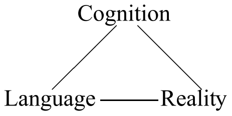


Figure 1.

Three different aspects, representing cognitively independent factors, are taken into account at constituting any language L as a tool of communication in which knowledge is formed and transmitted. They are: syntactic, semantic and pragmatic factors.

Reliability of cognition of reality by means of language L and truthfulness of its sentences are given by an agreement of syntactic and two kinds of (*intensional* and *extensional*) semantic knowledge, which correspond to three levels of knowledge about the components of the triad (cf. Wybraniec-Skardowska 2007c).

According to *Figure 1*, following Frege [17], Husserl [25] and other modern followers of *gramatica speculativa*, the meta-knowledge is the *knowledge referring to three realities* (*spaces*):

1. *language reality* **S** (the set of all *well-formed expressions* of L), in which results of cognitive activities such as concepts and propositions are expressed,
2. *conceptual reality* **C**, in which products of cognition of ordinary reality such as logical concepts and logical propositions (*meanings* of language expressions) are considered, and
3. *ontological reality* **O** which contains objects of cognition, among others, *denotations* of language expressions.

Applying the terms: 'language reality', 'conceptual reality' and 'ontological reality' we aim at distinguishing some models of language L which are necessary to define three different notions of truthfulness of its sentences. Thus, we depart from the classical notion of 'Reality' as an object of cognitive research. In particular, speaking further about indexation reality **I**, we mean certain metalinguistic space of objects (indices) serving the purpose of indication of categories of expressions of **S**, categories of conceptual objects of **C** and ontological categories of objects of **O**. The reality **I** forms categorial skeleton of language, conceptual and ontological realities.

Theoretical considerations are based on:

- **syntax** – describing **language reality S** related to L,
  and two kinds of semantics:
- **intensional** (**conceptual**) **semantics** – comprising the relationship between **S** and cognition – describing **conceptual reality C**, and
- **extensional** (**denotational**) **semantics** – describing the relationships between L and ordinary reality – **ontological reality O** to which the language refers (see Wybraniec-Skardowska 1991, 1998, 2007a, b, c).

The theoretical considerations take into account the **adequacy of the syntax and two kinds of semantics** of language L.

The **language reality S** is described by a theory of categorial syntax and the **conceptual** and **ontological realities** by its expansion to a theory of categorial semantics in which we can consider three kinds of **models** *of* L:

- one syntactic
  and
- two semantic (**intensional** and **extensional**).

For these models we can define three notions of **truthfulness**:

- one syntactic
  and
- two semantic employing the notion of *meaning* (*intension*) and the notion of *denotation* (*extension*), respectively.

2. Main Assumptions of the Theory of Syntax and Semantics

2.1. Categorial Syntax and Categorial Semantics

Any syntactically characterized language L is fixed if the *set* **S** *of all well-formed expressions* (briefly *wfes*) is determined. L is given here on the *type*-level, where all *wfes* of **S** are treated as expression-*types*, i.e. some classes of concrete, material, physical, *identifiable* expression-*tokens* used in definite linguistic-situational contexts. Hence, *wefs* of **S** are here abstract ideal syntactic units of L[1].

Language L can be exactly defined as a *categorial language*, i.e. language in which *wfes* are generated by a *categorial grammar* whose idea goes back to Ajdukiewicz (1935) and Polish tradition, and has a very long history[2]. Language L at the same time may be regarded as a linguistic scheme of

---

[1] Let us note that the differentiation *token-type* for linguistic objects originates from Charles Sanders Peirce (1931-1935). A formal theory of syntax based on this distinction is given in [49] and [51].

[2] The notion of categorial grammar originated from Ajdukiewicz (1935, 1960) was shaped by Bar-Hillel (1950, 1953, 1964). It was constructed under influence of Leśniewski's theory of semantic (syntactic) categories in his protothetics and ontology systems (1929, 1930), under Husserl's ideas of pure grammar (1900-1901), and under the influence of Russell's theory of logical types. The notion was considered by many authors: Lambek (1958, 1961), Montague (1970, 1974), Cresswell (1973, 1977), Buszkowski (1988, 1989), Marciszewski (1988), Simons (1989) and others. In this paper language L is generated by the so-called classical categorial grammar, the notion introduced and explicated by Buszkowski (1988, 1989) and the author (1985, 1989, 1991).

ontological reality **O**, keeping with Frege's ontological canons (1884), and of conceptual reality **C**.

Considerations are formalized on the ground of author's general formal-logical theory of **categorial syntax** and **categorial semantics** (1985, 1991, 1998, 1999, 2001a,b, 2006).

Every compound expression of L has a **functor-argument structure** and both it and its constituents (the main part – the main **functor** and its complementary parts – **arguments of that functor**) have determined:

- the *syntactic*, the *conceptual* and the *ontological categories* defined by the *functions* ι_L, ι_C, ι_O *of the indications of categorial indices* assigned to them, respectively,
- *meanings* (*intensions*), defined by the *meaning operation* μ,
- *denotations* (*extensions*), defined by the *denotation operation* δ.

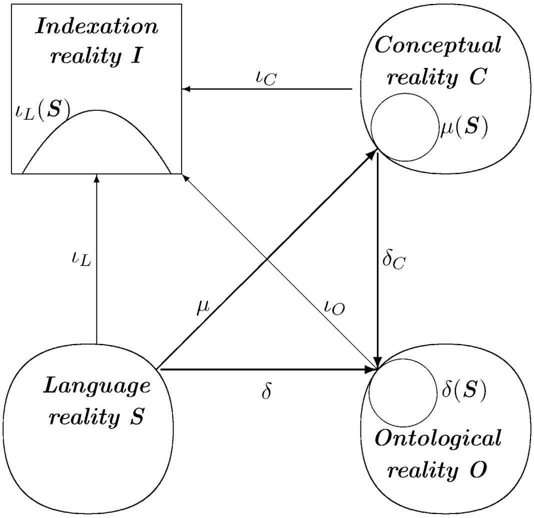


Figure 2.

It should be underlined that since *wefs* of **S** are understood as some abstract syntactic units of L, meanings of *wfes* are not their mental signification and denotations of *wfes* are not the same as object references of their concrete, material expression-*tokens* (cf. [57]).

2.2. Three referential relationships of wfes

We will concentrate on three referential relationships of *wfes of* **S** to three realities to which *wfes* refer:

- one syntactic: *metalinguistic* relationship connected with the above-mentioned **indexation reality I**, and
- two semantic: **conceptual** (**intensional**) and **denotational** (**extensional**) relationships connected with realities **C** and **O**, respectively. These relationships are illustrated in *Figure 2*.

2.3. Categorial indices

The theory of categorial syntax is a theory formalising the basic principles of Leśniewski's theory of semantic (syntactic) categories (1929, 1930) improved by Ajdukiewicz (1935) by introducing *categorial indices* assigned to expressions of language *L*.

Categorial indices belong to the indexation reality **I** and are metalanguage expressions corresponding to expressions of language *L*. They serve to defining the set **S** of all *wfes* of *L*. The set **S** is defined according to the *principle* (*SC*) *of syntactic connection* referring to Ajdukiewicz's approach (1935).

(*SC*) is the rule establishing the correspondence between the index of any functor-argument expression of *L* and indices: the index of its main functor and indices of its successive arguments. It states that:

(*SC*) *The index of the main functor of a functor-argument expression is a complex (functoral) index formed of the index of that expression and the successive indices of the successive arguments of that functor.*

2.4. Syntactic Operations

In the theory the *functions:* ι_L, ι_C, ι_O *of the indications of categorial indices* are certain syntactic operations from reality **S** or fragments of realties **C** and **O** into reality **I**, respectively, i.e.

- the syntactic operation $\iota_L\colon \mathbf{S}\rightarrow\mathbf{I}$,
- the ontological syntactic partial operation $\iota_O\colon \mathbf{O}\rightarrow\mathbf{I}$,
- the conceptual syntactic partial operation $\iota_C\colon \mathbf{C}\rightarrow\mathbf{I}$.

Categorial indices of **I** also serve to indicate *syntactic, conceptual* (*intensional*) and *ontological* (*denotational*) *categories*. These categories are included in realities **S**, **C** and **O**, respectively.

If $\xi \in \mathbf{I}$ then these categories are defined, respectively, as follows:

$$Cat_\xi = \{e \in \mathbf{S} \colon \iota_L(e) = \xi\}, \tag{1}$$

$$Con_\xi = \{c \in \mathbf{C} \colon \iota_C(c) = \xi\}, \tag{2}$$

$$Ont_\xi = \{o \in \mathbf{O} \colon \iota_O(o) = \xi\}. \tag{3}$$

In order to define *semantic categories* indicated by categorial indices, and also by conceptual and ontological categories, we have to take into consideration two semantic relationships and use some semantic operations.

2.5. Semantic Operations

In the theory of categorial semantics such notions as **meaning** and **denotation** of a *wfe* of L are considered.

As it was illustrated in *Figure 2* we consider three semantic operations defining **meanings** and **denotations** *of wfes*:

- the *meaning operation* $\mu \colon \mathbf{S} \to \mathbf{C}$,
- the *denotation operation* $\delta \colon \mathbf{S} \to \mathbf{O}$,
- the *conceptual denotation operation* $\delta_C \colon \mathbf{C} \to \mathbf{O}$.

Let us note that the semantic functions: μ, δ and δ_C, are defined on abstract objects of **S** (on *wfes-types*) and of **C** (on meanings: logical concepts, logical propositions, operations on them, operations on these operations and so on), respectively.

The notion of **meaning** as a value of the *meaning operation* μ on any *wfe* of L is a semantic-pragmatic one and it is defined as a manner of using *wfes* of L by its users in connection to the concept of meaning deriving from L. Wittgenstein (1953) and, independently, from K. Ajdukiewicz (1931, 1934); see Wybraniec-Skardowska (2005, 2007 a,b). So, the notion of meaning of any *wfe* of L is an abstract entity.

We take the standpoint that any *wfe-type* of **S** has an established meaning which determines its denotation, even if such an expression is understood as an indexical one in natural language (e.g. 'he', 'this', 'today')[3]. In this sense

---

[3] For example, let us note that the word-*type* 'today' understood as a class of all word-*tokens* identifiable with the word-*token*:

today

does not have a fixed meaning, but each of its sub-*types* consisting of *identifiable* tokens (utterances) of the word-*type* 'today' formulated on a given day is a meaningful *wfe-type* of English and determines by itself a denotation that is this day.

the approach presented here agrees with the classical Aristotelian position that the context has to be included somehow in the meaning; the manner of using *wfes* of L is in a way built into the meaning (cf. [57]).

The notion of **meaning** is differentiated from the notion of **denotation** in accordance with the distinction of G. Frege (1892) *Sinn* and *Bedeutung* and R. Carnap's distinction *intension-extension* (1947).

The *denotation operation* δ is defined as the composition of the operation μ and the *operation* δ_C *of conceptual denotation*, i.e.

$$(\delta_C) \qquad \delta(e) = \delta_C(\mu(e)) \quad \textit{for any } e \in \mathbf{S}.$$

So, we assume that denotation of the *wfe* e is determined by its meaning $\mu(e)$ and it is the value of the function δ_C of conceptual denotation for $\mu(e)$. Hence, we can state that:

If two wfes have the same meaning then they have the same denotation.

Formally:

Fact 1. $\mu(e) = \mu(e') \Rightarrow \delta(e) = \delta(e')$, *for any* $e, e' \in \mathbf{S}$.

It is well-known that the converse implication does not hold. So, the operation δ_C shows that something can differ **meaning** from **denotation**.

2.6. Knowledge and Cognitive Objects

The image $\mu(\mathbf{S})$ of $\mathbf{S}$ determined by the meaning operation μ is a fragment of conceptual reality $\mathbf{C}$ and includes all meanings of *wfes* of language L, so all components of **knowledge** (logical notions, logical propositions and operations between them, operations on the latter, and so on) and can be regarded as **knowledge** of relatively stable users of L about reality $\mathbf{O}$, codified by means of *wfes* of L.

The image $\delta(\mathbf{S})$ of $\mathbf{S}$ determined by the denotation operation δ is a fragment of ontological reality $\mathbf{O}$ and includes all denotations of *wfes* of language L, so all **objects of cognition** of $\mathbf{O}$ (things, states of things and operations between them) in cognitive-communicative process of cognition of reality $\mathbf{O}$ by relatively stable users of L.

We differentiate two kinds of semantic categories: intensional and extensional.

$$(4) \qquad Int_\xi = \{e \in \mathbf{S} \colon \mu(e) \in Con_\xi\}.$$

$$(5) \qquad Ext_\xi = \{e \in \mathbf{S} \colon \delta(e) \in Ont_\xi\}.$$

So, intensional categories consist of all *wfes* whose meanings belong to suitable conceptual categories, while extensional categories consist of all *wfes* whose denotations belong to suitable conceptual categories.

Adequacy of syntax and semantics required the syntactic and semantic agreement of *wfes* of L.

2.7. The principles of categorial agreement

In accordance with Frege's-Husserl's-Leśniewski's and Suszko's understanding of the *adequacy of syntax and semantics of language L*, syntactic and semantic (*intensional and extensional*) categories with the same index should be the same (see Frege, 1879, 1892; Husserl, 1900-1901; Leśniewski, 1929, 1930; Suszko, 1958, 1960, 1964, 1968).

This correspondence of the **categorial agreement** (denoted by (*CA1*) and (*CA2*)) – is here postulated by means of categorial indices that are the tool of coordination of language expressions and by two kinds of references that are assigned to them:

(*CA1*) $Cat_\xi = Int_\xi$.

(*CA2*) $Cat_\xi = Ext_\xi$.

From (1)–(5) and (*CA1*), (*CA2*) we get the following variants of the principles:

For any *wfe* e

(*C'A1*) $e \in Cat_\xi$ iff $\mu(e) \in Con_\xi$.

(*C'A2*) $e \in Cat_\xi$ iff $\delta(e) \in Ont_\xi$.

(*CA3*) $\iota_L(e) = \iota_C(\mu(e)) = \iota_O(\delta(e))$.

The condition (*C'A2*) is called the **principle of categorial agreement** and it is a formal notation of the principle originated by Suszko (1958, 1960, 1964; cf. also Stanosz and Nowaczyk 1976).

So, according to innovative Frege's ideas, the problem of adequacy of syntax and semantics of L is solved if:

Well formed expressions of L belonging to the same syntactic category correspond with their denotations, and more generally – with their two kinds of references (meanings and denotations) that are assigned to them, which belong to the same ontological, and more generally – to the same conceptual and ontological category.

2.8. Algebraic structures of categorial language and its correlates

The essence of the approach proposed here is considering functors of language expressions of L as mathematical functions mapping some language expressions of **S** into language expressions of **S** and as functions which correspond to some set-theoretical functions on extralinguistic objects – indices, meanings and denotations of arguments of these functors.

All functors of L create the set $\mathbf{F}$ included in $\mathbf{S}$.

The systems:

$$\mathbf{L} = \langle \mathbf{S}, \mathbf{F} \rangle \quad \text{and} \quad \iota_L(\mathbf{L}) = \langle \iota_L(\mathbf{S}), \iota_L(\mathbf{F}) \rangle$$

are treated as some syntactic algebraic structures, while the systems:

$$\mu(\mathbf{L}) = \langle \mu(\mathbf{S}), \mu(\mathbf{F}) \rangle \quad \text{and} \quad \delta(\mathbf{L}) = \langle \delta(\mathbf{S}), \delta(\mathbf{F}) \rangle$$

can be treated as some semantic algebras.

All these algebras are partial algebras[4].

The functors of $\mathbf{F}$ differ from other, basic expressions of $\mathbf{S}$ in that they have indices formed from simpler ones.

If e is a complex functor-argument *wfe* with the index a and its main functor is $f \in \mathbf{F}$ and its successive arguments are $e1, e2, \ldots, en$ with indices a_1, a_2, …, a_n, respectively, then the index b of f belonging to the set $\iota_L(\mathbf{F})$ is a *functoral* (complex) index formed from the index a and indices: a_1, a_2, …, a_n of its successive arguments.

The index b of the functor f can be noted as the quasi-fraction:

$$\iota_L(f) = b = a/a_1a_2\ldots a_n = \iota_L(e)/\iota_L(e1)\iota_L(e2)\ldots\iota_L(en).$$

We will show that indices, meanings and denotations of functors of the set $\mathbf{F}$ are algebraic, partial functions defined on images $\iota_L(\mathbf{S})$, $\mu(\mathbf{S})$, $\delta(\mathbf{S})$ of the set $\mathbf{S}$, respectively.

First we will note that in accordance with the principle (*SC*) the main functor f of e can be treated as a set-theoretical function satisfying the following formula:

(*Catf*) $f \in Cat_{a/a_1a_2\ldots a_n}$ iff

(*f*) $f\colon Cat_{a_1} \times Cat_{a_2} \times \cdots \times Cat_{a_n} \to Cat_a \;\&\; e = f(e1, e2, \ldots, en) \;\&$

(ι) $\iota_L(f)\colon \{(\iota_L(e1), \iota_L(e2), \ldots, \iota_L(en))\} \to \{\iota_L(e)\} \;\&$

(*PC1*) $\iota_L(e) = \iota_L(f(e1, e2, \ldots, en)) = \iota_L(f)(\iota_L(e1), \iota_L(e2), \ldots, \iota_L(en))$.

---

[4] Ideas about the algebraisation of language can already be found in Leibniz's papers. We can also find the algebraic approach to issues connected with syntax, semantics and compositionality in Montague's 'Universal Grammar' (1970) and in the papers of van Benthem (1980, 1981, 1984, 1986), Janssen (1996), Hendriks (2000). The difference between their approaches and the approach which we shall present here lies in the fact that carriers of the so-called *syntactic* and *semantic algebras* discussed in this paper include functors or, respectively, their suitable correlates, i.e. their ι_L- or some other semantic-function images. Simple functors and their suitable ι_L-, $\mu-$ or $\delta-$ images are partial operations of these algebras. They are set-theoretical functions determining these operations.

On the basis of the principles of categorial agreement we can state that semantic correlates of the functor f of the expression e are set-theoretical functions too, and deduce that they satisfy the following conditions:

$(Conf)$ $\mu(f) \in Con_{a/a_1 a_2 \ldots a_n}$ iff
(μ) $\mu(f)\colon Con_{a_1} \times Con_{a_2} \times \cdots \times Con_{a_n} \to Con_a$ &
$(PC2)$ $\mu(e) = \mu(f(e1, e2, \ldots, en)) = \mu(f)(\mu(e1), \mu(e2), \ldots, \mu(en))$;

$(Ontf)$ $\delta(f) \in Ont_{a/a_1 a_2 \ldots a_n}$ iff
(δ) $\delta(f)\colon Ont_{a_1} \times Ont_{a_2} \times \cdots \times Ont_{a_n} \to Ont_a$ &
$(PC3)$ $\delta(e) = \delta(f(e1, e2, \ldots, en)) = \delta(f)(\delta(e1), \delta(e2), \ldots, \delta(en))$.

2.9. Compositionality

The conditions (*PC1*), (*PC2*) and (*PC3*) are called the principles of compositionality of syntactic forms, meaning and denotation, respectively (cf. Partee et al. 1990; Janssen 1996, 2001; Hodges 1996, 1998, 2001). They have the following scheme of *compositionality* (*Ch*) for the function h representing:

1) the function ι_L, 2) the operation μ and 3) the operation δ:

(Ch) $h(e) = h(f(e1, e2, \ldots, en)) = h(f)((h(e1), h(e2), \ldots, h(en))$.

The scheme (*Ch*) says that: *1) the index, 2) the meaning and 3) the denotation of the main functor of the functor-argument expression e is a function defined on 1) indices, 2) meanings and 3) denotations of successive arguments of this functor.*

The suitable variants of compositionality are some requirement of homomorphisms between the mentioned partial algebras:

$$\mathbf{L} = \langle \mathbf{S}, \mathbf{F} \rangle \quad \xrightarrow[\iota_L]{hom} \quad \iota_L(\mathbf{L}) = \langle \iota_L(\mathbf{S}), \iota_L(\mathbf{F}) \rangle,$$
$$\mathbf{L} = \langle \mathbf{S}, \mathbf{F} \rangle \quad \xrightarrow[\mu]{hom} \quad \mu(\mathbf{L}) = \langle \mu(\mathbf{S}), \mu(\mathbf{F}) \rangle,$$
$$\mathbf{L} = \langle \mathbf{S}, \mathbf{F} \rangle \quad \xrightarrow[\delta]{hom} \quad \delta(\mathbf{L}) = \langle \delta(\mathbf{S}), \delta(\mathbf{F}) \rangle.$$

2.10. Concord between syntactic forms and their correlates

On the level of metatheory, it is possible to show the agreement between syntactic structures of *wfes* of the language reality **S** and their correlates in the conceptual reality **C** and in the ontological reality **O**.

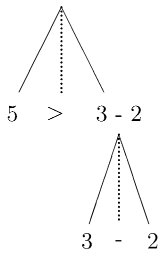
As *wfes* have *function-argument form*: all the functors (all their correlates) precede their arguments (correlates of their arguments as appropriate). Then the algebraic approach to language expressions corresponds to the tree method.

Example. Let us consider two *wfes* of language of arithmetic:

$$\text{a. } 5 > 3 - 2 \quad \text{and} \quad \text{b. } 3 - 2 > -1.$$

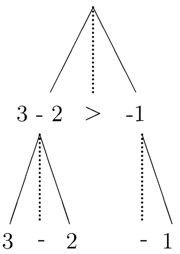
First we present parenthetical recordings a'. and b'. for a. and b. and diagrams of trees meant to explicate them. Diagrams Ta and Tb show a natural, phrasal, natural *functorial analysis* of these expressions. The dotted lines show functors.

Ta. 5 > 3 - 2



a'. (5) > ((3) - (2))

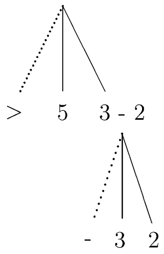
Tb. 3 - 2 > -1



b'. ((3) - (2)) > (- (1))

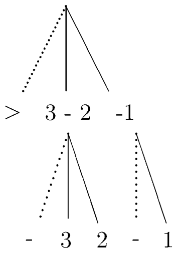
Appropriate function-argument recordings $\mathrm{a}_{f.}$ and $\mathrm{b}_{f.}$ and diagrams of trees: $T\mathrm{a}_{f.}$, $T\mathrm{b}_{f.}$ show "functional analysis" of expressions a., b. in Ajdukiewicz's prefix notation.

$T\mathrm{a}_{f.}$ 5 > 3 - 2



$\mathrm{a}_{f.}$ > (5, - (3, 2))

$T\mathrm{b}_{f.}$ 3 - 2 > -1



$\mathrm{b}_{f.}$ > (- (3, 2), - (1))

Let us note that the functorial analysis of a. and b. given here provides functional-argument expressions $\mathrm{a}_{f.}$ and $\mathrm{b}_{f.}$. It is unambiguously determined due to the semantic (denotational and intensional) functions of the signs '$>$' and '$-$': the first is a sign of two-argument operation on numbers, the second one in a. denotes a two-argument number operation, while in b. it also denotes a one-argument operation.[5] The mentioned signs, as functors, and thus as functions on signs of numbers, have as many arguments as their semantic correlates have.

Comparison of tree method and algebraic method based on compositionality shows one-to-one correspondence of constituents of any *wfe* of L with correlates in order to form and transmit our knowledge on reality $\mathbf{O}$ represented by L (see diagrams of trees $T\mathrm{b}_f$. and $T\mathrm{b}$. of the expression b. and corresponding to them diagrams of trees of categorial indices $T_{\iota_L}(\mathrm{b}_f)$ and $T_{\iota_L}(\mathrm{b})$ of b.).

Let us note that from the principle (*PC1*) and in accordance with the principle (*SC*), for $e = f(e_1, e_2, \ldots, e_n) \in S$ and $\iota_L(e) = a, \iota_L(f) = b$, $\iota_L(e_i) = a_i\,(i = 1, 2, \ldots, n)$, we obtain, on the basis of our theory, the following reconstruction of the *rule of cancellation of indices* used by Ajdukiewicz (1935):

$$(rc) \qquad a/a_1a_2\ldots a_n(a_1, a_2, \ldots, a_n) = a.$$

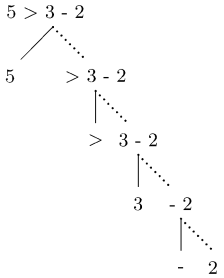
---

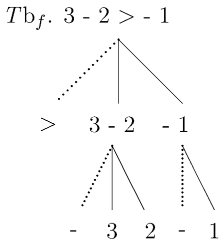
[5] Unambiguous "functorial analysis" is a feature of the languages of formal sciences. In relation to natural languages the analysis depends on linguistic intuition and often allows for a variety of possibilities (see e.g. Marciszewski 1981).

In this conception we do not state that "functoral analysis" of linguistic expressions must be determined unambiguously but we accept the statement that it is connected with expressions of a determined functor-argument structure.

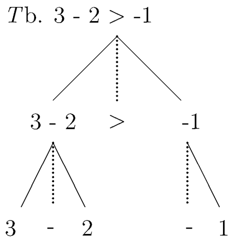
Let us also note that traditional phrasal linguistic analysis, formalized by Chomsky (1957) in his grammars of phrasal structures, takes into consideration grammatical phrasal analysis and only two parts of functoral parsing of expressions.

Let us consider, for instance, the expression a. and its functorial analysis illustrated by a derivation tree in Chomsky's sense.

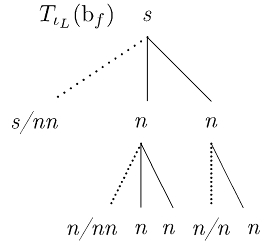




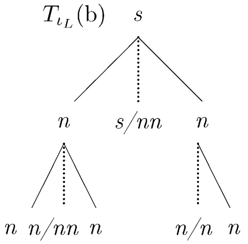
b_f. > (- (3, 2), - (1))



b'. ((3) - (2)) > (- (1))



$\iota_L(\mathrm{b}_f)$. $s/nn(n/nn(n,n),(n/n,n))$



$\iota_L(\mathrm{b}')$. $((n)n/nn(n))s/nn(n/n(n))$

The agreement between syntactic forms of *wfes* and their correlates is very important whenever we want to know whether our knowledge represented in language L is adequate to our cognition of reality.

Let e is any *wfe* of L and C_e is the *set of all constituents* of e. The concord between syntactic structure of e and its correlates is possible because the tree $\mathbf{T}(C_e)$ of constituents of e is isomorphic with trees:

$\mathbf{T}(\iota_L(C_e))$ of indices of all constituents of e,

$\mathbf{T}(\mu(C_e))$ of all meanings of all constituents of e and

$\mathbf{T}(\delta(C_e))$ of all denotations of those constituents.

These trees are formally defined as graphs by means of the set C_e and corresponding to it sets: $\iota_L(C_e)$, $\mu(C_e)$ and $\delta(C_e)$ of all constituents that are appropriate correlates of constituents of e. So,

$$\mathbf{T}(C_e) = \langle C_e, \approx\!> \rangle,$$
$$\mathbf{T}(h(C_e)) = \langle h(C_e), \approx\!>_h \rangle \quad \text{for } h = \iota_L, \mu, \delta,$$

where $\approx\!>$ is a linear ordering *relation of an earlier syntactic position in e* defined by means of the *relation* $\rightarrow$ *of syntactical subordination* (see Ajdukiewicz, 1960); $\approx>_h$ is h-image of the relation $\approx>$.

The mentioned isomorphisms of tree graphs are established by the functions h mapping every constituent of e in C_e that occupies in e a fixed syntactic position (place) onto its h-correlate that occupies in $h(e)$ the same position (place).

All notions introduced in this part can be defined formally.

Definition 1 (constituent of an expression e).

a. $t \in C_e^0 \Leftrightarrow e = t$.

A constituent of the order zero of a given *wfe* e is equal to the expression.

b. $t \in C_e^1 \Leftrightarrow$
$\exists_{n \geq 1} \exists_{f, t_0, t_1, \ldots, t_n \in S} \, (e = f(t_0, t_1, \ldots, t_n) \wedge \exists_{0 \leq j \leq n} (t = f \vee t = t_j))$.
t is a constituent of the first order of a given expression e iff e is a functor-argument expression and t is equal to the main functor of the expression or to one of its arguments.

c. $k > 0 \Rightarrow \left(t \in C_e^{k+1} \Leftrightarrow \exists_{r \in C_e^k} \, t \in C_r^1\right)$.

A *constituent of k+1-th order* of e, where $k > 0$, is a constituent of the first order of a constituent of *k-th* order of e.

d. $t \in C_e \Leftrightarrow \exists_n \, t \in C_e^n$.

A *constituent of a given expression* is a constituent of a finite order of that expression.

Definition 2 (constituent of e with the fixed syntactic position).

a. $t \in C_e^{(j_1)} \Leftrightarrow e$ is a functor-argument expression $\wedge$ t is the j_1-*th* constituent of C_e^1.

b. $k > 0 \Rightarrow \Big(t \in C_e^{(j_1, j_2, \ldots, j_{k+1})} \Leftrightarrow$ t is equal to the j_{k+1}-*th* constituent of a constituent of the set $C_e^{(j_1, j_2, \ldots, j_k)})\Big)$.

Definition 3 (relation of an earlier syntactic position in e).

a. $s \rightarrow s'$ iff $\underset{k,j}{\exists} s \in C_e^k \wedge s' \in C_e^j \wedge k \leq j$.

b. $s \approx> s'$ iff $s \rightarrow s' \vee$
$\left(\exists_{j_1, j_2, \ldots, j_k, n, m} \, (s \in C_e^{(j_1, j_2, \ldots, j_k, n)} \wedge s' \in C_e^{(j_1, j_2, \ldots, j_k, m)} \wedge n < m)\right)$.

s has in e an earlier syntactic position than s' iff s, s' are constituents of e and either s has the order lesser than or equal to the order of s' or s and s' are simultaneously constituents of some part e' of e with the same order $k > 0$ but s has in e' the position n while s' – the position $m > n$.

On the basis of the principles of compositionality it is easy to prove

Theorem 1. *For* $h = \iota_L, \mu, \delta$

$$\mathbf{T}(C_e) = \langle C_e, \approx\!>\rangle \quad \xrightarrow[isom]{h} \quad \mathbf{T}(h(C_e)) = \langle h(C_e), \approx\!>_h\rangle.$$

Uniformity of algebraic approach and tree approach allows to compare knowledge reference to three kinds of realities and to take into account the problem of its adequacy. It is connected with the problem of truthfulness of sentences of L representing knowledge.

3. Three notions of truthfulness

3.1. Three kinds of models of language and the notion of truth

We have treated the language reality **S** and corresponding to it $\iota_L-, \mu-$ and $\delta-$ images of **S**, i.e. $\iota_L(\mathbf{S})$ – a fragment of the indexation reality **I**, $\mu(S)$ – a fragment of the conceptual reality **C** and $\delta(S)$ – a fragment of the ontological reality **O** as some algebraic structures, as some partial algebras.

Let us distinguish in **S** the set of all *sentences* of L. Models of L are nonstandard **models**. They are the three mentioned algebraic structures (partial algebras) given as homomorphic images of algebraic structure $\mathbf{L}$=⟨**S, F**⟩ of language L:

$$\begin{aligned} \iota_L(\mathbf{L}) &= \langle \iota_L(\mathbf{S}), \iota_L(\mathbf{F})\rangle, \\ \mu(\mathbf{L}) &= \langle \mu(\mathbf{S}), \mu(\mathbf{F})\rangle, \\ \delta(\mathbf{L}) &= \langle \delta(\mathbf{S}), \delta(\mathbf{F})\rangle. \end{aligned}$$

They are determined by the fragments $\iota_L(\mathbf{S})$, $\mu(\mathbf{S})$ and $\delta(\mathbf{S})$ of the realities **I**, **C** and **O**, respectively. The first of them $\iota_L(\mathbf{L})$ is *syntactic* one and the next two are *semantic*: $\mu(\mathbf{L})$ – *intensional* and $\delta(\mathbf{L})$ – *extensional.*

3.2. Three notions of truthfulness

For the three models $\iota_L(\mathbf{L})$, $\mu(\mathbf{L})$ and $\delta(\mathbf{L})$ of the language L we define three notions of truthfulness. For this purpose we distinguish three nonempty subsets $T\iota_L, T\mu, T\delta$ of realities **I**, **C** and **O**, respectively:

- $T\iota_L$ consists only of the index of any true sentences,
- $T\mu$ consists of all meanings of sentences of L that are true logical propositions and
- $T\delta$ consists of all denotations of sentences of L that are states of affairs that obtain.

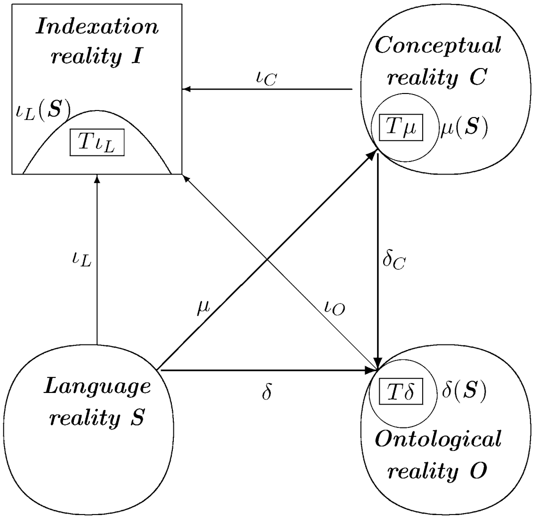


Figure 3.

All of the three definitions of a true sentence in one of the models $\iota_L(\mathbf{L})$, $\mu(\mathbf{L})$ and $\delta(\mathbf{L})$ of L are analogous and are substitutions of the following definition scheme:

Scheme of definitions (truthfulness): *For* $h = \iota, \mu$ and δ

The sentence e *is true in the model* $h(\mathbf{L})$ *iff* $h(e) \in Th$.

The definitions of a true sentence correspond to the **truth value principle** (cf. W. Hodges 1996). An expansion of the principle could be formulated as follows:

The correlate of a sentence (i.e. its index, meaning or denotation, respectively) *determines whether or not it is true in a suitable model.*

The three definitions of a true sentence can be given as follows:

- e *is syntactically true* iff $\iota_L(e) \in T\iota_L$,
- e *is intensionally true* iff $\mu(e) \in T\mu$,
- e *is extensionally true* iff $\delta(e) \in T\delta$.

From the above scheme of definitions of truthfulness of sentences we can easily get the following scheme of theorems:

METATHEOREM 1. *For* $h = \iota_L, \mu, \delta$

If e, e' *are sentences and* $h(e) = h(e')$, *then*
e *is true in* $h(\mathbf{L})$ *iff* e' *is true in* $h(\mathbf{L})$.

Metatheorem 1 is the scheme of the following three theorems our formal theory:

1) *If we have two sentences with the same index then they are syntactically true* iff *they have the same truth value in the syntactic model, i.e. their index is the index of all true sentences*,

2) *If two sentences have the same meanings then they are intensionally true* iff *they have the same truth value in the intensional model, i.e. their meanings are true logical propositions*,

3) *If two sentences have the same denotation then they have the same truth value in the extensional model, i.e. their denotations are the states of affairs that obtain.*

3.3. Reliability of cognition of reality

The main purpose of cognition is aiming at an agreement of truthfulness of sentences that are results of cognition in all three models: $\iota_L(\mathbf{L})$, $\mu(\mathbf{L})$ and $\delta(\mathbf{L})$ (cf. *Figure 3*).

Let us note that if a sentence is true in the extensional model $\delta(\mathbf{L})$ then it does not have to be true in the remaining models. So, in particular, a deductive knowledge that is included in the conceptual reality **C** cannot be in agreement with knowledge referring to the ontological reality **O**. There can be true sentences in $\delta(\mathbf{L})$ that are not deduced from the knowledge accepted earlier and cannot be true in the intensional model $\mu(\mathbf{L})$.

Considerations outlined in this paper point to a new aspect of the importance of Gödel's Incompleteness Theorem (1931): it explains why language cognition of reality illustrated by *Figure 3* can be incomplete.

Justification of these statements requires introducing some new notions.

3.4. Operations of replacement

The most important theorems which follow from the principles of compositionality (*PC1*), (*PC2*) and (*PC3*) use the syntactic notion of the three-argument *operation* π *of replacement of a constituent of a given wfe of L*. The operation π is defined by means of the *operation* π^n *of replacement of the constituents of n-th order*. The expressions $e' = \pi(p', p, e)$ and $e' = \pi^n(p', p, e)$ are read: *the expression* e' *is a result of replacement of the constituent p,* *respectively, the constituent* p *of* n*-th order, of* e *by the expression* p'. The definition of the operation π^n is inductive (see Wybraniec-Skardowska, 1991).

Definition 4 (operation of replacement). *Let* $e, e', p, p' \in \mathbf{S}$. *Then*

a. $e' = \pi^0(p', p, e)$ iff $p = e$ *and* $p' = e'$,

b. $e' = \pi^1(p', p, e)$ iff e *and* e' *are some functor-argument expressions of the set* $\mathbf{S}$ *with the same number of arguments of their main functors and differ from one another only by the same syntactic position when in* e *occurs the constituent* p *and in* e' *occurs the constituent* p',

c. $e' = \pi^{k+1}(p', p, e)$ iff $\exists_{q,q' \in \mathbf{S}} \big(e' = \pi^k(q', q, e) \ \& \ q' = \pi^1(p', p, q)\big)$,

d. $e' = \pi(p', p, e)$ iff $\exists_{n \geq 0}(e' = \pi^n(p', p, e))$.

We can define the operations of replacement $h(\pi)$ for the correlates *wfes* of $\mathbf{S}$ ($h = \iota_L, \mu, \delta$) in an analogous manner.

3.5. The most important theorems

In this part we will give some theorems of our deductive, formal-logical theory of syntax and semantics. They are logical consequences of the above-given definitions and principles of compositionality formulated earlier.

It is easy to justify three principles of compositionality with respect to the operation π. They are a substitution of the following metatheorem:

Metatheorem 2 (compositionality with respect to π). *For* $h = \iota_L, \mu, \delta$

$$(PC_\pi) \qquad h(\pi(p', p, e)) = h(\pi)(h(p'), h(p), h(e)).$$

We can also easily state that the theorems that we get from the next scheme are valid:

Metatheorem 3 (homomorphisms of replacement systems). *For* $h = \iota_L, \mu, \delta$

$$\langle \mathbf{S}, \pi, T \rangle \quad \xrightarrow[hom]{h} \quad \langle h(\mathbf{S}), h(\pi), h(T) \rangle,$$

where T *is the set of all true sentences of* L.

We can postulate that $T\iota_L = \iota_L(T), T\mu = \mu(T)$ and $T\delta = \delta(T)$.

Now, we will present theorems called replacement theorems.

Fact 2. *For* $h = \iota_L, \mu, \delta$
If $e = f(e_1, e_2, \ldots, e_n), e' = f'(e'_1, e'_2, \ldots, e'_n) \in \mathbf{S}$
then $h(e) = h(e')$ *iff* $h(f) = h(f')$ *and* $h(e_i) = h(e'_i)$ *for any* $i = 1, \ldots, n$.

By means of Fact 2 we can easily obtain the one *fundamental syntactic replacement theorem* and two *fundamental semantic replacement theorems* which are the suitable substitutions of the following metatheorem of our theory:

METATHEOREM 4 (replacement principles). *For* $h = \iota_L, \mu, \delta$

If $e, e' \in \mathbf{S}$ *and* $e' = \pi(p', p, e)$ *then* $(h(p) = h(p')$ *iff* $h(e) = h(e'))$.

So: *Two expressions have the same correlate (the same categorial index – the syntactic category, the same meanings, the same denotation, respectively) if and only if by the replacement of one of them by the other in any wfe of L we obtain a wfe of L which has the same correlate (the same categorial index – the same syntactic category, the same meaning, the same denotation, respectively), as the expression from which it was derived.*

COROLLARY 1. *If* $e, e' \in \mathbf{S}$ *and* $e' = \pi(p', p, e)$, *then*

$$\begin{aligned} \exists_\zeta(p, p' \in Cat_\zeta) \quad &iff \quad \exists_\zeta(e, e' \in Cat_\zeta), \\ \exists_\zeta(p, p' \in Con_\zeta) \quad &iff \quad \exists_\zeta(e, e' \in Con_\zeta), \\ \exists_\zeta(p, p' \in Ont_\zeta) \quad &iff \quad \exists_\zeta(e, e' \in Ont_\zeta). \end{aligned}$$

The next theorems are connected with the true value principles.

METATHEOREM 5 (referring to the truth value principles). *For* $h = \iota, \mu, \delta$
If e, e' *are sentences of L and* $e' = \pi(p', p', e)$ *and* $h(p) = h(p')$, *then*
e is true in $h(\mathbf{L})$ *iff* e' *is true in* $h(\mathbf{L})$.

The three theorems that we get from the above metatheorem together state that:

Replacing in any sentence its constituent by an expression which has the same correlate (the same index, the same meaning, the same denotation, respectively), never alters the truth value of the replaced sentence in the given syntactic, intensional, extensional, respectively, model.

If we accept the following axiom:

AXIOM: *If e is a sentence and* $\mu(e) \in T\mu$, *then* $\delta(e) \in T\delta$,

then from the above metatheorem, for $h = \mu$, we get:

FACT 3. *If* e, e' *are sentences,* $e' = \pi(p', p', e)$ *and* $\mu(p) = \mu(p')$, *then*
if e is true in $\mu(\mathbf{L})$ *then* e' *is true in* $\delta(\mathbf{L})$.

So: *Replacing in any true sentence in the intensional model its constituent by an expression that has the same meaning, we get a sentence which is true in the extensional model.*

STRONGER METATHEOREM (referring to truth value principles)
For $h = \iota, \mu, \delta$.
If e, e' *are sentences and* $e' = \pi(p', p, e)$, *then*
$h(p) = h(p')$ *iff* (e *is true in* $h(\mathbf{L})$ *iff* e' *is true in* $h(\mathbf{L})$).

The recognition of the above metatheorem requires accepting the three axioms which are connected with Leibniz's principles (cf. Gerhard 1890, p. 280, Janssen 1996, p.463) and have the same scheme:

SCHEME OF LEIBNIZ'S AXIOMS *For* $h = \iota, \mu, \delta$.
If e, e' *are sentences and* $e' = \pi(p', p, e)$, *then*
if (e *is true in* $h(\mathbf{L})$ *iff* e' *is true in* $h(\mathbf{L})$) *then* $h(p) = h(p')$.

Leibniz's Axioms together state that:

If replacing in any sentence its constituent p *by an expression* p' *never alters the truth value of the replaced sentence in the syntactic, in the intensional, in the extensional, respectively, model, then* p *and* p' *have the same categorial index, the same meaning, the same denotation, respectively.*

Three theorems which follow from *Stronger Metatheorem* (referring to truth value principles) together say that (cf. Hodges 1996):

Two expressions of the language L have the same correlates (the same categorial index – syntactic category or form, the same meaning – intension, the same denotation – extension, respectively), if and only if replacing one of them by another in any sentence never alters the truth value of the replaced sentence in the syntactic, intensional, extensional, respectively, model of the language L.

4. Final remarks

- We have tried to give a description of meta-knowledge in connection with three references of logical knowledge to:
  - language,
  - conceptual reality and
  - ontological reality.
- Thanks to it we could define three kinds of models of language and three kinds of truthfulness in these models.
- These models are not standard models; in particular the notion of truth does not employ the notions of satisfaction and valuation of variables used for formalized languages.

- Adequacy of language knowledge to cognitive objects of reality is understood as an agreement of truthfulness in these three models.
- It is possible to give a generalization of the notion of meta-knowledge in communication systems in order to apply it to knowledge in text systems but the solution of this problem requires more time and is solved by my co-worker Edward Bryniarski.

4.1. Acknowledgements

I thank unknown referees and my colleagues Edward Bryniarski and Marek Magdziak for their stimulating comments, remarks and suggestions which allowed me to complete or improve some fragments of my paper.

References

[1] Ajdukiewicz, K., 1931, 'O znaczeniu' ('On meaning of expressions'), *Księga Pamiątkowa Polskiego Towarzystwa Filozoficznego we Lwowie, Lwów*,

[2] Ajdukiewicz, K., 1934, 'Sprache und Sinn', *Erkenntnis*, IV: 100–138.

[3] Ajdukiewicz, K., 1935, 'Die syntaktische Konnexität', *Studia Pholosophica, Leopoli*, 1: 1–27. English translation: 'Syntactic connection', in McCall, S. (ed.), *Polish Logic 1920–1939*, Clarendon Press, Oxford, pp. 207–231.

[4] Ajdukiewicz, K., 1960, 'Związki składniowe między członami zdań oznajmujących ('Syntactic relation between elements of declarative sentences')', *Studia Filozoficzne*, 6(21): 73–86. First presented in English at the International Linguistic Symposium in Erfurt, September 27–October 2, 1958.

[5] Bar-Hillel, Y., 1950, 'On syntactical categories', *Journal of Symbolic Logic*, 15: 1–16; reprinted in Bar-Hillel, Y., *Language and Information, Selected Essays on Their Theory and Applications*, Addison-Wesley Publishing Co., Reading, Mass, 1964, pp. 19–37.

[6] Bar-Hillel, Y., 1953, 'A quasi-arithmetical notation for syntactic description', *Language*, 63: 47–58; reprinted in *Aspects of Language*, Jerusalem, pp. 61–74.

[7] Bar-Hillel, Y., 1964, *Language and Information, Selected Essays on Their Theory and Applications*, Addison-Wesley Publishing Co., Reading, Mass.

[8] Buszkowski, W., 1988, 'Three theories of categorial grammar', in Buszkowski, W., Marciszewski, W., van Benthem, J. (eds.), pp. 57–84.

[9] Buszkowski, W., 1989, *Logiczne Podstawy Gramatyk Kategorialnych Ajdukiewicza-Lambeka (Logical Foundations of Ajdukiewicz's-Lambek's Categorial Grammars), Logika i jej Zastosowania*, PWN, Warszawa.

[10] Buszkowski, W., Marciszewski, W., van Benthem, J. (eds.), 1988, *Categorial Grammar*, John Benjamis Publishing Company, Amsterdam-Philalelphia.

[11] Cresswell, M.J., 1973, *Logics and Languages*, Mathuen, London.

[12] Cresswell, M.J., 1977, 'Categorial languages', *Studia Logica*, 36: 257–269.

[13] Carnap, R., 1947, *Meaning and Necessity*, University of Chicago Press, Chicago.

[14] Chomsky, N., 1957, *Syntactic Structure*, Mouton and Co., The Hague.

[15] Frege, G., 1879, *Begriffsschrift, eine der arithmetischen nachbildete Formelsprache des reinen Denkens*, Halle; reprinted in Frege, G., *Begriffsschrift und andere Ausätze*, Angelelli, I. (ed.), Wissenschaftliche Buchgesellschaft/G. Olms, Darmstadt-Hildesheim, 1964.

[16] Frege, G., 1884, *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*, W. Koebner, Breslau.

[17] Frege, G., 1892, 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und pilosophishe Kritik*, 100: 25–50.

[18] Frege, G., 1964, *Begriffsschrift und andere Ausätze*, Angelelli, I. (ed.), Wissenschaftliche Buchgesellschaft/G. Olms, Darmstadt-Hildesheim.

[19] Gerhard, C.I. (ed.), 1890, *Die philosophische Schriften von Wilhelm Leibniz, vol. 7*, Weidmansche Buchhandlung, Berlin.

[20] Gödel, K., 1931, 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik*, 38: 173–198.

[21] Hendriks, H., 2000, 'Compositional and model-theoretic interpretation', *Artificial Intelligence Preprint Series*, Preprint nr 020.

[22] Hodges, W., 1996, 'Compositional semantics for language of imperfect information', *Logic Journal of the IGPL*, 5(4): 539–563.

[23] Hodges, W., 1998, 'Compositionality is not the problem', *Logic and Logical Philosophy*, 6: 7–33.

[24] Hodges, W., 2001, 'Formal features of compositionality', *Journal of Logic, Language and Information 10*, pp. 7–28, Kluwer Academic Publishers.

[25] Husserl, E., 1900–1901, *Logische Untersuchungen*, vol. I, Halle 1900, vol. II, Halle, 1901.

[26] Janssen, T.M.V., 1996, 'Compositionality', in van Benthem, J., ter Muelen, A. (eds.), *Handbook of Logic and Language*, Chapter 7, Elsevier Science, Amsterdam–Lausanne–New York, pp. 417–473.

[27] Janssen, T.M.V., 2001, 'Frege, contextuality and compositionality', *Journal of Logic, Language and Information*, 10: 115–136.

[28] Lambek, J., 1958, 'The mathematics of sentence structure', *American Mathematical Monthly*, 65: 154–170.

[29] Lambek, J., 1961, 'On the calculus of syntactic types', in Jakobson, R. (ed.), *Structure of Language and its Mathematical Aspects*, Proceedings of Symposia in Applied Mathematics, vol. 12, AMS, Providence, Rhode Island.

[30] Leśniewski, S., 1929, 'Grundzüge eines neuen Systems der Grundlagen der Mathematik', *Fundamenta Mathematicae*, 14: 1–81.

[31] Leśniewski, S., 1930, 'Über die Grundlagen der Ontologie', *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, Classe II, vol. 23, Warszawa, pp. 111–132.

[32] Marciszewski, W., 1988, 'A chronicle of categorial grammar', in Buszkowski, W., et al. (eds.), *Categorial Grammar*, John Benjamis Publishing Company, Amsterdam-Philalelphia, pp. 7–22.

[33] Montague, R., 1970, 'Universal grammar', *Theoria*, 36: 373–398.

[34] Montague, R., 1974, 'Formal Philosophy', Thomason, R.H. (ed.), *Selected Papers of Richard Montague*, Yale University Press, New Haven, Conn.

[35] Partee, B.H., ter Meulen, A., Wall, R.E., 1990, *Mathematical Methods in Linguistics*, Kluwer Academic Publishers, Dordrecht.

[36] Peirce, Ch.S., 1931–1935, *Collected Papers of Charles Sanders Peirce*, Hartshorne C., Meiss P. (eds.), vols. 1–5, Cambridge, Mass.

[37] Simons, P., 1989, 'Combinators and categorial grammar', *Notre Dame Journal of Formal Logic*, 30(2): 241–261.

[38] Stanosz, B., Nowaczyk, A., 1976, *Logiczne Podstawy Języka (The Logical Foundations of Language)*, Ossolineum, Wrocław-Warszawa.

[39] Suszko, R., 1958, 'Syntactic structure and semantical reference', Part I, *Studia Logica*, 8: 213–144.

[40] Suszko, R., 1960, 'Syntactic structure and semantical reference', Part II, *Studia Logica*, 9: 63–93.

[41] Suszko, R., 1964, 'O kategoriach syntaktycznych i denotacjach wyrażeń w językach sformalizowanych' ('On syntactic categories and denotation of expressions in formalized languages'), in *Rozprawy Logiczne* (*Logical Dissertations*) (to the memory of Kazimierz Ajdukiewicz), Warszawa, pp. 193–204.

[42] Suszko, R., 1968, 'Ontology in the tractatus of L. Wittgenstein', *Notre Dame Journal of Formal Logic*, 9: 7–33.

[43] Van Benthem, J., 1980, 'Universal algebra and model theory. Two excursions on the border', *Report ZW-7908*. Department of Mathematics, Groningen University.

[44] Van Benthem, J., 1981, 'Why is semantics what?', in Groenendijk, J., Janssen, T., Stokhof, M. (eds.), *Formal Methods in the Study of Language*, Mathematical Centre Tract 135, Amsterdam, pp. 29–49.

[45] Van Benthem, J., 1984, 'The logic of semantics', in Landman, F., Veltman F., (eds.), *Varietes of Formal Semantics*, GRASS series, vol. 3, Foris, Dordrecht, pp. 55–80.

[46] Van Benthem, J., 1986, *Essays in Logical Semantics*, Reidel, Dordrecht.

[47] Wittgenstein, L., 1953, *Philosophical Investigations*, Blackwell, Oxford.

[48] Wybraniec-Skardowska, U., Rogalski, A.K., 1999, 'On universal grammar and its formalisation', *Proceedings of 20th World Congress of Philosophy*, Boston 1998, http://www.bu.edu/wcp/Papers/Logi/LogiWybr.htm.

[49] Wybraniec-Skardowska, U., 1985, *Teoria Języków Syntaktycznie Kategorialnych (Theory of Syntactically-Categorial Languages)*, PWN, Wrocław–Warszawa.

[50] Wybraniec-Skardowska, U., 1989, 'On eliminatibility of ideal linguistic entities', *Studia Logica*, 48(4): 587–615.

[51] Wybraniec-Skardowska, U., 1991, *Theory of Language Syntax. Categorial Approach*, Kluwer Academic Publisher, Dordrecht–Boston–London.

[52] Wybraniec-Skardowska, U., 1998, 'Logical and philosophical ideas in certain approaches to language', *Synthese*, 116(2): 231–277.

[53] Wybraniec-Skardowska, U., 2001a, 'On denotations of quantifiers', in Omyła, M. (ed.), *Logical Ideas of Roman Suszko, Proceedings of The Wide-Poland Conference of History of Logic* (to the memory of Roman Suszko), Kraków 1999, Faculty of Philosophy and Sociology of Warsaw University, Warszawa, pp. 89–119.

[54] Wybraniec-Skardowska, U., 2001b, 'Three principles of compositionality', *Bulletin of Symbolic Logic*, 7(1): 157–158. The complete text of this paper appears in *Cognitive Science and Media in Education*, vol. 8, 2007.

[55] Wybraniec-Skardowska, U., 2005, 'Meaning and interpretation', in Beziau, J.Y., Costa Leite, A. (eds.), *Handbook of the First World Congress and School on Universal Logic*, Unilog'05, Montreux, Switzerland, 104.

[56] Wybraniec-Skardowska, U., 2006, 'On the formalization of classical categorial grammar', in Jadacki, J., Paśniczek, J. (eds.), *The Lvov-Warsaw School—The New Generation*, Poznań Studies in the Philosophy of Sciences and Humanities, vol. 89, Rodopi, Amsterdam-New York, NY, pp. 269–288.

[57] Wybraniec-Skardowska, U., 2007a, 'Meaning and interpretation', Part I, *Studia Logica*, 85: 107–134. http://dx.doi.org/10.1007/s11225-007-9026-0.

[58] Wybraniec-Skardowska, U., 2007b, ''Meaning and interpretation', Part II, *Studia Logica*, 85: 263–276. http://dx.doi.org/10.1007/s11225-007-9031-3.

[59] Wybraniec-Skardowska, U., 2007c, 'Three levels of knowledge', in Baaz, M., Preining, N. (eds.), *Gödel Centenary 2006: Posters*, Collegium Logicum, vol. IX, Kurt Gödel Society, Vienna, pp. 87–91.

Urszula Wybraniec-Skardowska
Autonomous Section of Applied Logic and Methodology of Sciences
Poznań School of Banking, Department in Chorzów
Wandy 66
41–500 Chorzów
Poland
`uws@uni.opole.pl`