LNCS 10148

Helle Hvid Hansen
Sarah E. Murray
Mehrnoosh Sadrzadeh
Henk Zeevat (Eds.)

Logic, Language, and Computation

11th International Tbilisi Symposium, TbiLLC 2015
Tbilisi, Georgia, September 21–26, 2015
Revised Selected Papers

Springer

Lecture Notes in Computer Science 10148

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

FoLLI Publications on Logic, Language and Information

Subline of Lectures Notes in Computer Science

More information about this series at http://www.springer.com/series/7407

Helle Hvid Hansen · Sarah E. Murray
Mehrnoosh Sadrzadeh · Henk Zeevat (Eds.)

Logic, Language, and Computation

11th International Tbilisi Symposium, TbiLLC 2015
Tbilisi, Georgia, September 21–26, 2015
Revised Selected Papers

Springer

Editors
Helle Hvid Hansen
Faculty of Technology, Policy and Management
Delft University of Technology
Delft, Zuid-Holland
The Netherlands

Mehrnoosh Sadrzadeh
School of Electronic Engineering and Computer Science
Queen Mary University of London
London
UK

Sarah E. Murray
Department of Linguistics
Cornell University
Ithaca, NY
USA

Henk Zeevat
Heinrich-Heine-Universität
Düsseldorf
Germany

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-662-54331-3 ISBN 978-3-662-54332-0 (eBook)
DOI 10.1007/978-3-662-54332-0

Library of Congress Control Number: 2017932127

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

The 11th International Tbilisi Symposium on Logic, Language, and Computation was held at the Tbilisi State University in Tbilisi, Georgia, during September 21–26, 2015. The symposium was organized by the Centre for Language, Logic and Speech at the Tbilisi State University, the Georgian Academy of Sciences, and the Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam. The biennial conference series and the proceedings are representative of the aims of the organizing institutes: to promote the integrated study of logic, information, and language. While the conference is open to contributions from any of the three fields, it aims to foster interaction among them by achieving stronger awareness of developments in the other fields, and of work that embraces more than one field or belongs to the interface between fields. The scientific program consisted of tutorials, invited lectures, contributed talks, and two workshops.

The symposium offered three tutorials, given on each of the three major disciplines of the conference and aimed at students as well as researchers working in the other areas. The tutorial speakers were Brunella Gerla, Lisa Matthewson, and Joel Ouaknine. Six invited lectures were delivered at the symposium: two on logic by Melvin Fitting and George Metcalfe, two on language by Rajesh Bhatt and Sarah Murray, and two on computation by Helle Hvid Hansen and Mehrnoosh Sadrzadeh. The workshop on How to Make Things Happen in the Grammar: The Implementation of Obligatoriness, organized by Rajesh Bhatt and Vincent Homer featured invited talks by Omer Preminger and Ivy Sichel as well as six contributed talks. The workshop on Automata and Coalgebra was organized by Helle Hvid Hansen and Alexandra Silva and featured invited talks by Bartek Klin, Clemens Kupke, and Stefan Milius.

This volume contains the abstracts for the tutorials and the invited lectures followed by 18 papers that were selected after a rigorous, two-stage refereeing process during which each paper was reviewed by at least two anonymous referees. Here we give a brief overview of their contributions.

Rusiko Asatiani contributes to the semantics of the Georgian verbal morphology by specifying an algorithm for deciding between the active and passive voice. In Georgian, voice has semantic consequences that, however, do not boil down to a simple semantics of voice. The algorithm predicts on the basis of input cognitive features which voice to choose and consequently also predicts which features can be relevant on a given occasion.

Anja Goldschmidt, Thomas Gamerschlag, Wiebke Petersen, Ekaterina Gabrovska, and Wilhelm Geuder investigate the semantics of the German verb *schlagen* (to hit, to beat) using modification by manner adverbs to discover more about the force component in the meaning and agent-oriented adverbs for interactions between force and agentivity. The investigation leads to an analysis of the verb *schlagen* in frame semantics.

Justyna Grudzińska and Marek Zawadowski show how a version of dependent type theory can give a more uniform account of the many readings of the indefinite in examples like: *Not every linguist studied every solution that some problem might have*. The diversity of the many readings is a challenge to the traditional generalized quantifier account.

Petr Homola presents a novel method for translating Hobbs approach to interpretation as abduction into answer set programming. This approach overcomes the weaknesses of previous approaches to abduction, which did not allow for automatically rejecting the inconsistent proofs. The current translation pairs abduction with an inference system, where inconsistent proofs are automatically rejected and search space is naturally reduced.

Dawei Jin gives an account of weak and strong intervention effects for *why*-questions in Chinese, based on a distinction between monotone decreasing quantifiers (strong effects) and quantifiers that are indefinite plurals (weak effects).

Liana Lortkipanidze, Nino Amirezashvili, Anna Chutkerashvili, Nino Javashvili, and Liana Samsonadze document the design and implementation of their syntactic annotation of the Georgian Literary Corpus. The program tools offer modules for the morphologic, syntactic, and semantic levels. The paper gives the description of the automatic syntactic analyzer.

Sebastian Löbner extends the frame formalism and its model-theoretic semantics by first-order comparators. These are two-place attributes that capture basic comparison relationships between objects of the same type. The extension is used for giving a general frame decomposition for punctual verbs of change and a number of special cases of such verbs, using a direct implementation of Allen's calculus of temporal intervals.

Ralf Naumann and Wiebke Petersen use logical systems for default reasoning and belief revision to capture semantic prediction and for discarding faulty interpretational hypotheses. The paper thereby gives a logical account of the neurophysiological research findings in which sentence comprehension relies strongly on semantic prediction and, as a result of this, on the retraction of errors.

Peter Sutton and Hana Filip solve the problem of variation between mass and count conceptualization for the same nouns by distinguishing four classes: prototypical count (bird), compound artifacts (furniture), granular (sand), and substance (mud). These are compared from two perspectives: individuation and consistency, where the context forces one to take priority over the other. This leads to variation for the compound artifacts and the granular nouns.

Henk Zeevat defines a direct semantic interpretation of (augmented) dependency graphs, using ideas from frame semantics and from discourse representation theory. Dependency graphs as conceived in the paper are thereby not just useful for evaluating stochastic parsers, but can also be used for disambiguation by semantic methods.

Richard Zuber provides an algebraic characterisation of the so-called reflexive and anaphoric determiners (as in: *John admires most linguists, including himself* or *John and Mary like no authors, except each other*). Such determiners turn out to be substantially different from ordinary determiners in the algebraic approach to NL semantics going back to Keenan and Faltz.

Alexandru Baltag, Nick Bezhanishvili, Aybüke Özgün, and Sonja Smets generalize their previous topological semantics for belief by interpreting belief as the interior of the closure of the interior operator. Their resulting belief logic is strictly stronger than KD4 and strictly weaker than KD45. They encode in these spaces the semantics for conditional beliefs and updates. Relevant soundness and completeness theorems are proved.

Nick Bezhanishvili, Dick de Jongh, Apostolos Tzimoulis, and Zhiguang Zhao provide a universal model for the positive fragment of intuitionistic logic. A representation for characterizing positive formulae is presented and the universal model is formulated using this characterization. An alternative proof of a theorem by Jankov is provided, where it is shown that the intermediate logic KC, the logic of weak excluded middle, is maximal with regard to intuitionistic propositional calculus.

Pietro Codara and Diego Valota use formal concept analysis to provide an intuitive semantics for the Gödel–Dummett many-valued logic. The connecting point is the use of a Heyting algebra, which provides a basis for formal concept analysis and an algebraic variety for the class of Gödel–Dummett logics. A characterization of Gödel implication and negation is developed in terms of concepts, and a Gödel algebra of concepts is presented.

Zoltán Ésik (RIP) provides a representation theorem for stratified lattices. These are lattices endowed with a certain sequence of preorder relations, representing infinite supplies of truth values, and they have been developed as a framework for solving fixpoint equations of non-monotone operators. The representation theorem is based on the inverse limits of continuous lattices, and has as a corollary that fixpoints of certain weakly monotone functions exist.

Christian Fermüller and Ondrej Majer relate Hintikka's game semantics of independence friendly (IF) logic to Giles's game developed as semantics for Lukasiewicz logic. The results are based on interpreting the expected payoffs of IF games as the fuzzy truth values from the interval [0,1]. It is shown that any rational number is the value of a propositional IF logic formulae and a logic with both fuzzy and IF connectives is developed.

Melvin Fitting provides a new algorithm for connecting modal logics to justification logics. The first such algorithm was developed by Artemov. The current algorithm differs from that of Artemov in two ways. First, it works on the steps of the proof rather than the proof as a whole. Second, the algorithm has two parts, one of which is specific to the modal logic in consideration, in this case S4, the other general to all modal proofs. This two-stage process is novel in the literature. The process is automated in Prolog.

Dick De Jongh and Fatemeh Shirmohammadzadeh Maleki develop a Hilbert-style proof system for the basic sub-intuitionistic logic F, introduced by Corsi and Restall, and prove weak and strong completeness theorems. This provides an alternative to the Kripke semantics of logic F, whose frames lack certain fundamental properties. Further, the authors show that intuitionistic propositional logic is conservative over the logic F and also over Visser's basic logic.

TbiLLC 2015 was extra special since during the symposium week, Professor Dick de Jongh (University of Amsterdam) and Professor Matthias Baaz (Vienna University of Technology) were awarded honorary doctorates of the Ivane Javakhishvili Tbilisi

State University in recognition of their contributions to the Georgian school of logic, mathematics, and linguistics over the past two decades. In particular, they have both been instrumental in the success of the TbiLLC symposium series and the Tbilisi Summer School in Logic and Language not only through their scientific contributions, but also by organizing and promoting these events, and by obtaining funding to ensure their continuity. The award ceremony took place on September 25, 2016, and it was attended by the symposium participants as well as numerous local staff and students. The opening words were spoken by David Gabelaia (TSU Razmadze Mathematical Institute), followed by speeches by the rector of the university and member of the Georgian Academy of Sciences, Vladimer Papava, by the renowned Georgian linguist, member, and former president of the Georgian Academy of Sciences, Thamaz Gamkrelidze, and by Nick Bezhanishvili, scientific descendant of Dick de Jongh, currently assistant professor at the ILLC, University of Amsterdam. The ceremony concluded with the acceptance speech by Dick de Jongh. A full transcript of all speeches and photos from the ceremony can be found on the TbiLLC 2015 website.

This proceedings volume has the sad honor to contain one of the last papers by Zoltán Ésik, who was a well-known and respected member of the theoretical computer science community. We learned the shocking news of Zoltán's passing during the reviewing phase. As both reviewers recommended acceptance and remarked that the manuscript was "written with extreme care" and could be "published virtually as is", we very much wanted to include it in the volume. We are thankful to his widow and son, who kindly provided us with their consent to publish the paper. We are also grateful to Dexter Kozen (Cornell University), who solved the practical issue of obtaining the LaTeX source from the PDF manuscript. The preface is followed by a short essay by Dexter Kozen in commemoration of Zoltan Ésik.

We would like to thank all the authors for their contributions, and the anonymous reviewers for their high-quality reports. We would also like to express our gratitude to the organizers of the symposium, who made the event an unforgettable experience for all of its participants. The Tbilisi symposia are renowned not only for their high scientific standards, but also for their friendly atmosphere, and heartwarming Georgian hospitality, and the 11th symposium was no exception. Finally, we thank the ILLC (University of Amsterdam), Sebastian Löbner, and Johan van Benthem for their generous financial support for the symposium.

December 2016

Helle Hvid Hansen
Sarah Murray
Mehrnoosh Sadrzadeh
Henk Zeevat

Organization

Organizing Institutions

Centre for Language, Logic and Speech, Tbilisi State University, Georgia.
Georgian Academy of Sciences.
Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands.

Organizing Committee

Rusiko Asatiani	Tbilisi State University, Georgia
Zoé Christoff	University of Amsterdam, The Netherlands
Anna Chutkerashvili	Tbilisi State University, Georgia
David Gabelaia	Tbilisi State University, Georgia
Kristina Gogoladze	University of Amsterdam, The Netherlands
Paula Henk	University of Amsterdam, The Netherlands
Julia Ilin	University of Amsterdam, The Netherlands
Marina Ivanishvili	Tbilisi State University, Georgia
Nino Javashvili	Tbilisi State University, Georgia
Ramaz Liparteliani	Tbilisi State University, Georgia
Liana Lortkipanidze	Tbilisi State University, Georgia
Peter van Ormondt	University of Amsterdam, The Netherlands
Aybüke Özgün	University of Lorraine, France
Khimuri Rukhaia	Tbilisi State University and Sokhumi State University, Georgia

Program Committee

Daniel Altshuler (Co-chair)	Heinrich-Heine-University Düsseldorf, Germany
Nick Bezhanishvili	University of Amsterdam, The Netherlands
Agata Ciabattoni	Technical University of Vienna, Austria
Petr Cintula	Czech Academy of Sciences, Czech Republic
Hana Filip	Heinrich-Heine-University Düsseldorf, Germany
Davide Grossi	University of Liverpool, UK
Dag Haug	University of Oslo, Norway
Achim Jung	University of Birmingham, UK
Dexter Kozen	Cornell University, USA
Marcus Kracht	Bielefeld University, Germany
Clemens Kupke	University of Strathclyde, UK
Alda Mari	Institut Jean Nicod, France

Roumi Pancheva	University of Southern California, Los Angeles, USA
Jessica Rett	University of California, Los Angeles, USA
Alexandra Silva	University College London, UK
Alex Simpson	University of Edinburgh, UK
Sonja Smets	University of Amsterdam, The Netherlands
Luca Spada (Co-chair)	University of Amsterdam, The Netherlands, and University of Salerno, Italy
Mark Steedman	University of Edinburgh, UK
Matthew Stone	Rutgers University, USA
Henriette de Swart	Utrecht University, The Netherlands
Judith Tonhauser	Ohio State University, USA
Eytan Zweig	University of York, UK

Standing Committee

Rusiko Asatiani	Tbilisi State University, Georgia
Matthias Baaz	Technical University of Vienna, Austria
Guram Bezhanishvili	New Mexico State University, USA
Peter Bosch	University of Osnabrück, Germany
George Chikoidze	Georgian Technical University, Georgia
Dick de Jongh (Chair)	University of Amsterdam, The Netherlands
Paul Dekker	University of Amsterdam, The Netherlands,
Hans Kamp	University of Stuttgart, Germany
Manfred Krifka	Zentrum für Allgemeine Sprachwissenschaft Berlin, Germany
Temur Kutsia	Johannes Kepler University, Austria
Sebastian Löbner	Heinrich-Heine-University Düsseldorf, Germany
Barbara Partee	University of Massachusetts Amherst, USA

Tutorials

Lisa Matthewson	University of British Columbia, Canada
Brunella Gerla	University of Insubria, Italy
Joel Ouaknine	University of Oxford, UK

Invited Speakers

Rajesh Bhatt	University of Massachusetts Amherst, USA
Sarah E. Murray	Cornell University, USA
Melvin Fitting	City University of New York, USA
George Metcalfe	University of Bern, Switzerland
Helle Hvid Hansen	Delft University of Technology, The Netherlands
Mehrnoosh Sadrzadeh	Queen Mary University of London, UK

Workshop Organizers

Rajesh Bhatt	University of Massachusetts Amherst, USA
Helle Hvid Hansen	Delft University of Technology, The Netherlands
Vincent Homer	University of Massachusetts Amherst, USA
Alexandra Silva	University College London, UK

Invited Speakers at Workshops

Bartek Klin	Warsaw University, Poland
Clemens Kupke	University of Strathclyde, UK
Stefan Milius	Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Omer Preminger	University of Maryland, USA
Ivy Sichel	Hebrew University of Jerusalem, Israel

Zoltán Ésik (1951–2016)

Through the years, Zoltán and I often agreed to disagree. Our tumultuous relationship was at times magnificent, at time exasperating. Perhaps this is the normal state of affairs with two strong-willed researchers with competing ideas and similar interests. Nevertheless, I thoroughly relished my interactions with him, and I miss him terribly.

Zoltán's work with Steve Bloom on iteration theories stands as a remarkable and comprehensive achievement by any measure. It is a truly wide-ranging and fundamental theory with far reaching implications. Zoltán and I were competitors in a very real sense: My own work on Kleene algebra attempts to address many of the same issues from a somewhat different perspective. Both systems purport to capture the essence of imperative computation at a very basic equational level. We often argued about the relative merits of the two systems, mostly inconclusively. My student Konstantinos Mamouras, in his PhD thesis, discovered what was needed to reconcile the two systems and managed to produce a unification; but I regret never having had the pleasure of discussing this work with Zoltán or getting his reaction. I am sure he would have had much to say.

Zoltán and I started working together on a topic of mutual interest during the summer of 2013 at the LICS conference in New Orleans. We developed the basic ideas and wrote down some rough notes, but we did not have the occasion to return to it until shortly before his passing. He produced the penultimate version of a paper, which I am currently in the process of revising for publication. I am grateful to have been a part of one of his last works, although I am sure the final version would have been much improved with his collaboration.

In spite of our occasional disagreements, I admired Zoltán greatly for his consummate scholarship, clarity of thought, and good taste. I was always inspired by his talks and his ideas. I am truly honored to be one of the last to have shared scientific ideas with him. His untimely passing is a great tragedy and loss for our field. He will be missed.

December 2016 — Dexter Kozen

Ithaca, New York, USA

Abstracts of Tutorials and Invited Lectures

Many-Valued Logic

(Tutorial in Logic)

Brunella Gerla
University of Insubria, Varese, Italy

The topic of my tutorial is many-valued logic. By many-valued logic we generally mean a (propositional) logic in which there are more than two truth values. This is a very general setting. We restrict it by considering only logics in which the connectives are truth functional, that is, they can be described by a sort of many-valued truth tables. So we investigate propositional logics in which the connectives are interpreted in [0,1] (real unit interval): the conjunction is interpreted by a t-norm, that is, a commutative, associative binary operation on [0,1] having 1 as neutral element and 0 as absorbent element, and the implication is interpreted by the associated residuum that exists whenever the t-norm is left continuous. The logic of formulas holding true (i.e. equal to 1) when interpreted by any left continuous t-norm is called Monoidal t-norm based Logic (MTL) and can be axiomatized by a finite number of propositional axioms. Adding axioms to MTL we get a whole hierarchy of many-valued logics, including Łukasiewic, Gödel, Product, Nilpotent Minimum logics and, of course, classical propositional logic. In our tutorial we will take a look at the axiomatic extensions of MTL and then we shall focus on some of them. In particular, one of the aims is to show the connections between logical systems and associated algebraic structures, generalizing the existing connection between classical propositional logic and Boolean algebras, or between intuitionistic logic and Heyting algebras. For instance, the characterization of truth tables of some of the infinitely-valued propositional logics of the MTL hierarchy can be seen from the algebraic side as the characterization of free algebras in the related varieties.

Outline of the tutorial:

- Many-valued logics, t-norms, MTL and BL.
- Axiomatic extensions of MTL.
- Łukasiewicz logic and MV-algebras: merging lattice structure with groups.
- Gödel logic and Nilpotent Minimum logic: where only the order of truth valus counts.
- (If time allows) Many-valued propositional formulas as events for non-classical probabilities.

Cross-Linguistic Semantics: Methods, Results, and Theoretical Implications

(Tutorial in Language)

Lisa Matthewson
The University of British Columbia, Vancouver, Canada

This tutorial series provides an introduction to the methods of cross-linguistic semantic study, showcases some recent results from research on endangered languages of North America, and situates the results with respect to the core enterprise of uncovering linguistic universals and diversity.

Session 1 introduces the empirical and theoretical goals of cross-linguistic semantics, and overviews both standard and recently-discovered methodologies for this type of research. Session 2 is a case study on modals. Session 3 presents a case study on quantifiers. We conclude with a discussion of what is known about universals and variation in the semantic component of the grammar.

By default, the tutorial will be structured as three presentations with time for discussion and questions. However, it is possible to run this as a more fully interactive event. Interested participants should feel free to contact the convener at lisa.matthewson@ubc.ca if they have requests.

Program Termination – Survey and Challenges

(Tutorial in Computation)

Joel Ouaknine
University of Oxford, Oxford, UK

In the quest for program analysis and verification, program termination — determining whether a given program will always halt or could execute forever — has emerged as a pivotal component. Unfortunately, this task was proven to be undecidable by Alan Turing eight decades ago, before the advent of the first working computers! In recent years, however, great strides were made in the automated analysis of termination of programs, from simple counter machines to Windows device drivers.

Perhaps surprisingly, from a theoretical standpoint the study of termination involves advanced techniques from a variety of mathematical fields, including analytic and algebraic number theory, Diophantine geometry, real algebraic geometry, model theory, and Ramsey theory.

In this tutorial, we will survey a cross section of topics ranging from the history of program termination to the development of modern tools such as Microsoft Research's Terminator, presenting in the process some of the theoretical challenges that remain and that we expect will drive the field for the foreseeable future.

Justification Logics, Realization, and an Implementation

(Invited Lecture in Logic)

Melvin Fitting
City University of New York, New York City, USA

Gödel began a project to find an arithmetic semantics for intuitionistic logic, but did not complete it. It was finished by Sergei Artemov, in the 1990's. As an essential step in this work, Artemov introduced the first justification logic, LP, (standing for logic of proofs). LP is a modal-like logic, with an infinite family of proof or justification terms, and can be seen as an explicit version of the familiar modal logic S4. Since then, many other justification logic/modal logic pairs have been investigated, and justification logic has become a subject of independent interest, going beyond the original connection with intuitionistic logic. It is now known that there are infinitely many modal logics with justification logic counterparts, but the exact extent of the family is not known. Justification logics are connected with their corresponding modal logics via a Realization Theorem. A Realization Theorem connecting LP and S4 has a constructive proof, but there are other cases in which realization holds, but it is not known if a constructive proof exists. I will discuss Realization Theorems in general, and LP/S4 in particular. The realization proof I will talk about has a two part structure, going through a quasi-realization stage. This gives some additional insight into the phenomenon. I will conclude by demonstrating a computer implementation of the algorithm behind realization for LP/S4.

(Uniform) Interpolation in Logic and Algebra

(Invited Lecture in Logic)

George Metcalfe
University of Bern, Bern, Switzerland

Interpolation is a property of logical systems, first introduced and proved for first-order classical logic by William Craig in the 1950s. For a propositional logic *L*, it says, very roughly, that if a formula *A* entails another formula *B* in *L*, then there is an "interpolating" formula *C* whose variables occur in both *A* and *B* such that *A* entails *C* and *C* entails *B* in *L*. Remarkably, there exists a (well-known) close connection between interpolation and the algebraic property of amalgamation, which allows algebraic structures to be suitably combined so that some common substructure is preserved. In this talk, I will describe this relationship – a bridge between logic and algebra – in the context of non-classical logics and their corresponding algebraic semantics. I will also describe (new) algebraic characterizations of stronger interpolation properties, where the interpolating formula *C* may be chosen "uniformly" based only on the formula *A* (or *B*) and a subset *X* of its variables, providing then an interpolant for any *B* (or *A*) whose overlapping variables with *A* (or *B*) occur in *X*.

PPIs and Movement in Hindi-Urdu

(Invited Lecture in Language)

Rajesh Bhatt and Vincent Homer
University of Massachusetts Amherst, Amherst, MA, USA

Typically, Positive Polarity Items (PPIs), e.g. *would rather*, cannot be interpreted in the scope of a clausemate negation (barring rescuing or shielding) (Baker 1970, van der Wouden 1997, Szabolcsi 2004 a.o.):

(1) a. John would rather leave. b. *John wouldn't rather leave.

The scope of most of them is uniquely determined by their surface position. But PPI indefinites are special: they can surface under negation and yet yield a grammatical sentence under a wide scope interpretation:

(2) John didn't understand something. ✓ SOME $\gg$ NEG;*NEG $\gg$ SOME

Here we address the question of the mechanism through which a PPI of the *some* type takes wide scope out of an anti-licensing configuration. One possibility is (covert) movement, another is mechanisms that allow indefinites to take (island-violating) ultra-wide scope such as choice functions (Reinhart 1997). The relevant configurations that have motivated choice functions for other languages can be set up for Hindi-Urdu too.

(3) logõ=ko lagtaa hai [ki Saima aur kuch tabalcii gaayab ho people=DAT seem is that Saima and some.PL tabla.player disappear be gaye haĩ]
go.PFV.MPL are
'People think that Saima and some tabla players have disappeared.' (ok: SOME $\gg$ THINK)

We can therefore assume that a device that generates wide-scope for indefinites without movement is available in Hindi-Urdu too. We show that in Hindi-Urdu at least, this device is unable to salvage PPIs in the relevant configuration. Only good old fashioned overt movement does the needful. If we think of overt movement in Hindi-Urdu as being the analogue of covert movement elsewhere, then the Hindi-Urdu facts are an argument that it is movement, albeit covert, that salvages PPIs in English too, not alternative scope-shifting devices. We explore whether the conclusion from Hindi-Urdu does in fact extend to English.

Complex Connectives

(Invited Lecture in Language)

Sarah E. Murray
Cornell University, Ithaca, NY, USA

This talk discusses the interpretation and analysis of several sentential connectives found in Cheyenne (Algonquian), drawing on the author's fieldwork as well as several collections of texts. Coordinating connectives in English, including *and* (conjunction), *but* (contrastive conjunction), and *or* (disjunction), are monomorphemic. In Cheyenne, the basic form used for conjunction is *naa*. Other connectives are morphologically complex, formed by combining *naa* with other morphemes, all of which have independent uses.

These complex connectives, and certain uses of *naa* alone, complicate a compositional, truth-functional analysis of the Cheyenne connectives. In particular, though disjunction is logically weaker than conjunction, the two forms for disjunction – *naa matȯ=héva* and *naa mó=héá'e* – each contain the conjunction *naa*. Recently, several authors have proposed analyses of related data from other languages, arguing the basic element is not true conjunction. However, the data in these languages differ from Cheyenne in crucial ways. Building on these analyses, and other proposals on the semantics of disjunction, this talk proposes an analysis of the Cheyenne connectives that preserves *naa* as conjunction.

Brzozowski Minimization via Dual Adjunctions

(Invited Lecture in Computation)

Helle Hvid Hansen
Delft University of Technology, Delft, The Netherlands

Brzozowski's minimisation algorithm [1] seems to work like magic. Given a (non) deterministic automaton $\mathcal{A}$, reverse its transitions, swap initial and final states, make it deterministic and reachable; then do it all again. The result is a minimal deterministic automaton that accepts the same language as $\mathcal{A}$.

A direct correctness proof is not difficult, but a more insightful perspective was given in [2] where Brzozowski's algorithm was described in terms of a duality between reachability and observability, and generalised to Moore automata and weighted automata. At a more abstract level, the algorithm was explained by a dual adjunction of automata. Around the same time, a similar minimisation construction for automata over various base categories was presented in [3] in terms of dual equivalences between coalgebras (automata) and algebras (logic).

The authors of these two lines of work have since joined forces with the aim of finding a unifying framework and new examples. In this talk, I will give an overview of our current insights, and show how to obtain a minimal deterministic automaton from a finite alternating automaton via the duality between complete atomic Boolean algebras and sets.

References

1. Brzozowski, J.A.: Canonical regular expressions and minimal state graphs for definite events. In: Mathematical Theory of Automata. MRI Symposia Series, vol. 12. Polytechnic Press (1962)
2. Bonchi, F., Bonsangue, M.M., Hansen, H.H., Panangaden, P., Rutten, J.J.M.M., Silva, A.: Algebra-Coalgebra duality in Brzozowski's Minimization Algorithm. ACM Trans. Comput. Logic **15**(1) (2014)
3. Bezhanishvili, N., Panangaden, P., Kupke, C.: Minimization via duality. In: Ong, L., de Queiroz, R. (eds.) Logic, Language, Information and Computation. LNCS, vol. 7456, pp. 191–205. Springer, Berlin (2012)

A Multilinear Algebraic Computational Model of Natural Language

(Invited Lecture in Computation)

Mehrnoosh Sadrzadeh
Queen Mary University of London, London, UK

Computational models of natural language can be be categorised into syntax, semantics, and pragmatics. Syntactic models include Chomsky's work on generative grammars and its algebraic type-categorial counterparts (Ajdukiewicz, Lambek). Semantic models mainly span around Montague's translation of natural language sentences to higher order logic formulae. These models do not deal with data-driven sources of text such as Wikipedia, books, and news. The pragmatic models (Harris, Firth) argue that representations of words should be based on the contexts in which they often occur. Here, various statistical measures are developed to retrieve information from data and to reason about them. A popular formal framework thereof is that of vector spaces. These provide a solid base for word meanings, but it is less clear how to extend them to phrases and sentences.

We provide a multilinear algebraic setting inspired by category theoretical models of Lambek's pregroup grammars, and of vector spaces and tensor products, and develop a solution. Here, data from corpora and empirical evaluations provide essential resources.

In this talk I will present our original framework (with Clark and Coecke [1], with Grefenstette [2]), its extension to unified sentence spaces (with Kartsaklis and Pulman [3]), to relative pronouns (with Clark and Coecke [4, 5]), and to quantifiers (with Rypacek [6] and Hedges [7]). I will also go through experimental results as relevant to each extension.

References

1. Coecke, B., Sadrzadeh, M., Clark, S.: Mathematical foundations for a compositional distributional model of meaning. Linguistic Analysis, Volume dedicated to Lambek's 90th Birthday, J. van Benthem, M. Moortgat, W. Buszkowski, pp. 345–384 (2010)
2. Grefenstette, E., Sadrzadeh, M.: Concrete models and empirical evaluations for the categorical compositional distributional model of meaning. Comput. Linguist. **41**, 71–118 (2015)
3. Kartsaklis, D., Sadrzadeh, M., Pulman, S.: A unified sentence space for categorical distributional-compositional semantics: theory and experiments. In: Proceedings of 24th International Conference on Computational Linguistics. COLING 2012, pp. 549–558 (2012)

4. Sadrzadeh, M., Clark, S., Coecke, C.: Frobenius anatomy of word meanings I: subject and object relative pronouns. J. Logic Comput. **23**, 1293–1317 (2013)
5. Sadrzadeh, M., Clark, S., Coecke, B.: Frobenius anatomy of word meanings 2: possessive relative pronouns. J. Logic Comput. **26**, pp. 3–29 (2016)
6. Rypacek, O., Sadrzadeh, M.: A low-level treatment of generalised quantifiers in categorical compositional distributional semantics. In: Joint Proceedings of the Second International Workshop on Natural Language and Computer Science (NLCS'14) and First International Workshop on Natural Language Services for Reasoners (NLSR 2014). de Paiva, Neuper, Quaresma, Retoré, Moss, Saludes (eds.), TR 2014/02, Center for Informatics and Systems of the University of Coimbra, pp. 165–177 (2014)
7. Hedges, J., Sadrzadeh, M.: A Generalised Quantifier Theory of Natural Language in Categorical Compositional Distributional Semantics with Bialgebras. Corr, abs/1602.01635, presented in QPL2016 (2016)

Contents

Logic and Computation

Language and Logic

An Algorithm Defining the Choice of ‘Active~Passive’ Formal Paradigms in Georgian

Rusudan Asatiani(✉)

Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia
rus_asatiani@hotmail.com

Abstract. Two different formal paradigms traditionally referred to as ‘Active’ and ‘Passive’ are clearly distinguished in Georgian; however, there are many cases in which a simple semantic-functional interpretation of the paradigms cannot be given inasmuch as the constructions pointed out as ‘Active’ or ‘Passive’ can actually represent a variety of verb semantics: non-conversive passives (both dynamic and static), active intransitive processes, reflexives, reciprocals, potentials, deponents, etc. Thus, the problem with these paradigms is that it is difficult to predict the meaning from the form and, to such an extent, traditional terms ‘Active’ and ‘Passive’ actually have a conventional character. This paper suggests a cognitive model based on certain semantic features that define the choice of either the passive or the active formal paradigms for grammatical representations of so-called ‘medial’ verbs. The process of choice is organized as an algorithm with four stages of implicational rules and mirrors the hierarchically organized optimal dynamic process of linguistic structuring of an active ~ passive continuum.

Keywords: Georgian · Active ~ passive opposition · Medial verbs · Continuum of active ~ passive opposition · Algorithmic rules in grammar

1 Introduction: Posing the Problems

Two different paradigms establishing the formal opposition between verb forms are clearly distinguished in Georgian. The main morphosyntactic features that define the formal opposition and establish the paradigms can be summarized as in Table 1 below.

Georgian grammarians interpret A-, B-, C-arguments correspondingly as S, DO and IO functions[1] and discuss the opposition of paradigms in terms of the arguments' functional changes;[2] consequently, the distinguished paradigms are referred to as ‘Active’ (resp. I-paradigm) and ‘Passive’ (resp. II-paradigm). The Active paradigm mostly represents active transitive constructions (see below (a)-examples), while their

[1] The verb forms in the examples are glossed according to this tradition.

[2] This assumes: Passive constructions are considered as conversive ones of the corresponding active constructions, where a Patient is promoted to the subject position along the following string of hierarchically organized functional relations: S > DO > IO, while an Agent is demoted and transformed into a prepositional phrase; therefore, it no longer represents a core argument defined by a verb valency.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 3–17, 2017.
DOI: 10.1007/978-3-662-54332-0_1

Table 1. Morphosyntactic Features of Formal Paradigms

	Features	I-Paradigm	II-Paradigm
1.	Vowel prefixes	–	*i-, e-, -d, Ø*
2.	A-ARG.3.SG suffix (in Present)	*-s*	*-a*
3.	A-ARG.3.PL suffix (in Aorist)	*-es*	*-nen*
4.	Present tense marker	*-Ø*	*-i*
5.	Thematic marker (in I-series TAM forms[a])	*-eb-, -ob-, -op-, -av-, -am-, -i-, -Ø-;*	*-eb-*
6.	Imperfective marker	*-d-, (-od-)*	*-od-*
7.	A-ARG's case	NOM (in I-series) ERG (in II-series) DAT (in III-series)	NOM
8.	B-ARG's case	DAT (in I-series) NOM (in I- and II-series)	DAT
9.	C-ARG's case	DAT	DAT

[a]There are 11 different verb forms expressing various combinations of Tense/Mood/Aspect (TAM) categories. According to Georgian grammarians, they are called *mts'k'rivi*. Based on morphosyntactic features the *screeves* are grouped and organized into three series. I-series includes: Present Indicative, Imperfect, Present Subjunctive, Future Indicative, Conditional; Future Subjunctive; II-series consists of: Aorist Indicative and Optative; and III-series comprises: Perfect, Pluperfect, and Perfect Subjunctive [8].

conversive passives (see below (b)-examples), as a rule, are grammaticalized by the second, Passive-paradigm [8][3].

(1) a. *k'ac-**i***(f.7)[4] *m-xat'-**av***(f.5)-*s*(f.2) *me*(f.8)
man-NOM DO.1.SG-paint-THM-PRS.ACT.S.3.SG 1.SG.DAT
'A man is painting me.'

b. *me*(f.7) *v-**i***(f.1)-*xat'-**eb***(f.5)-***i***(f.4) *k'ac-is* *mier.*
1.SG.NOM S.1-PASS-paint-THM-PRS.PASS.SG man-GEN by
'I am being painted by a man.'

(2) a. *k'ac-**i***(f.7) *a-šen-**eb***(f.5)-***s***(f.2) *saxl-**s***(f.8).
man-NOM NV-build-THM-PRS.ACT.S.3.SG house-DAT
'A man builds a house.'

b. *saxl-**i***(f.7) *šen-**d***(f.1)-***eb***(f.5)-***a***(f.2) *k'ac-is* *mier.*
house-NOM build-PASS-THM-PRS.PASS.S.3.SG man-GEN by
'A house is built by a man.'

[3] The main distinguishing formal features are bolded and, in this paper, are conventionally numbered according to the Table 1 as f.1, f.2 …f.9.

(3) a. *k’ac-**i***(f.7) *u-gzavn-**i***(f.5)-***d***(f.6.)-*a* *c’eril-**s***(f.8) *kal-**s***(f.9).
man-NOM IO.3.OV-send-THM-IMP-S.3.SG letter-DAT woman-DAT
‘A man was sending a letter to a woman.’

b. *c’eril-**i***(f.7) ***e***(f.1)-*gzavn-**eb***(f.5)-***od***(f.6)-*a* *kal-**s***(f.9) *k’ac-is mier.*
letter-NOM IO.3.PASS-send-THM-IMP-S.3.SG woman-DAT man-GEN BY
‘A letter is being sent to a woman by a man.’

(4) a. *deda*(f.7) *g-a-tb-**ob***(f.5)-***s***(f.2) *šen*(f.8).
mother.NOM DO.2-NV-warm-THM-PRS.ACT.S.3.SG 2.SG.DAT
‘A mother warms you.’

b. *šen*(f.7) *tb-**eb***(f.5)-***i***(f.4) *ded-is* *mier*
2.SG.NOM warm-THM-PRS.PASS.S.2.SG mother-GEN by
‘You are warmed (by a mother).’

(5) a. *k’ac-eb-**ma***(f.7) *a-u-šen-**es***(f.3) *saxl-**i***(f.8) *megobar-**s***(f.9).
man-PL-ERG PV-IO.3.OV-build-AOR.S.3.S house-NOM friend-DAT

b. *saxl-**i***(f.7) *a-u-šen-**d***(f.1)-*a* *megobar-**s***(f.9)
house-NOM PV-IO.3.OV-build-PASS-AOR.S.3.SG friend-DAT
k’ac-eb-is mier.
man-PL-GEN by
‘A house was built for a friend by men.’

(6) a. *deda-**m***(f.7) *ga-u-zard-a* *švil-eb-**i***(f.8)
mother-ERG PV-IO.3.OV-bring up-AOR.S.3.SG child-PL-NOM
*samšoblo-**s***(f.9).
homeland-DAT
‘A mother brought up children for the homeland.’

b. *švil-eb-**i***(f.7) *ga-**e***(f.1)-*zard-**nen***(f.3) *samšoblo-**s***(f.9)
child-PL-NOM PV-IO.3.PASS-bring up-AOR.S.3.PL homeland-DAT
ded-is *mier*
mother-GEN by
‘the children have been brought up for the homeland by a mother.’

(7) a. *k’ac-**s***(f.7) *a-u-šen-**eb***(f.5)-*i-a*
man-DAT (PV-SINV.3.CV-build-THM-0-OINV.3[SINV.3.SG]).PRF.ACT
*saxl-**i***(f.8)
house-NOM
‘A man has built *a* house.’

b. *saxl-**i***(f.7) *a-šen-**eb***(f.5)-*ul-a* *k’ac-is* *mier.*
house-NOM (PV-build-THM-PRT-BE.PRS.S.3.SG).PRF.PAS man-GEN by
‘A house has been built by a man.’

However, there are a lot of problems in this respect, both from formal and from semantic-functional points of view [3]; all the features taken together obviously distinguish an opposition between the active and the passive constructions, although none of them, taken independently, can be regarded as a simple marker. The following examples illustrate this.

1. The main function of -*s* and -*a* suffixes is to mark S.3.SG. Once this function is identified, examples can be found in various verb forms. For example, -*s* expresses S.3.SG

– in the subjunctive mood of passive verb forms:
  i(PASS)-*xat'* (paint)-*eb*(THM)-*od*(IMP)-*e*(SUBJ)-*s*(S.3.SG) 'it would be painted',
  i(CV)-*dg*(stand)-*e*(SUBJ)-*s*(S.3.SG) 'it would stand'
– in some static verbs:
  zi(sit.PRS)-*s* (S.3.SG) 's/he sits'
– in intransitive-active verbs:
  cxovr(live)-*ob*(THM)-*s*(S.3.SG) 's/he lives'
  pikr(think)-*ob*(THM)-*s*(S.3.SG) 's/he thinks'

The suffix-*a* can be a marker of active verbs' S.3.SG in past tenses as well (see examples (6-a), (7-a)).

2. The main function of -*eb*- is to mark out dynamic verb forms. In expressing this func-tion, -*eb*- also occurs with some active transitive verbs (see examples (2-a), (7-a)).
3. The vowel prefixes, too, are polyfunctional: In general, they represent derivational changes of verb valency – either the increase or decrease of syntactically linked verb arguments [1]. For instance, -*i*- expresses such categories as, e.g.:

   – the subjective version:
     i(SV)-*c'er*(write)-*s*(S.3.SG) 's/he writes smth. for him/herself'
     i(SV)-*šen*(build)-*eb*(THM)-*s*(S.3.SG) 's/he builds smth. for her/himself'
   – the reflexive:
     i(CV)-*ban*(wash)-*s*(S.3.SG) 's/he takes a bath'
     i(CV)-*p'ars*(shave)-*av*(THM)-*s*(S.3.SG) 'he has a shave'
   – potentials (see examples (10));
   – deponents (see examples (8));
   – the future tense of some intransitive-active verbs:
     [*i*(CV)-*cxovr*(live)-*eb*(THM)]:FUT-*s*(S.3.SG) 's/he will live'
     [*i*(CV)-*mγer*(sing)-*eb*(THM)]:FUT-*s*(S.3.SG) 's/he will sing'.[4]

4. The Nominative case characterizes subjects of some non-conversive passives (both, dynamic and static) in present and in past tense forms as well; for example,

   is(s/he.NOM.SG) *dg*(stand)-*a*(PRS)-*s*(S.3.SG) 's/he stands'
   is(s/he.NOM.SG) *i*(CV)-*dg*(stand)-*a*(PST.S.3.SG) 's/he stood'

[4] For the polyfunctionality of the *i*-prefix see [2].

is(s/he.NOM.SG) *gd*(lay.strewn)-*i*(PRS)-*a*(S.3.SG) ‘s/he lies strewn’
is(s/he.NOM.SG) *e*(CV)-*gd*(lay.strewn)-*o*(PST.S.3.SG) ‘s/he lay strewn’.

The Ergative (or the Dative) case can also mark the subjects of intransitive, yet active, verbs which show active, dynamic processes; for example:

man(s/he.ERG.SG) *i*(CV)-*cxovr*(live)-*a*(AOR.S.3.SG) ‘s/he lived’
man (s/he.ERG.SG) *i*(CV)-*pikr*(think)-*a*(AOR.S.3.SG) ‘s/he thought’
mas (s/he.DAT.SG) [*u*(SINV.3.CV)-*cek’*(dance)-*v*(THM)-*i*(0)]:PRF.ACT-*a*(OINV.3. [SINV.3. SG]) ‘(supposedly) s/he has danced’
mas(s/he.DAT.SG) [*u*(CV)-*muš*(work)-*av*(THM)-*i*(0)]:PR-*a*(OINV.3[S.INV.3. SG]) ‘(supposedly) s/he has worked’.

Georgian morphosyntactically marked passive constructions do not always express the conversion of corresponding active ones; in fact they can express a variety of meanings. The following examples illustrate this point.

Active semantics:[5]

(8) a. *k’ac-i* *e-kač-eb-a* *magida-s.*
man-NOM CV-tug-THM-S.3.SG table-DAT
‘A man tugs hard at a table.’

b. *k’ac-i* *a-c’v-eb-a* *mankana-s.*
man-NOM CV-push-THM-S.3.SG car-DAT
‘A man pushes a car.’

c. *k’ac-i* *i-gin-eb-a.*
man-NOM CV-curse-THM-S.3.SG
‘A man curses.’

Dynamic actions:

(9) a. *gogo* *adre* *dg-eb-a.*
girl.NOM early stand-THM-S.3.SG
‘A girl gets up early in the morning.’

b. *bič’-i* *advil-ad* *tvr-eb-a.*
boy-NOM easy-ADV get_drunk-THM-S.3.SG
‘A boy gets drunk easily.’

c. *gogona* *mze-ze* *tb-eb-a*
little_girl.NOM sun-on warm-THM-S.3.SG
‘A little girl gets warm in the sun.’

[5] Verbs having passive form, but active semantics (so-called deponents) are analyzed in [11].

Potentials:

(10) a. *es* *xorc-i* *ar* *i-č'm-eb-a*
this meat-NOM not CV-eat-THM-S.3.SG
'This meat is not edible.'

b. *ak* *c'q'al-i* *i-sm-eb-a*
here water-NOM CV-drink-THM-S.3.SG
'Here the water is drinkable.'

c. *es* *adgil-i* *karg-ad* *i-k'itx-eb-a*
this place-NOM good-ADV CV-read-THM-S.3.SG
'This place is well-readable.'

Reciprocals:

(11) a. *deda* *švil-s* *e-tamaš-eb-a*
mother.NOM child-DAT CV-play-THM-S.3.SG
'A mother plays with her child' = 'they play together.'

b. *k'ac-i* *kal-s* *e-cek'v-eb-a*
man-NOM woman-DAT CV-dance-THM-S.3.SG
'A man dances with a woman' = 'they dance together.'

c. *bič'-i* *megobar-s* *e-č'idav-eb-a*
boy-NOM friend-DAT CV-wrestle-THM-S.3.SG
'A boy wrestles with a boy' = 'they wrestle with each other.'

The verbs without an active counterpart can produce corresponding active-transitive semantics only by special derivational paradigms. Their base forms are semantically intransitive verb forms, while for conversive-passives the base forms are transitive ones.[6]

Thus, the problem with the paradigms is that it is difficult to predict the meaning from the form and/or vice versa; and, to such an extent, the traditional terms 'Active' and 'Passive' actually have only a conventional character. It is obvious that the question as to what the real function of the morphosyntactically differentiated paradigms is still needs to be answered. These formal paradigms cannot be interpreted simply, and their semantic and/or functional analysis definitely requires further investigation.

[6] For some structural features of the so-called passive forms and their semantic interpretations see [7].

2 A Continuum of Active~Passive Opposition

In many languages, as in Georgian, the active ~ passive constructions do not always simply express syntactically defined converses, and the passive formal paradigm is also used to mark other related constructions. In general, there are languages in which the passive formal paradigm marks, for instance, reflexives and reciprocals in Russian and deponents in Latin. In some languages, this paradigm goes further and expresses other grammatical relations as well. In Japanese, for example, it is the formal representation for potential, honorific, and spontaneous voice; moreover, in some other languages it is used for plurals as well [9]. Consequently, attempts at new theoretical approaches were undertaken to explain such cross-linguistic phenomena.

Shibatani's [9, 10] interpretation seems to be efficient from this point of view. He considers the active ~ passive opposition as a continuum, where polar dimensions fit in with prototypical active and passive constructions, while non-polar, inter-medial constructions share only some semantic-categorical features of the categories characteristic for the prototypical ones.

Languages apply various strategies for the formal representation of such non-polar, medial forms; they either develop new formal paradigms, or come to an optimal decision and choose from the existing ones a paradigm that is conventionally regarded as the most appropriate and conceptually proximate according to certain semantic-categorical features (Fig. 1).

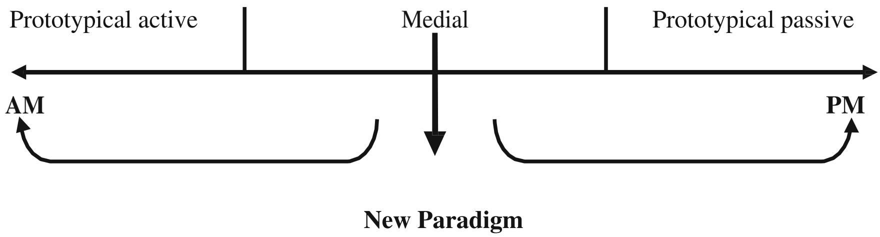


Fig. 1. A continuum of active ~ passive opposition

In such cases, it is much more difficult, and sometimes even impossible, to come up with simple, either functional (due to the changes of syntactic functions) or semantic (due to the distinguishing active ~ passive semantics), taxonomic interpretations concerning the formal paradigms.

3 The Georgian Active~Passive Continuum

Georgian active ~ passive opposition could be interpreted as a continuum where a prototypical active corresponds to active transitive constructions with effected objects (including causatives), which are represented by the active paradigm. Prototypical passives correspond to conversive forms of an active construction, in turn represented by the passive paradigm. All other verbs can be qualified as medial ones.

Medial verbs expressing static events (e.g., *q'ri-a* 'lie.scattered/strewn-s.3.SG', *peni-a* 'is.spread.out-s.3.SG', *k'idi-a* 'is.hanging.on-s.3.SG', *c'eri-a* 'is.written-s.3.SG', *xat'i-a* 'is.drawn-s.3.SG', *abi-a* 'is.tied.(on)-s.3.SG', etc.) show an auxiliary conjugation in present tense.

(12)
me *v-gd-i-v-ar*
1.SG.NOM S.1-lie_about-PRS-S.1-be.PRS.SG
šen *gd-i-x-ar*
2.SG.NOM S.2-lie_about-PRS-S.2-be.PRS.SG
is *gd-i-a*
3.SG.NOM S.3-lie_about-PRS-S.3-be.PRS.SG
'I/you$_{SG}$ /s(he)/it is/are lying about'
čven *v-gd-i-v-ar-t*
1.PL.NOM S.1-lie_about-PRS-S.1-be.PRS.PL
tkven *gd-i-x-ar-t*
2.PL.NOM S.2-lie_about-PRS-S.2-be.PRS.PL
isini *gd-i-an*
3.PL.NOM S.3-lie_about-PRS-S.3-be.PRS.PL
'we/you$_{PL}$/they are lying about'

Thus, they introduce the new paradigm expressing specific static morphology; that is, the paradigm with auxiliary conjugation. These empirical data indicate that the opposition **'Dynamic-Static'** takes a distinct role in the process of formal representation of medial forms. All other verbs (including medial ones) representing non-static, dynamic events instead of the auxiliary conjugation show specific thematic markers in I-series forms (see line 5 in Table 1) and follow either the A- or P-paradigm.

So, the process of the grammaticalization of medial verbs can be explained by the following general cognitive tendency:

In the process of the formal representation of medial forms, Georgian establishes the new paradigm for static verbs, while for other medial dynamic forms applies either the active or the passive formal paradigm. The strategy of choice is defined by the specific conventionally accepted linguistic 'decision' as to which categorical-semantic features of the prototypical constructions are regarded as central.

In order to demonstrate such categorical-semantic features, we must take into account the following linguistic empirical facts observed during the process of the formal representation of some intransitive, dynamic medial forms:

If a medial, prototypically non-active and/or non-passive, verb is ***telic***,[7] *it selects the passive formal paradigm. If a medial verb is* ***atelic***, *then it chooses the active formal paradigm of representation.*

[7] The feature 'telicity' was used by Dee Ann Holisky [6] for some intransitive-active verbs in Georgian, but we suppose that it is decisive for the whole process of formal representation of an active ~ passive continuum.

A preverb[8] could be considered as a marker of telicity, and a formal representation of the above choice becomes fairly simple:

If a verb with medial semantics can take at least one preverb showing perfective aspect, then the verb has the 'passive form.'

Compare, for instance, the examples in (13) to the ones in (14):

(13) *dg*(stand)-*eb*(THM)-*a*(S.3.SG) 's/he is getting up'
a(PV:FUT)-*dg*(stand)-*eb*(THM)-*a*(S.3.SG) 's/he will stand up'
gada(PV:FUT)-*dg-eb*(THM)-*a*(S.3.SG) 's/he will stand elsewhere'
c'ar(PV:FUT)-*dg-eb*(THM)-*a*(S.3.SG) 's/he will step forward';
e(CV)-*mal*(hide)-*eb*(THM)-*a*(S.3.SG) 's/he is hiding from smth./smb.'
da(PV:FUT)-*e*(CV)-*mal-eb*(THM)-*a*(S.3.SG) 's/he will hide from smth./smb.';
a(CV)-*c'v*(press)-*eb*(THM)-*a*(S.3.SG) ' s/he is pressing down'*mi*(PV:FUT)-
a(CV)-*c'v-eb*(THM)-*a*(S.3.SG) 's/he will push against smth./smb.'
da(PV:FUT)-*a*(CV)-*c'v-eb*(THM)-*a*(S.3.SG) 's/he will lie down on smth./smb.'

(14) *cxovr*(live)-*ob*(THM)-*s*(S.3.SG) 's/he lives'
pik'(think)-*ob*(THM)-*s*(S.3.SG) 's/he thinks'
arseb(exist)-*ob*(THM)-*s*(S.3.SG) 's/he/it exists'
k'ank'al(shiver)-*eb*(THM)-*s*(S.3.SG) 's/he hivers'
gor(roll)-*av*(THM)-s(S.3.SG) 's/he/it rolls'
suntk(breathe)-*av*(THM)-*s*(S.3.SG) 's/he breathes'

The first set of medial verbs is telic. All these verbs (having preverbs) follow the P-paradigm. The second set of atelic verbs (not having preverbs) follows the A- paradigm.

4 The Hierarchically Organized Dynamic Paradigm

Linguistic representations of active, passive, and medial verb forms can be reinterpreted as a hierarchically organized cognitive process that defines the choices of either the active (AM) or passive (PM) formal paradigms. The **decision** of which paradigm will be the most appropriate one for the concrete medial verb semantics is taken step by step, conventionally based on the optimal cognitive interpretations originating from some crucial semantic features.

[8] In Georgian so-called preverbs are preverbal affixes that show a direction/orientation of an action sometimes producing new semantics of a verb as well. Additionally, they form the future tense for transitive and conversive-passive verb forms as well as the perfective forms [8]. Inasmuch as the telicity is the property of a verb or verb phrase that presents an action or event as being complete in some sense (resp. perfective), preverbs formally represent telicity as well. Thus, telic verbs can distinguish the opposition between perfective and imperfective aspect represented in Georgian by preverbs, while for atelic verbs this is semantically excluded.

Step 1: Prototypically active and prototypically passive relations are represented by the main formal paradigms, respectively, by the A- and P-paradigms.

Step 2: Medial (non-prototypical) relations are marked according to two different strategies:

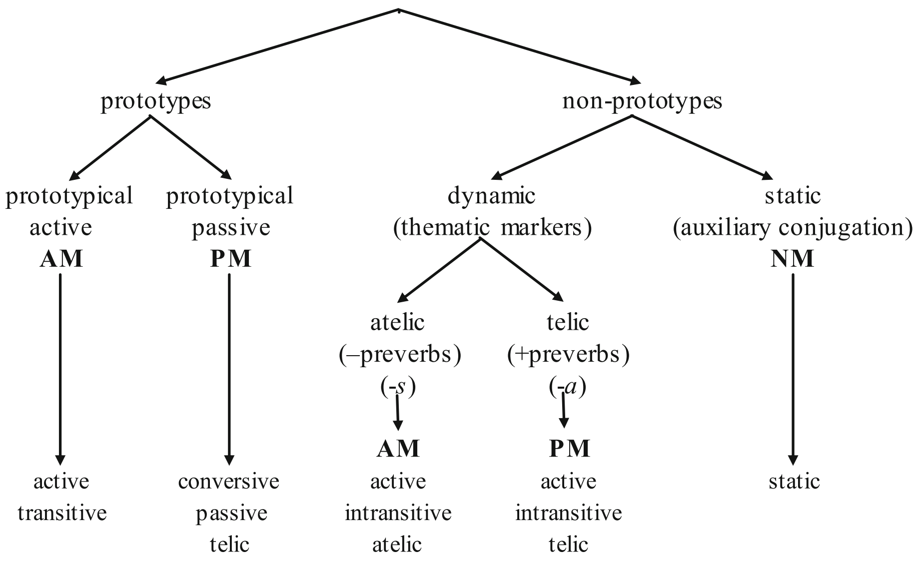
Strategy 1. The new paradigm (NM) is established.
Strategy 2. Either active or passive paradigms of representation are chosen

Further specific cognitive processes and semantic features define which of the above strategies is chosen. First of all, the feature **Dynamic~Static** plays a crucial role – verbs expressing **static** states are marked according to the strategy 1, and the new paradigm of conjugation with the auxiliary verb *to be* is chosen. On the other hand, verbs expressing **dynamic** action choose either the active or passive formal paradigm of representation (Strategy 2).

Step 3: For the **dynamic** subgroup, further choices are determined by the semantic feature **Telicity:**

Telic medial verbs choose the passive paradigm, atelic medial verbs - the active paradigm.[9]

We can represent the process as a productive-generative tree-structure.



The given classification of verbs that is based on the prototypical approach differs from the traditionally distinguished verb classes, according to which class I contains active transitive verbs (including derived causatives), class II – dynamic-passives,

[9] In other words, the morphosyntactic features are absolutely identical with the morphosyntactic features characteristic for the transitive and/or conversive-passive verb forms.

Table 2. Semantic interpretation of traditionally distinguished verb classes

Verb classes	Dynamic	Telic	Transitive
I-class	+	+	+
II-class	+	+	–
III-class	+	–	–
IV-class	–	–	–

class III – active-intransitive verbs, and class IV – affective and static verbs.[10] These four classes can be described and reinterpreted by the semantic features of Transitivity, Dynamicity, and Telicity as well.

However, such an interpretation does not explain why some dynamic and telic verbs actually expressing active semantics (see above examples (8)-(9)-(11)) follow the P-paradigm. The distinguishing feature cannot be either Transitivity or Telicity and/or Dynamicity. It is possible to add the fourth semantic feature of Activity, which makes the semantic analysis more sufficient. In such a case, the conversive-passives defined by the combination {intransitive, telic, dynamic, passive} would be opposed to other "passives" expressed by the combination {intransitive, telic, dynamic, active}; but this "addition" does not resolve the formal problems inasmuch as different combinations of the features would be yet again represented by the same P-paradigm. The prototypical approach is more efficient from this point of view. According to it, II class verbs are divided into two subclasses and the process of formal representation is hierarchically organized:

On the first stage, conversive-passives (being –[TRANSITIVE]) are formalized by P-paradigm, while on the second and the third stages all other, so-called medial, forms (also being –[TRANSITIVE]) are checked according to the features Dynamicity and Telicity as was described above. Thus, given prototypical approach resolves the problem of A-/P-paradigms choices and to such an extent it seems more appropriate.

The whole process of formal paradigm choices is governed by a general, cognitively defined 'conventional linguistic decision':

> *The definite paradigm representing some core semantics serves better to represent certain marginal semantics as well.*

5 The Algorithm Defining the Choice of 'Active' and 'Passive' Formal Paradigms

The whole process can also be reinterpreted and represented as an algorithm which describes the dynamic process of formal paradigm choices.

A generative strategy is based on the decisive semantic features defining the choice: Dynamicity, Telicity, and Aorist. Instead of the feature Transitivity we prefer to describe the whole process by semantic roles of verb arguments inasmuch as the

[10] See [8] and compare with [5].

semantic roles are decisive for Transitivity[11] as well as for arguments case patterns, which are also very important for the definition of A-/P-paradigms. Case patterns – either the Nominative, or the Ergative, or the Dative constructions – are distinguished on the basis of possible case marking combinations of arguments (see features 7-8-9 in Table 1) that are actually realized and conditioned in Georgian by the categories of tense, aspect, and mood.[12] For the binary organization of the algorithm we apply the feature Volitionality,[13] which is regarded as the crucial, main one for describing and distinguishing the semantic roles: An argument whose volition is included in an event is pointed out as +[VL] (resp. Ag → S), while an argument whose 'VL' is not included in an event is pointed out as −[VL] (resp. Ad, Rec, Benef → IO). As far as an argument is concerned that is semantically 'undergoer' and does not exist independently of an event (or not at all) [4] (resp. Patient → DO), the feature VL seems to be redundant for it; that is, it might be structurally qualified as an argument with a priori *zero* VL. Thus, it is pointed out as Ø.

An algorithm with the four stages of implicational rules mirrors the hierarchically organized optimal generative process of linguistic structuring of an active ~ passive continuum in Georgian as follows.

I. {
- −[DYNAMIC] → Nominative constr., static morphology (IV class verbs)[15]
- −[VOLITIONAL] → Dative case
- {(−[VL]), Ø} → Nominative constr., passive morphology (II class verbs)
- {−[VL]$_{EXP}$ (Ø)} → Dative constr., static or passive morph. (IV class$_{AFFECTIVE}$)
- +[VL] → {
  - **II.** Ø → **III.** {
    - +[AORIST] → Ergative, active morph. (I class)
    - −[AORIST] → Nomin., active morph. (I class)
  - **IV.** {
    - +[TELIC] → Nomin., passive morph. (II class)
    - −[TELIC] → **III.** (III class)

Certain combinations of rival features are formalized according to the following strategy:

1. *−[DYNAMIC] (resp. static) verbs follow the passive paradigm without any restrictions (I stage); also, no restrictions are decisive for −[VOLITIONAL] arguments which are always represented by the dative case. Affective verbs with experiencer A-argument always are characterized by the dative construction as well, while their morphology could be either the passive or static (I stage).*
2. *+[DYNAMIC] verbs must be necessarily checked:*
   a. *Do they equate with events including Ø-argument or not? (II stage)*

---

[11] That is, if the arguments structure of a verb includes a patient, a verb is transitive, if not, intransitive.

[12] A case pattern, in actual fact, is the main syntactic feature defining the grouping *screeves* into I, II and III series; see [8] and especially [5].

[13] For the feature VL see [5, 6].

b. *Do they represent events in Past (resp. +[AORIST]) or not? (III stage)*
c. *Do they represent events or actions as being complete in some sense (resp. + [TELIC]) or not? (IV stage)*
d. *In case they are –[TELIC], the algorithm recursively returns to III stage, continuing the checking according to the feature AORIST.*

The algorithm stops after the application of one additional rule:

A verb which chooses the ergative construction with the active morphology must choose the dative construction with the active morphology for III-series TAM forms as well.

6 Conclusions

The proposed analysis is an attempt to describe Georgian verb forms within the cognitive, dynamic approach.

- The active∼passive opposition is reinterpreted as a continuum;
- The continuum is structuralized according to choices between the formal paradigms;
- The choices are defined as a dynamic process;
- The process is based on a conventional cognitive decision;
- The decisions are formulated as implicational rules
- The rules are determined by the semantic features Dynamicity, Volitionality, Telicity, and Aorist.

Glossary

0: zero

1: 1st person

3: 3rd person

2: 2nd person

ACT: Active

ADV: adverbial case

AOR: aorist

CV: characteristic vowel

DAT: dative

DO: direct object

ERG: ergative

FUT: future

GEN: genitive

IMP: imperfect

IO: indirect object

NOM: nominative

NV: neutral version

OINV: inverted object

OV: objective version

PASS: passive

PL: plural

PRF: perfect

PV: preverb

PRS: present

PST: past

PRT: participle

S: subject

SG: singular

SINV: inverted subject

SUBJ: subjunctive

SV: subjective version

THM: thematic suffix

References

1. Asatiani, R.: zmnur p'repiksul xmovanta punkcionaluri k'valipik'acia kartvelur enebši (The Functional Qualification of Verbal Vowel Prefixes in the Kartvelian languages). Macne, Tbilisi (1987)
2. Asatiani, R.: Conceptual structure of reflexive and middle. In: Proceedings of 4th International Symposium on Language, Logic and Computation, pp. 5–16. ILLC scientific publications, Amsterdam (2001)
3. Asatiani, R.: The information structure and typological peculiarities of the georgian passive constructions. In: Bezhanishvili, G., Löbner, S., Marra, V., Richter, F. (eds.) TbiLLC 2011. LNCS, vol. 7758, pp. 17–34. Springer, Heidelberg (2013). doi:10.1007/978-3-642-36976-6_4
4. Dowty, D.: Thematic proto-roles and argument selection. Language **67**(3), 547–619 (1991). Linguistic Society of America
5. Harris, A.C.: Georgian Syntax. Cambridge University Press, Cambridge (1981)

6. Holisky, D.A.: Aspect and Georgian Medial Verbs. Caravan Books, Delmar (1981)
7. Ivanishvili, M., Soselia, E.: A morphological structure and semantics of the georgian so-called passive forms. In: de Jongh, D., Zeevat, H., Nilsenova, M. (eds.) Proceedings of the 3rd and 4th International Symposium on Language, Logic and Computation, Batumi, Georgia, 12–16 September (1999). Borjomi, Georgia, September 23–28 (2001)
8. Shanidze, A.: kartuli enis gramat’ik’is sapudzvlebi (Fundamentals of the Georgian Language Grammar). TSU Press, Tbilisi (1973)
9. Shibatani, M.: Passives and related constructions: a prototype analysis. Language **61**(4), 821–848 (1985)
10. Shibatani, M.: On the Conceptual Framework for Voice Phenomena. Linguistics **44**(2), 217–269 (2006)
11. Tuite, K.: Deponent Verbs in Georgian. In: Bublitz, W., Boeder, W., von Roncador, M., Vater, H. (eds.) Philologie, Typologie und Sprachstruktur: Festschrift Für Winfried Boeder zum 65. Geburtstag, pp. 375–589. Peter Lang Verlag, Frankfurt am Main (2002)

Towards Verb Modification in Frames
A Case Study on German *Schlagen* (to hit)

Anja Goldschmidt[1(✉)], Thomas Gamerschlag[2], Wiebke Petersen[3], Ekaterina Gabrovska[2], and Wilhelm Geuder[2]

[1] UiL OTS, Utrecht University, Utrecht, The Netherlands
A.Goldschmidt@uu.nl
[2] CRC 991, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[3] Department of Linguistics and Information Science, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Abstract. *Hit*-verbs have three basic meaning components, namely movement, contact and force (e.g. [12], Levin 1993), which interact with the verbs' argument structure in various ways. In this paper, we map out the different grammatical constructions of the German verb *schlagen* (usually, though loosely, translated as 'hit'; also 'beat', 'strike') and their restrictions on agentivity and the force component. Using modification by pure manner adverbs as a tool to test for possible default values of the force component, and agent-oriented adverbs to discover possible interactions with agentivity, we show that German *schlagen* is rather liberal with respect to its force component. Crucially, the force component may not only be modified by standard, force-denoting manner adverbs such as *lightly* and *hard*, but also through agent-oriented adverbs such as *playfully*, via a defeasible inference. We show further that our findings can be profitably modelled in Frame Semantics, a framework which is especially well suited for modelling a fine-grained decomposition of word meaning, including the manner-related components of verbs.

Keywords: Hit-verbs · Modification · Force · Agentivity · Frame semantics

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) with a grant to the Collaborative Research Centre (SFB) 991 "The Structure of Representations in Language, Cognition, and Science", Projects B09 and C10 (Sebastian Löbner and Wiebke Petersen). Additionally, part of the work was funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) (to Anja Goldschmidt). The final version of the paper has profited from the comments of two anonymous reviewers, and from discussion with Sebastian Löbner on the frame modelling. We also thank the audience at the 11th TbiLLC Symposium for their feedback on an earlier version. We would like to acknowledge the help of the Utrecht Institute of Linguistics' (UiL OTS) statistics advisor, Kirsten Schutter-Pessoa, with the questionnaire data. The order of the authors is chronological, as they joined in at different stages of the project.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 18–36, 2017.
DOI: 10.1007/978-3-662-54332-0_2

1 Introduction

In the literature on English verb classes, one class has been especially discussed by syntacticians as well as semanticists: 'verbs of contact by impact'. The name of this class of verbs can be traced back at least to Levin (1993) [12], although already Fillmore (1970) [5] discusses the verb *hit* from this perspective. In most of the analyses of *hit*, the meaning components 'motion', 'contact' and 'force' are identified as basic (cf. [12], Levin 1993, [6], Gao and Cheng 2003, among others).

Hit-verbs are described by Levin (1993:150) [12] as involving the movement of one entity leading to contact with another entity. Although Levin does not speak explicitly of a force component associated with these verbs, the choice of the name for the class, 'contact by impact', clearly indicates that she takes the presence of a (high) force to be crucial, especially as there is also a class of 'verbs of contact' ([12], Levin 1993:155f). Furthermore, authors like Erteschik-Shir and Rapoport (2010:59) [4] a.o. use notions like "forceful contact" in analyzing verbs like *hit*. Gao and Cheng (2003:494) [6] also observe that English verbs of contact by impact have a force component which "is specified in all the verbs" as again already indicated by the name of the class. What is more, the authors also state that actions referred to as hitting are characterized by the exertion of high force ([6], Gao and Cheng 2003:494). This force needs a source, which, on a cognitive linguistic view, is typically the agent represented by the subject in English ([10], Kim 2009:46f). Besides the source or the subject, there is also a patient/an object receiving the force, which makes *hit* a standard case of a transitive verb (cf. [10], Kim 2009:50; [5], Fillmore 1970:128).

When looking for 'contact by impact' verbs in German, the most prominent representative is the verb *schlagen*, which is commonly treated as the translational equivalent of *hit*, but may often correspond more closely to English verbs like *beat*, *strike*, *knock*, *deliver a blow* (which are usually neglected in the literature on *hit*). There are five relevant constructional variants of *schlagen* (other variants as intransitive or particle verbs may have additional meaning components or figurative meanings) which can be differentiated (following Geuder and Gabrovska, ms. [7]):

Unaccusative construction: A theme argument (i.e., a moving entity, esp. in ballistic movement) is realized as subject, typically combined with a PP that encodes a target.

(1)	Die	Gitarre	schlug	gegen	die	Tischkante.
	The	guitar	hit(PAST)	against	the	edge.of.the.table
	'The guitar hit the edge of the table.'					

Agentive-resultative construction: An agent argument is realized as subject co-occurring with an accusative object introducing the theme and a resultative predicate (often a directional PP).

(2)	Er	schlug	die	Gitarre	gegen	die	Tischkante
	He	hit(PAST)	the	guitar	against	the	edge.of.the.table
	'He hit the guitar against the edge of the table.'						

Simple transitive construction: Prototypically an agentive subject plus a patient (i.e. receiver of a blow, frequently animate), realized as accusative object. A transitive construction also occurs with a number of idiomatic meanings which we do not consider here.

(3) Wenn ein Bauer einen Esel hat, dann schlägt er ihn
If a farmer a.ACC donkey has then beats he him
'If a farmer has a donkey, then he beats it.'

Agentive oblique construction: Agentive subject, with adverbial goal complement (mostly PP).

(4) Er schlug (mit der Faust) auf den Tisch.
He hit(PAST) with the fist on the table
'He hit the table (with his fist).'

Double complement construction: Agentive subject, with accusative[1] or dative patient plus PP goal complement.

(5) a. Er schlug mich auf den Rücken
He hit(PAST) me.ACC on the back
b. Er schlug mir auf den Rücken
He hit(PAST) me.DAT on the back
'He hit me on the back.'

In this paper we concentrate on the first and the third of these variants, where either two inanimate and therefore non-agentive entities are involved (first variant) or two animate entities, including a volitional agent (third variant). These variants represent the two most extreme cases with respect to agentivity.

Based on a questionnaire study we will show that in German, the meaning component 'force' is not always specified with a high value, contrary to what is assumed for English *hit*. We propose that *schlagen* comes with a force attribute which can receive any value: a high one (prototypical in the simple transitive construction), or a low one (not prototypical in any construction, but possible with all of them). As will be seen, the value of the force attribute is subject to influences from adverbial modification and contextual inference.

The paper is structured as follows: In Sect. 2, we will explain how we use adverbial modification to tease apart the meaning components of *schlagen*. Furthermore we present and analyse the empirical data from the questionnaire study, which illustrate the behavior of *schlagen* with respect to the force component in the above constructions. In Sect. 3, we will give a first sketch of a frame semantic analysis of *schlagen* along the lines of e.g. [14], Petersen 2015, integrating the basic meaning components 'force' and 'motion/contact' and the argument structure of the verb, as well as illustrating the effects of adverbial modification.

[1] Following Vogel (2016) [17], the variant with an accusative object plus a goal PP can be subsumed here as an extended version of the simple transitive construction.

2 *Schlagen* and the Force Component

2.1 Adverbial Modification as a Tool to Tease Apart Verb Meanings

The German verb *schlagen* can appear in a number of grammatical constructions (cf. Sect. 1). In order to map out its behaviour with respect to the force component, we conducted a questionnaire study focusing (i) on the unaccusative construction, involving an inanimate entity as theme argument in subject position and a target-encoding PP (cf. 6), and (ii) on the simple transitive construction with an animate agent in subject position and an animate patient as accusative object (cf. 7). These two variants represent the two most extreme cases with respect to agentivity (cf. Sect. 1). We chose them in order to explore effects of agentivity on the force component.

(6) Die Gitarre schlägt gegen die Tischkante.
the guitar hits against the edge.of.the.table
'The guitar hits the edge of the table.'

(7) Sophia schlägt Simon.
Sophia hits Simon
'Sophia hits Simon.'

Since *hit*-verbs are generally described as 'verbs of forceful contact', one might assume that the two variants should display the same force feature. On the other hand, the construction in (7) is said to encode an action that especially affects the object (more detailed discussion in [7], Gabrovska and Geuder ms.), hence a difference in strength or kind of force might ensue. The first thing to test, therefore, is whether the force feature is uniformly present with agentive and non-agentive constructions.

We propose that questions like this can be addressed by examining the patterning of modifiers that occur with a verb. In this paper, we use adjectives (in adverbial function, which in German remains morphologically unmarked) whose lexical meaning specifies features like force and agentivity. We presume that modifiers can be divided, just like verb meanings, along these lines into those that are specified for agentive traits and others that target a pure force feature (Schäfer 2013 [16] speaks of a class of 'pure manner adverbs'). Hence, we can compare modifier-verb pairs of the type *hart/leicht schlagen* (hit hard/lightly) — in which only the force feature should be addressed — with constructions like *spielerisch schlagen* (hit playfully) in which an agentive feature of the verb is addressed, since playing requires an agent. We expect compatibility restrictions that derive from the semantic representations, for instance, the agentive modifier *spielerisch* (playfully) should be unable to occur with the unaccusative construction (cf. 6).

A more tricky question is what to expect of modifiers that target different values on the forcefulness scale, i.e. *hart/leicht*. It is immediately apparent that modifiers which denote low force are in principle compatible with *hit*-verbs, although standard descriptions of the verb meaning do not really seem to leave room for this. This leads us to the question of prototypical expectations, and generally speaking, to the role of inferencing in the interpretation of modifiers. Hence, over and above simple semantic compatibility, we want to test whether verb-modifier pairs lead to preferential assumptions in the course of interpretation (which can be overridden). A tool to test this is the so-called denial-of-expectation construction with the conjunction *but* (cf. [11], Lakoff 1971).

In order to address these questions, we conducted a questionnaire study that examined the following contrasts. First, we are interested in the behaviour of the pure manner adverbs:

Question A: Do the constructional variants of *schlagen* differ with respect to the meaning component "forceful contact" — especially in the sense that the transitive construction is specified for high force?

Question B: Do the constructional variants give rise to inferences about forces in different ways?

These questions can be tested in terms of the following hypotheses.

2.2 Hypotheses

Firstly, in the denial-of-expectation construction, of the pure manner adverbs we expect only *leicht* (lightly) in a simple transitive sentence to be acceptable, i.e. no expectations that can be contrasted through the use of *aber* (but) should arise in unaccusative sentences. The only expectation that we assume to arise is that of high force in simple transitive sentences. This expectation can then felicitously be "denied" by the use of *leicht* in combination with *aber* (cf. Table 1).

Table 1. Contrasts to be investigated with pure manner adverbs *leicht* (lightly) and *hart* (hard)

	(A) testing for semantic compatibility of modifiers	(B) testing for a (default) expectation
Agentive, transitive cases	Sie schlägt ihn √leicht/√hart She hits him √lightly/√hard	Sie schlägt ihn, aber √leicht/??hart She hits him, but √lightly/??hard
Non-agentive, unacc. case	Die Gitarre schlägt √leicht/√hart gegen die Tischkante The guitar hits the edge of the table √lightly/√hard	Die Gitarre schlägt gegen die Tischkante, aber ??leicht/??hart The guitar hits the edge of the table, but ??lightly/??hard

Secondly, we examine the behaviour of *spielerisch* (playfully) as an example of an agent-oriented modifier. While the reference to agentivity should be part of its semantic representation, we presume that *spielerisch* is also able to indicate a low amount of force used when combined with *schlagen*. This leads to the more specific question of whether this "force effect" is a feature of the semantic representation or an inference triggered by meaning components of other types. That the reference to agentivity is part of the semantic representation of *spielerisch* can easily be confirmed with our method:

Agentivity Hypothesis: Agent-oriented manner adverbs such as *spielerisch* (playfully) can only apply to *schlagen* in the case of an animate agent in the simple transitive construction (and not in the unaccusative construction with a non-agentive theme as subject).

The predictions that can be derived from the agentivity hypothesis, illustrated with some example sentences, can be found in Table 2 (the sentences with *leicht* (lightly) are given for comparison, they should be acceptable with both constructions).

Table 2. Predictions derivable from the agentivity hypothesis

Simple transitive construction	Unaccusative construction
√ Sie schlägt ihn spielerisch	# Die Gitarre schlägt spielerisch gegen die Tischkante
√ She hits him playfully	# The guitar hits the edge of the table playfully
√ Sie schlägt ihn leicht	√ Die Gitarre schlägt leicht gegen die Tischkante
√ She hits him lightly	√ The guitar hits the edge of the table lightly

Moreover, we are testing the assumption that the indication of low force in the case of *spielerisch* is a defeasible inference:

Force Inference Hypothesis: Modifiers of the type of *spielerisch* (playfully), when combined with *schlagen*, have an effect on the force component of *schlagen*, i.e. indicate a low value. However, this is an inferential process, and hence defeasible.

For the test sentences relating to the force inference hypothesis, we again make use of the denial-of-expectation construction with *aber* (but). The idea is that an adverb such as *spielerisch*, which results in a force decrease inference, cannot be opposed to an adverb that also indicates a decrease of the force magnitude such as e.g. *leicht* (lightly). Since both modify the force magnitude in the same direction, they should not be contrastable in a denial-of-expectation construction with *aber* (but). However, since *spielerisch* triggers a defeasible and hence cancellable inference on the force component, it should be acceptable with an adverb modifying the magnitude of the force in the opposite direction, e.g. *hart* (hard). In this case, the inference should be cancelled. The predictions derivable from the force inference hypothesis are given in Table 3.

Table 3. Predictions derivable from the force inference hypothesis

Contrast in opposite direction	√ Sie schlägt ihn spielerisch, aber doch recht hart
	√ She hits him playfully, but still rather hard
Contrast in same direction	?? Sie schlägt ihn spielerisch, aber doch recht leicht
	?? She hits him playfully, but still rather lightly

2.3 Questionnaire Design and Materials

The questionnaire comprised 95 test sentences, distributed over seven questionnaires á 21–22 sentences, including two control sentences that were direct contradictions (e.g. *hit hard and lightly*). The sentences were randomized, and all questionnaires were distributed among German native speakers in two versions, one of which contained the test sentences in reversed order.

The sentences had to be rated on a 4-point Likert scale, where a 4 means "clearly good", a 3 "maybe good", a 2 "maybe bad" and a 1 "clearly bad". This way, speakers were forced to make a commitment as to whether a sentence was more on the acceptable side or more on the unacceptable side. The rating task was preceded by an introduction, which included an example sentence from an unrelated domain (speed) and asked speakers to rate sentences according to their first intuition. Following the rating task, information about speakers' language background was collected via four questions relating to their language(s) and place(s) they have been raised/lived.

15–20 participants were tested for each version of all seven questionnaires. Participants who rated either of the direct contradictions in the two control sentences higher than 1 were excluded from the analysis, as were participants whose native language was not German. 165 participants in total were included in the analysis.

2.4 Data and Results

An overview of the results can be found in Table 4.

At first glance, all of our expectations have been confirmed. The sentences testing expectations arising about the force magnitude in either simple transitive construction or unaccusative construction, making use of pure manner adverbs *hart* (hard) and *leicht* (lightly) as well as the contrastive conjunction *aber* (but), have received a visibly lower percentage of ratings 3 "maybe good" and 4 "clearly good" than their counterparts without *aber* (65% and 44% vs. > 90%). The exception are sentences of the type *Sie schlug ihn, aber leicht* (She hit him, but lightly), which were judged just as good as their counterparts without *aber* (100% vs. 93%).

This confirms our prediction that *schlagen* in transitive construction with an animate agent prototypically denotes high force, and that no such default interpretation is available for *schlagen* in the unaccusative construction with an inanimate entity as subject.

It also seems true that adverbs of the type *spielerisch* (playfully) can only be used to modify *schlagen* if the verb appears in a simple transitive construction with an animate agent, and not if it is used in the unaccusative construction (88% ratings 3 & 4 vs. 32%).

Finally, we can see that sentences of the type *spielerisch, aber doch recht hart* (playfully, but hard) receive much higher ratings than sentences of the type *spielerisch, aber doch recht leicht* (playfully, but lightly) (76.5% ratings 3 & 4 vs. 28%). This shows that modifiers of the type of *playfully* do indeed result in

Table 4. Percentages of ratings 3 "maybe good" and 4 "clearly good" for all hypotheses (observed, not estimated)

Hypothesis	Example sentences	%
Force expectations transitive case	Sie schlägt ihn, aber leicht	100%
	She hits him, but lightly	
	Sie schlägt ihn, aber hart	65%
	She hits him, but hard	
	Sie schlägt ihn leicht	92.9%
	She hits him lightly	
	Sie schlägt ihn hart	95%
	She hits him hard	
Force expectations unaccusative case	Die Gitarre schlägt gegen den Tisch, aber leicht	65.8%
	The guitar hits the table, but lightly	
	Die Gitarre schlägt gegen den Tisch, aber hart	44.1%
	The guitar hits the table, but hart	
	Die Gitarre schlägt leicht gegen den Tisch	90.4%
	The guitar hits the table lightly	
	Die Gitarre schlägt hart gegen den Tisch	92,5%
	The guitar hits the table hard	
Agentivity hypothesis	Sie schlägt ihn spielerisch	88%
	She hits him playfully	
	Sie schlägt ihn leicht	89.5%
	She hits him lightly	
	Die Gitarre schlägt spielerisch gegen den Tisch	32%
	The guitar hits the table playfully	
	Die Gitarre schlägt leicht gegen den Tisch	87%
	The guitar hits the table lightly	
Force inference hypothesis	Sie schlägt ihn spielerisch, aber doch recht hart	76.5%
	She hits him playfully, but still rather hard	
	Sie schlägt ihn spielerisch, aber doch recht leicht	28%
	She hits him playfully, but sill rather lightly	

a force decrease inference, as they can felicitously be contrasted with *hard* (a force increasing adverb), but not with *lightly* (a force decreasing adverb).

In order to test whether the observed results are significant, we have run a general linear mixed effects model for all hypotheses. Unfortunately, we had too few observations to be able to create a general linear mixed effects model testing the force expectations in the transitive case (cf. first hypothesis in Table 4).

The model testing the force expectations in the unaccusative case (cf. second hypothesis in Table 4), while confirming the general trend observable in Table 4 above, shows that the differences are not significant: the odds of rating a sentence with contrastive *but* 3 (maybe good) or 4 (clearly good) are 0.042 times (*but hard*, p = .06) and 0.116 times (*but lightly*, p = .21) the odds of rating a sentence without contrastive *but* 3 or 4. That means there is a trend that participants liked sentences with *but* less than sentences without *but*, i.e. there don't seem to be any expectations about the force magnitude that can be contrasted ("denied") with *hard* or *lightly* in the unaccusative construction with an inanimate entity as subject.

The model for the agentivity hypothesis yields significant results: the odds of giving a sentence with *schlagen* ratings 3 (maybe good) or 4 (clearly good) are significantly higher for a sentence in unaccusative construction with a force-related adverb such as *lightly* ($p < .001$) or for a sentence in simple transitive construction with any adverb ($p = .002$): more than 30 times the odds of giving a sentence with *schlagen* in unaccusative construction and with an adverb of the type *playfully* ratings 3 or 4. I.e. participants mostly did not accept agent-oriented manner adverbs with inanimate subjects in unaccusative construction.

Lastly, the model for the differences in rating for the force inference hypothesis, while confirming the trend observable in Table 4, also does not show significant results: the odds of giving ratings 3 (maybe good) or 4 (clearly good) for sentences of type *playfully, but lightly* are 0.063 ($p = .057$) times the odds of giving sentences of type *playfully, but hard* ratings 3 or 4. That means that sentences of the type *playfully, but lightly* are less acceptable than sentences of the type *playfully, but hard*, which provides evidence for the prediction that the low force reading of *playfully* is a cancellable inference.

To sum up: while we have evidence for all our predictions, it seems that the observations about expectations relating to the magnitude of the force of *schlagen* in transitive and unaccusative constructions (without modification) are not as strong as expected. On the other hand, the force decrease inference that playfully-type adverbs trigger when combined with *schlagen* is clearly observable in our data (only just not significant). And it is very clearly the case that agent-oriented manner adverbs cannot combine with *schlagen* in the unaccusative construction with an inanimate theme in subject position.

In the next section, we will present a model of *schlagen* in the framework of Frame Semantics à la Petersen (2015) [14], which can integrate these findings about *schlagen* in both constructions (simple transitive and unaccusative), as well as explicitly model its other meaning components (movement and contact).

3 A Frame Semantic Model of *hit*-verbs

In this section, we discuss the modelling of our findings in a Frame representation, a relational model of conceptual structure that is built on functional attributes ([14], Petersen 2015). In contrast to lexical decomposition models that focus on event structure (ultimately elaborating on the insights of Dowty, 1979 [2]), a Frame model is able to include a detailed analysis of the manner component of a verb's meaning and the way it relates to arguments, including implicit arguments, of the event. The manner component of *schlagen* will be characterised as based on notions of force exertion. As already pointed out by Levin (1993) [12], however, movement is another component that has to be factored in. Hence, the meaning of *schlagen* will involve at least two entities, dubbed here THEME and RECIPIENT, and a movement of the theme towards the recipient (or: 'target', 'patient'), leading up to contact, the whole process being marked by a notion of force transmission.

There are various different scenarios of *schlagen* that would require variants of the representation, but for our purposes here, we concentrate on a single prototypical case in which a number of parameters is fixed that would have to be variable to yield a fully general account. It is possible to integrate a frame model in a compositional semantics with a fully-developed syntax-semantics interface, but we are bypassing this aspect for simplicity (see e.g. [9], Kallmeyer & Osswald 2013, for compositional aspects of frame theory). Our main goal with the following model is merely to sort out which attributes are basically involved in the concept *schlagen*, how they are interrelated, and how, in principle, modifiers are able to create the effects demonstrated above in our empirical study.

The final result will be a complex frame (cf. Fig. 1). In the following, we will discuss each part of the frame in some more detail, and then explain the effects of the various modifiers.

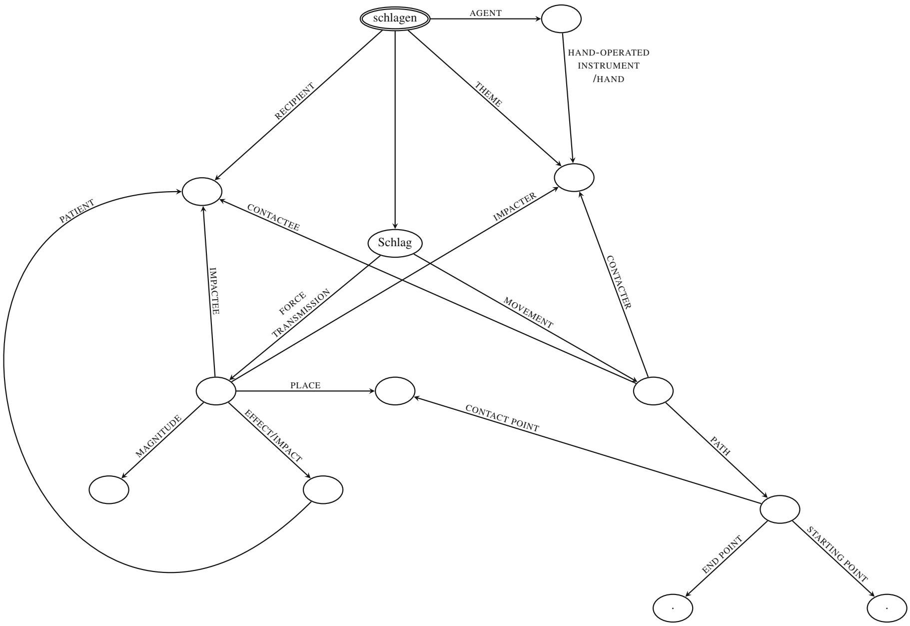


Fig. 1. The complete frame of the verb *schlagen* (to hit)

3.1 Argument Roles

At a coarse level of analysis, we can identify the scene as an interaction between individuals (but note that this view will be refined presently). As pointed out in the introduction, different uses of *schlagen* differ in the way how arguments are realised. It is also possible to gather more arguments than were seen in our standard examples (8a/b), cf. (8c), which may indicate that sometimes implicit arguments have to be taken into account in a full representation even of simple examples like (8a):

(8) a. Sophia schlägt Simon. [Agent, Recipient/Patient]
Sophia hits Simon
b. Die Gitarre schlägt gegen die Tischkante.
The guitar hit (against) the edge.of.the.table
[Theme, Recipient/Target]
c. Katja schlägt die Flasche gegen den Tisch.
Katja hit the bottle against the table
[Agent, Theme, Recipient/Target]

A role label like 'agent' can directly be used as an attribute in the sense of Frame theory – its status as a functional notion is already evident in the standard event-semantic notation agent(e,x) (cf. [13], Parsons 1990), which means a mapping of events onto individuals. In the graphic frame representation (adopted in [14], Petersen 2015) attributes are shown as labels of the arcs of a graph, and their values as nodes (cf. Fig. 1). Hence, the attribute AGENT leads to a node that introduces the relevant individual. Due to the recursive nature of the attribute-value structure used in Frame Semantics, more information can be added as a next step, e.g. if the agent also controls an instrument, is in a particular intentional state in the event, etc. Conversely, the same individuals can be the value of other attributes, too. This latter case becomes important as soon as an event description is more finely decomposed: the classic thematic roles may in fact sum up information from different aspects of the description (mirroring a set of "proto-role entailments" in the sense of Dowty 1991 [3]). This is why we set up the participant roles as a separate array, beside the core description of the event. The thematic roles are focal parts of the representation but not primitives: they can be linked in various ways to different parts of the decomposed event description. In the frame fragment in Fig. 2, AGENT and THEME are shown as arguments of *schlagen* (the central node). But THEME has at the same time incoming arrows from (i.e., is the value of) diverse attributes, stating that the moving object may simultaneously also be an instrument in the sense that it is manipulated by an agent ("hit the table with the bottle").

We will next specify the movement/contact and force transmission components of *schlagen* in the frame.

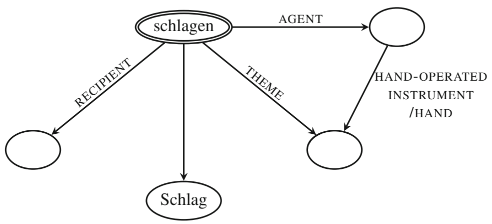


Fig. 2. The argument roles in the frame of *schlagen*

3.2 Components of the Event Description

In line with standard views of *hit*-verbs, we distinguish between a 'movement' component and a 'force transmission' component in their description. The illustration in Fig. 3 shows these as attributes of the node "Schlag" (hit); the values in each case are events. The force-transmission component specifies force-related attributes of the participants (in terms of the roles of IMPACTER and its force-dynamic antagonist, called IMPACTEE). Similarly, the movement component assigns its own semantic roles which we have dubbed here, with some amount of foresight, 'contacter' and 'contactee'.

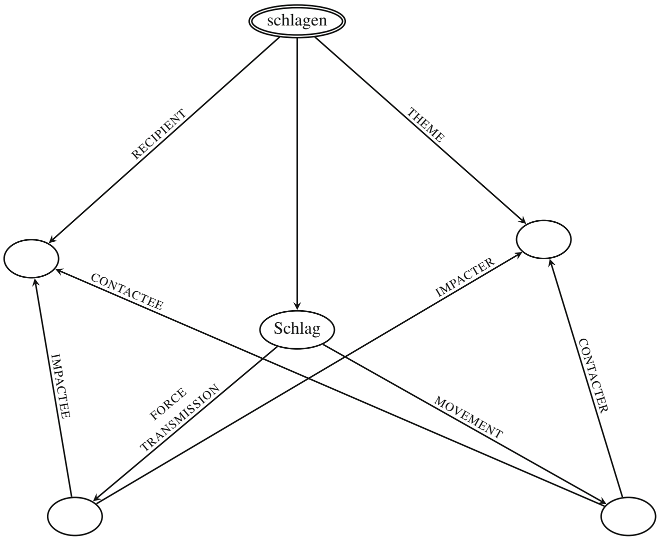


Fig. 3. The argument roles of the MOVEMENT and FORCE TRANSMISSION components

The contactee role derives from the presence of a reference object that serves to localise the movement path, as in "hit the bottle *against the table*". However, in order to simplify the discussion, we do not represent the semantic composition of *schlagen* with prepositional phrases, rather, we present the movement node as having already inherited all the relevant information about a movement that leads up to contact. The resulting network of attributes now says that the moving object is at the same time the IMPACTER, the source of a forceful impact, and that the goal of the movement is the IMPACTEE, the target of the impact. As shown before, the moving object may also be under the control of an agent in some particular scenario, making it also the agent's instrument.

The two subcomponents MOVEMENT and FORCE TRANSMISSION can now be considered in more detail (cf. Fig. 4).

MOVEMENT is described in terms of the two arguments just shown and the PATH. The path is described as a linear order of points in space, as is standardly done (e.g. [18], Zwarts 2005). Some kind of path is always present due to the verb meaning, irrespective of the addition of PPs in the syntax. Its linear ordering of points specifies a designated starting point and endpoint, among other things, which we can encode as attributes of the path. Furthermore, there is one point on the path on which an impact takes place, i.e. a force transmission event between impacter and impactee. In the typical case, this would be the endpoint (as e.g. specified in the path description "against the table"), but our representation leaves it open in principle whether this or some other point of the path will be identified with the value of the attribute CONTACT POINT.

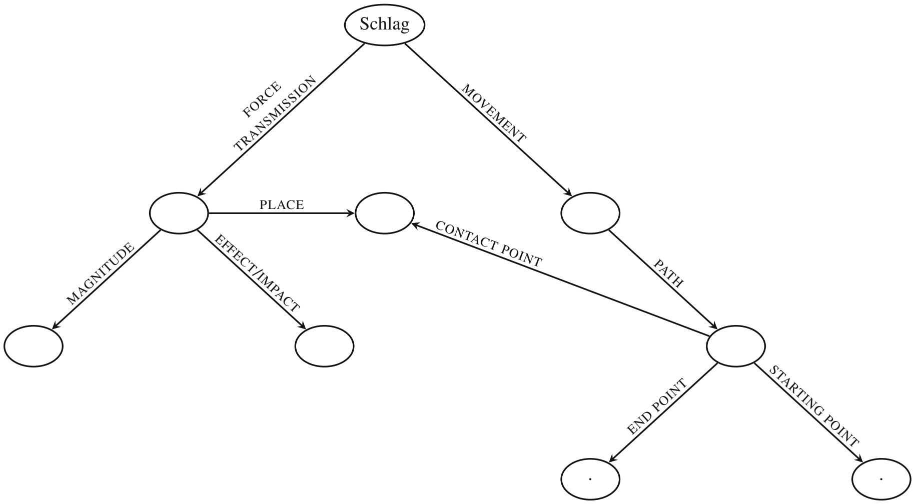


Fig. 4. The MOVEMENT and FORCE TRANSMISSION components

On the left hand side of the representation in Fig. 4, the force transmission event is described in terms of the attributes PLACE (of impact, as just mentioned), potential causal effects, the participant roles (cf. Fig. 3), and, as we additionally assume here, a measure function that directly maps the force transmission event onto a value for the magnitude/strength of its impact (which has to be distinguished from the question of effects such as doing damage or not). This, in sum, is a preliminary proposal for encoding the distinction of the two domains force dynamics and spatial description in *schlagen* events, and their interaction. Let us now consider how the functioning of modifiers could be understood on such a basis.

3.3 Modification

In general, (manner) modification in frames can be understood as a mechanism that narrows down admissible values of the attributes in an event description, but it may also lead to the addition of new attributes (a sketch of adverbial modification in frames can be found in [8], Geuder 2006). The process is driven by the lexical semantics of the modifiers, i.e. the adjectives that underlie adverbial forms. Additionally, however, many effects of manner modification are not due to hard-wired semantic features, but to inferential processes. This could be observed in Sect. 2 above, where we showed that adverbial modifiers like *spielerisch* (playfully), when combined with *schlagen*, lead to a defeasible inference of low force. Hence there are two things that we want to explain here: How do we get from the lexical meaning of a modifier to the effect of manner modification, and how do we distinguish between inferred and hard-wired effects of modification?

For reasons of space, we will only look into modification with *spielerisch* in some detail. For cases like *hart schlagen* (hit hard) let us simply point out that the lexical meaning of this adjective indeed seems to consist in a specification of a (high) force value of some impact, and remain vague about potential further meaning components, like the kind of interaction of two surfaces or materials (but especially German *hart* seems to suggest some specific kind of mechanical interaction of two non-elastic bodies, which may be less prominent in the case of its English cognate). Hence, we have to formulate a rule such that the modifier *hart*, by its lexical meaning, interacts with the attribute MAGNITUDE (of a force transmission) so as to restrict the set of feature values admissible here.[2] In other words, it is a subsective modifier acting on a feature set (instead of on an extension of a predicate, as it would be in a neo-Davidsonian framework; for some more details we refer the reader to [8], Geuder 2006).

A more tricky case is the adverb *spielerisch*. We have seen that it presupposes an animate agent (cf. the agentivity hypothesis in Sect. 2). Furthermore, it does have an effect on the magnitude of the force posited for the hit, but only a defeasible one, in contrast to *hart*. So we conclude, in the first place, that it contributes a property of the event's agent, and the force-dynamic effects must be inferred on the basis of the adjective's lexical representation and its effect on the network of attributes in the frame of *schlagen*. Hence, we now need a second frame representation of the adjective in order to combine it with the event frame, and a preliminary inspection of corpus data shows us that this adjective is highly variable in its meaning as a modifier. We will therefore confine ourselves to formulating an approximation of the meaning that it assumes in the present context together with the verb *schlagen*, without attempting a more generalisable lexical representation. In the context of *spielerisch schlagen* (hit playfully), the outcome is obviously the description of an action that is a hitting, but one that is not an attack and is not marked by an intention to harm the patient. Rather, it transports a communicative intention to evoke the possibility of a real, aggressive hit which constitutes a kind of joke. Hence, we want to take serious the meaning of the stem *play* that is present in the adjective. Playing is arguably an activity that takes place in the real world but whose relevance and goals reside in a fictional representation. In this way, playing is often (though not always) an activity that simulates something else. Here, the overall contribution

[2] Technically, a feature can be defined as an attribute-value pair.

of this use of *spielerisch* involves an intention not to produce the full, real-life effects of the event that is modified by this adjective, i.e. the “Schlag” or hit. What the adjective as such is about is rather that the agent intends a “play” consisting of the simulation of a hit.

Therefore, we propose that the simplified representation in Fig. 5 contains the essential aspects of this modifier meaning.

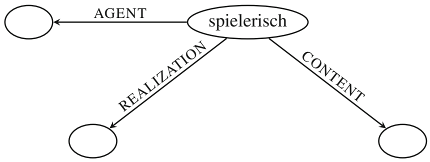


Fig. 5. A frame representation for *spielerisch* (playfully)

First of all, we assume that *spielerisch* selects an argument which is an agent. Thus, *spielerisch* always implies that some agent is playing something. Secondly, we assume an argument role, dubbed here REALISATION, which introduces the activity that is really performed in playing. And finally, there is the aspect that playing aims at a fictional sphere, this is what the activity “means”, shown here as the CONTENT attribute. Now, let us combine this representation with the event frame for *schlagen* (cf. Fig. 1) and see what we get from it (cf. Fig. 6).

Resolving the agentivity condition in the modification looks straightforward: The player must be identified with the hitter. Second, the activity that constitutes the play is the hitting (the REALISATION attribute). When we say *jemanden spielerisch schlagen* (to hit somebody playfully), then in this case *spielerisch* seems to point to a communicative act, a joking activity. Therefore, for this specific case we take the CONTENT feature of the play to be the aggressive act, which is communicated as an absurd possibility and therefore as a friendly joke. (In other contexts, the application of the basic concept of playing may lead to different results). What we get for our specific case is that the value of the attribute CONTENT of *spielerisch* is another event description of the type *Schlag*, but this time the fully-fledged, aggressive one (which, as just said, is being simulated by the playful hitting). Therefore, it has all the attributes that are seen in the main part of the frame (for reasons of space not represented twice in Fig. 6, i.e. the empty node of the CONTENT attribute is taken to be the whole frame of *schlagen* again, from the node ‘Schlag’ downwards). If the hit that really occurs is a playful version of a fully-fledged hit, we can posit a correlation: we can reasonably expect that the force magnitude of the playful hit will be lower than that of the simulated fully-fledged hit. This use of correlations between values of certain attributes is part and parcel of frame theory (cf. Barsalou 1992 [1] who introduces correlation as a technical term) and plays an important role throughout in the analysis of manner modification (as also pointed out in [8], Geuder 2006; see also [15], Petersen & Gamerschlag 2014, for a slightly different application). Such

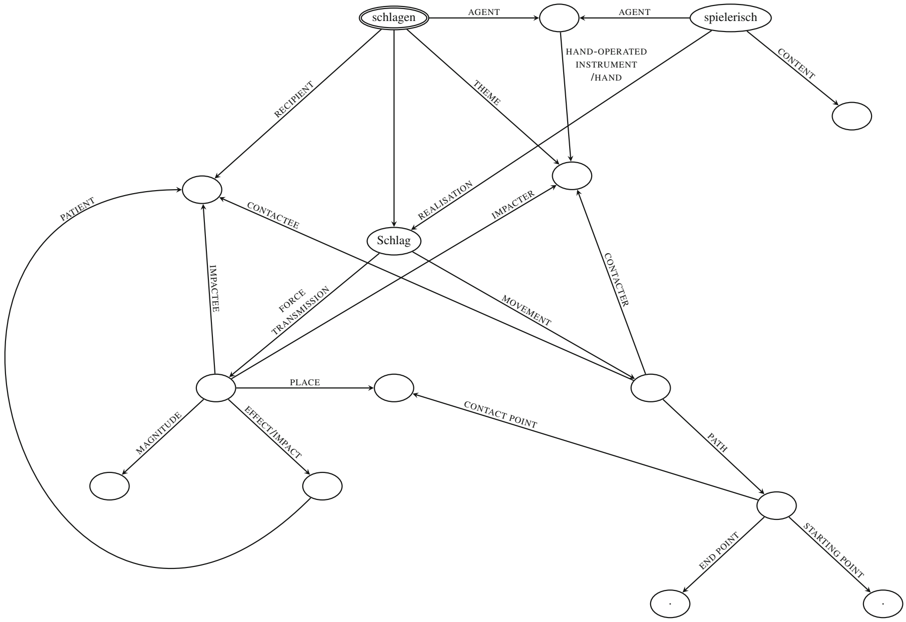


Fig. 6. Combination of the frames for *spielerisch* (playfully) and *schlagen* (hit)

correlations can be due to strict laws (of nature, for instance), but others can also be typical correspondences which are defeasible. We are dealing here with the latter case. It is simply the felicitous course of events if the playful hitting does not do any harm. In our questionnaire study above, it was demonstrated that this is a defeasible inference.

Hence *spielerisch* as a manner modifier works differently from *hart* as a manner modifier, and we have now shown the reason for this: *hart* applies directly to the force magnitude to change its value. In contrast, *spielerisch*, while it does have an influence on the same value, does so only via a prototypical correlation. The parts of the frame that it directly applies to are different ones than those that *hart* directly applies to. However, the frame representation also shows the way in which the correlation plays itself out: this is the link (arc) from *spielerisch* to the node "Schlag", stating that the hit is realised as a play.

This concludes our sketch of how Frame theory is applied to modelling adverbial modification, both with respect to the semantic representation and to the explanation of inferential effects. Of course our present account only presents a single and fairly narrow case study, but the mechanisms demonstrated here can be exploited in a more general way. And while we have glossed over a number of finer points, these could be integrated thanks to the flexibility of the framework that allows one to 'zoom in' and add more details. For example, we have left implicit the semantic intuition that the hitting in our example would have to be an intentional hitting by the agent, and that likewise playing is an intentional activity; but such points could be added in more fine-grained versions.

4 Summary and Conclusion

In this paper, we have presented a couple of observations about the force component and modification of German *hit*-verbs, and how these can be modelled within Frame Semantics.

Firstly, in Sect. 2, we showed how the denial-of-expectation test with *but* can be used to test native speakers' expectations about certain defaults. We were able to show that when *schlagen* (hit) is used in the unaccusative construction with an inanimate theme in subject position, no expectations arise as to the magnitude of the force (cf. the relative oddness of *Die Gitarre schlägt gegen die Tischkante, aber leicht/hart* (The guitar hits the edge of the table, but lightly/hard)). But when *schlagen* is used in the simple transitive construction with an animate agent and patient, there is a tendency to expect the hitting to be done with high force (cf. the acceptability of the sentence *Sie schlägt ihn, aber leicht* (She hits him, but lightly)). Thus, we have shown that German *schlagen* does not lexically specify a high amount of force.

Secondly, we were interested in the interaction between the force component and two distinct types of modifiers, pure manner adverbs such as *hart/leicht* (hard/lightly) on the one hand, and agent-oriented adverbs such as *spielerisch* (playfully) on the other hand. Crucially, both are able to modify the force component of the verb, though they do so via different mechanisms. When *schlagen* is combined with modifiers such as *spielerisch* (playfully), a defeasible inference arises that the hitting was done with little force. This was again tested through the use of *but* in denial-of-expectation construction (cf. the force inference hypothesis in Sect. 2). However, this effect is only observable for the simple transitive construction, since agent-oriented modifiers of the type of *spielerisch* can only apply to *schlagen* in this construction (with an animate agent, cf. the agentivity hypothesis in Sect. 2).

These observations were modelled within Frame Semantics (cf. Sect. 3), a form of meaning representation based on recursive attribute value structures. Frame Semantics allows to combine the various meaning components of *hit*-verbs, such as force and movement/contact, with the general argument structure of the verb. We showed that the different grammatical constructions of *schlagen* can all be modelled in one frame, and that this mode of representation makes explicit the connections between the lexical content of the verb and other words in the sentence (e.g. the recipient of the hit is also characterised as the contactee of the movement/contact component, the impactee of the force component, and the patient of the force impact). Furthermore, we were able to integrate the frame for *spielerisch* into the verb frame and show explicitly how the defeasible inference about low force is computed, rooting it in the semantics of the modifier itself.

A Appendix — Example Sentences from the Questionnaires

Sentences testing expectations arising about the force magnitude in the transitive and unaccusative constructions:

– Sophia schlägt Simon hart.
  'Sophia hits Simon hard.'
– Chris schlägt Alex leicht.
  'Chris hits Alex lightly.'
– Julia schlägt Tobias, aber hart.
  'Julia hits Tobias, but hard.'
– Tobias schlägt Maike, aber leicht.
  'Tobias hits Maike, but lightly.'
– Die Gitarre schlägt hart gegen die Tischkante.
  'The guitar hits the edge of the table hard.'
– Die Gitarre schlägt leicht gegen die Tischkante.
  'The guitar hits the edge of the table lightly.'
– Der Zweig schlägt gegen die Hauswand, aber hart.
  'The branch hits the wall of the house, but hard.'
– Der Zweig schlägt gegen die Hauswand, aber leicht.
  'The branch hits the wall of the house, but lightly.'
– Die Wellen schlagen hart gegen den Deich.
  'The waves hit the dyke hard.'
– Die Wellen schlagen gegen den Deich, aber leicht.
  'The waves hit the dyke, but lightly.'

Sentences testing predictions of the agentivity hypothesis:

– Andrea schlägt Jan spielerisch auf den Arm.
  'Andrea hits Jan playfully on the arm.'
– Chris schlägt Alex leicht auf den Arm.
  'Chris hits Alex lightly on the arm.'
– Die Gitarre schlägt spielerisch gegen die Tischkante.
  'The guitar hits the edge of the table playfully.'
– Der Zweig schlägt leicht gegen die Hauswand.
  'The branch hits the wall of the house lightly.'

Sentences testing predictions of the force inference hypothesis:

– Andrea schlägt Jan spielerisch, aber doch recht leicht, auf den Arm.
  'Andrea hits Jan playfully, but still rather lightly, on the arm.'
– Andrea schlägt Jan spielerisch, aber doch recht hart, auf den Arm.
  'Andrea hits Jan playfully, but still rather hard, on the arm.'

References

1. Barsalou, L.: Frames, concepts, and conceptual fields. In: Lehrer, A., Kittay, E.F. (eds.) Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization, pp. 21–74. Lawrence Erlbaum Associates Publishers, Hillsdale (1992)
2. Dowty, D.: Word Meaning and Montague Grammar. Reidel, Dordrecht (1979)
3. Dowty, D.: Thematic proto-roles and argument selection. Language **67**(3), 547–619 (1991)
4. Erteschik-Shir, N., Rapoport, T.: Contacts and other results. In: Rappaport-Hovav, M., Doron, E., Sichel, I. (eds.) Syntax, Lexical Semantics, and Event Structure, pp. 59–75. Oxford University Press, Oxford (2010)
5. Fillmore, C.J.: The grammar of hitting and breaking. In: Jacobs, R., Rosenbaum, P. (eds.) Readings in English Transformational Grammar, pp. 120–133. Ginn, Waltham (1970)
6. Gao, H., Cheng, C.C.: Vverb of contact by impact in english and their equivalents in mandarin chinese. Lang. Linguist. **4**(3), 485–508 (2003)
7. Geuder, W., Gabrovska, E.: Verbs and their modifiers - a pilot study on German schlagen, ms. Univ. of Düsseldorf (2016)
8. Geuder, W.: Manner modification of states. In: Ebert, C., Endriss, C. (eds.) Proceedings of Sinn und Bedeutung, vol. 10, pp. 111–124. ZAS, Berlin (2006)
9. Kallmeyer, L., Osswald, R.: Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. J. Lang. Model. **1**(2), 267–330 (2013)
10. Kim, Y.: Event construal and its linguistic encoding: towards an extended semantic map model. Ph.D. thesis, University of Oregon (2009)
11. Lakoff, R.: If's, and's and but's about conjunction. In: Fillmore, C., Langedoen, D. (eds.) Studies in Linguistic Semantics, pp. 114–149. Holt, Rinehart and Winston Inc., New York (1971)
12. Levin, B.: English Verb Classes and Alternations: A Preliminary Investigation. University of Chicago Press, Chicago (1993)
13. Parsons, T.: Events in the Semantics of English. A Study in Subatomic Semantics. MIT Press, Cambridge (1990)
14. Petersen, W.: Representation of concepts as frames. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Meaning, Frames, and Conceptual Representation, pp. 43–67. Düsseldorf University Press, Düsseldorf (2015)
15. Petersen, W., Gamerschlag, T.: Why chocolate eggs can taste old but not oval: a frame-theoretic analysis of inferential evidentials. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Frames and Concept Types. SLP, vol. 94, pp. 199–218. Springer, Heidelberg (2014). doi:10.1007/978-3-319-01541-5_9
16. Schäfer, M.: Positions and Interpretations. German Adverbial Adjectives at the Syntax-Semantics Interface. De Gruyter Mouton, Berlin (2013)
17. Vogel, R.: Optimal constructions. In: Géraldine Legendre, M.P., Zaroukian, E. (eds.) Advances in Optimality Theoretic-Syntax and Semantics. Oxford University Press, New York (2016)
18. Zwarts, J.: Prepositional aspect and the algebra of paths. Linguist. Philos. **26**(6), 739–779 (2005)

Whence Long-Distance Indefinite Readings? Solving Chierchia's Puzzle with Dependent Types

Justyna Grudzińska[1](✉) and Marek Zawadowski[2]

[1] Institute of Philosophy, University of Warsaw, Warsaw, Poland
j.grudzinska@uw.edu.pl
[2] Institute of Mathematics, University of Warsaw, Warsaw, Poland
zawado@mimuw.edu.pl

Abstract. Indefinites (e.g. *a man*, *some woman*) have given rise to a number of puzzles concerning their scopal and dynamic behavior. One such puzzle about long-distance indefinites seems to be unsettled in the literature [3]. In this paper we show how Chierchia's puzzle of long-distance indefinites can be handled in semantics with dependent types. The proposal builds on our formal system combining generalized quantifiers [1,17,22] with dependent types [19,20,24] in [9,11].

Keywords: Quantifier scope · Long-distance indefinite · Dependent type

1 Introduction

It has been observed that indefinites (e.g. *a man*, *some women*) often behave more like referring expressions (e.g. *John*, *Mary*) and differ from the so-called standard quantifier expressions (e.g. *every man, most women*) with respect to their scopal and anaphoric properties. This has led to the abandonment of the uniform treatment of all quantifier expressions in the form of generalized quantifiers [1] and development of a battery of mechanisms for modeling indefinites: individual/plural variables [7,12–14,23], choice/Skolem function variables [15,25,29,32], dynamic existential quantification [8,31]. As will be argued in this paper, adopting a dependent type theoretical approach to generalized quantification allows us to restore some of this lost uniformity. In our previous work, we have defined a new uniform algorithm to account for a wide range of anaphoric (dynamic) effects associated with natural language quantification [9,11]. In this paper, we will show how Chierchia's puzzle about two kinds of long-distance indefinites [3,27,30] can be handled in our semantics with dependent types, without giving up the quantificational treatment of indefinites.

The paper is organized as follows. Section 2 introduces informally the main features of our semantics with dependent types: (i) types, dependent types, and their interpretation; and (ii) the notion of generalized quantification extended

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 37–53, 2017.
DOI: 10.1007/978-3-662-54332-0_3

to dependent types. In Sect. 3, we use an example of the behavior of indefinites in 'donkey anaphora' contexts to briefly sketch what can be achieved with our algorithm for the interpretation of anaphoric (dynamic) effects. Section 4 shows how one can use dependent types to tackle Chierchia's puzzle of long-distance indefinites. We propose to credit the problematic long-distance readings to the presence of (possibly pragmatically induced) dependencies. In the Conclusion, we compare our proposal to the two recent quantificational accounts of indefinites in Schwarzschild [28] and Brasoveanu & Farkas [2]. Appendix I and II provide detailed analyzes of the linguistic examples used in the paper.

2 Dependent Type Semantics

In our previous work, we have developed a new dependent type theoretical semantics for natural language quantification [9,11]. Our approach combines elements from the two semantic frameworks: classical Montague-style semantics and dependent type theories [4,6,18,24]. Like in the classical Montague-style semantics, our approach makes essential use of generalized quantifiers [1,17,22]. But in the spirit of the dependent type theoretical framework, we adopt a many-typed analysis (in place of a standard single-sorted analysis). Like in the dependent type theories, we have type dependency in our system [19,20]. But whereas the existing dependent type theoretical approaches have been proof-theoretic [4,6,18,24], our semantics is model-theoretic with truth and reference being basic concepts. In the following, we will introduce the main elements of our framework.

2.1 Types, Dependent Types and Their Interpretation

The variables of our semantic system are always typed. We write $x : X$ to denote that the variable x is of type X and refer to this as a type specification of the variable x. Types are interpreted as sets. We write the interpretation of the type X as $\|X\|$. Types can depend on the variables of other types. Thus, if we already have a type specification $x : X$, then we can also have type $Y(x)$ depending on the variable x and we can declare a variable y of type $Y(x)$ in the context $x : X$, and form an extended context

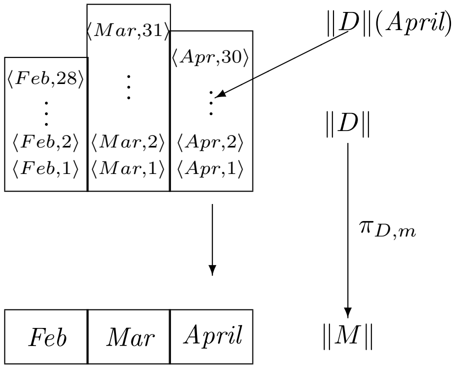
$$x : X, y : Y(x)$$

The fact that Y depends on X is modeled as a function

$$\pi : \|Y\| \rightarrow \|X\|.$$

One example of such a dependence of types is that if m is a variable of the type of months M, there is a type $D(m)$ of the days in that month

$$m : M, d : D(m)$$



If we interpret type M as a set $\|M\|$ of months, then we can interpret type D as a set of the days of the months in $\|M\|$, i.e. as a set of pairs

$$\|D\| = \{\langle a, k\rangle : k \text{ is (the number of) a day in month } a\}$$

equipped with the projection $\pi_{D,m} : \|D\| \to \|M\|$. The particular sets $\|D\|(a)$ of the days of the month a can be recovered as the *fibers* of this projection (the preimages of $\{a\}$ under $\pi_{D,m}$)

$$\|D\|(a) = \{d \in \|D\| : \pi_{D,m}(d) = a\}.$$

Our system makes no use of assignment functions. Variables serve to determine dependencies and contribute to the semantics in an indirect way. Our semantics is defined by directly combining interpretations of quantifier phrases and predicates via some 'algebraic' operations and variables serve as an auxiliary syntactic tool to determine how these operations are to be applied.

2.2 Generalized Quantifiers on Dependent Types

While Montague-style semantics is single-sorted in the sense that it includes one type **e** of all entities (strictly speaking, it has two basic types: type **e** and type **t** of truth values, and a recursive definition of functional types), our semantics is many-sorted in the sense that it includes many basic types. On the Montague-style analysis, a quantifier phrase like *some woman* is interpreted over the universe of all entities E, i.e. *some woman* denotes the set of subsets of E

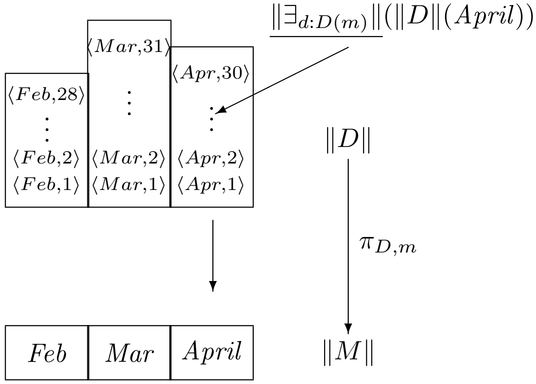
$$\|\exists x : woman\ x\| = \{X \subseteq E : \|woman\| \cap X \neq \emptyset\}.$$

As a consequence of our many-sorted analysis, we have a polymorphic interpretation of quantifiers. On our analysis, a quantifier phrase like *some woman* is interpreted over the type *Woman*, i.e. *some woman* denotes the set of all non-empty subsets of the set of women

$$\|\exists_{w:Woman}\| = \{X \subseteq \|Woman\| :\ X \neq \emptyset\}.$$

As a result of our many sorted-analysis, we also have a polymorphic interpretation of predicates. On our analysis, a predicate like *love* is interpreted over types (e.g. *Man*, *Woman*, ...), and not over the universe of all entities.

The main novelty of our proposal is in combining generalized quantifiers with dependent types and thus introducing quantification over fibers, e.g. existential quantification over the fiber of the days of April $\|D\|(April)$ (as in *some days of April*)



In this sense, fibers are considered 1st class citizens of our semantics, i.e. we allow for quantification over fibers on a par with quantification over sets interpreting any other types (e.g. *Man*, *Woman*, ...). As will become evident below, the fact that we have dependent types and quantification over fibers in our semantics proves crucial to our solutions to some of the key puzzles involving anaphoric and scopal properties of indefinites (unbound anaphora, exceptional scopes).

3 Indefinites and 'Donkey Anaphora'

One puzzle about indefinites relates to the phenomenon of 'donkey anaphora' (an instance of unbound anaphora)

(1) Every farmer who owns a donkey beats it.

On the so-called universal reading, sentence (1) is understood to mean that every farmer who owns a donkey beats every donkey he owns. The sentence is considered problematic, for the indefinite antecedent (*a donkey*) is contained inside a relative clause and the pronoun (*it*) is outside that clause (i.e. the pronoun is not syntactically bound its quantifier antecedent) but is related anaphorically to the antecedent. The problem of 'donkey anaphora' has been dealt with in a number of semantic paradigms (DRT, dynamic semantics, Skolem function accounts, dependent type theoretical approaches). Our analysis of 'donkey sentences' makes crucial use of dependent types and quantification over fibers. As will be briefly explained below, the advantage of our analysis is in allowing a uniform and natural solution to some of the main difficulties surrounding 'donkey sentences' ('proportion problem', ambiguities claimed for 'donkey sentences').

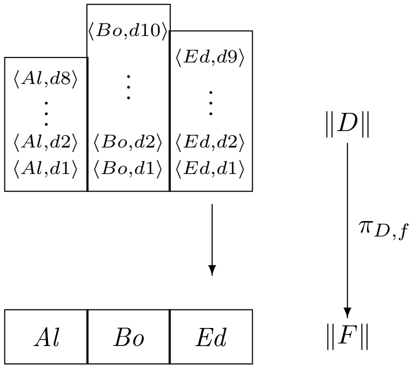
The overall interpretational architecture of our system is two-dimensional [9,11]. The two dimensions to the meaning of a sentence in our system are: the truth value of a sentence and the dynamic effects introduced by the sentence (dynamic extensions of context). Context for us is a sequence of type specifications of the individual variables

$$x : X, y : Y(x), z : Z(x, y), \dots$$

A sentence extends context by some possibly dependent types. In our work, we have defined a new algorithm for the interpretation of the main kinds of dynamic effects (dynamic extensions of context) associated with natural language quantification [9,11]. In the 'donkey anaphora' case, the modified common noun *farmer who owns a donkey* of sentence (1) extends the context by adding two newly formed types (more precisely, by adding new variable specifications on two newly formed types)

$$f : F, d : D(f).$$

The main clause quantifies universally over the interpretations of the respective types (unbound anaphoric pronouns are treated as universal quantifiers in our system)



– the type F interpreted as $\|F\|$ (the set of farmers who own some donkeys),
– the dependent type D interpreted for the farmer a in $\|F\|$ as $\|D\|(a)$ (the set of donkeys owned by the farmer a),

yielding the desired truth conditions: every farmer who owns a donkey beats every donkey in the corresponding fiber of the donkeys owned.

Importantly, this solution does not run into the 'proportion problem'. Since we quantify over farmers and the respective fibers of the donkeys owned (and not over $\langle farmer, donkey\rangle$ pairs), a sentence like *Most farmers who own a donkey beat it* comes out false if there are ten farmers who own one donkey and never beat them, and one farmer who owns twenty donkeys and beats all of them. Furthermore, 'donkey sentences' have been also claimed to be ambiguous between the so-called (i) strong (universal) reading: *Every farmer who owns a* *donkey beats* EVERY *donkey he owns*, and (ii) weak (existential) reading: *Every farmer who owns a donkey beats* AT LEAST ONE *donkey he owns*. Our analysis can accommodate this observation by taking the weak reading to simply employ the quantifier *some* in place of *every*: every farmer who owns a donkey beats at least one donkey in the corresponding fiber of the donkeys owned. Here, we can follow [16] in assuming that in the unmarked case 'donkey pronouns' are 'subject to a maximality constraint' — then, on our analysis, they are treated as universal quantifiers. Pragmatic factors (world knowledge, discourse context), however, can sometimes override this maximality constraint, and then 'donkey pronouns' can be also treated as existential quantifiers. Finally, since our semantics allows for quantification over fibers, our analysis can be extended to account for some more complicated 'donkey sentences' such as *Every farmer who owns donkeys beats most of them.*

4 Indefinites and Chierchia's Puzzle

Yet another puzzle about indefinites concerns their scopal properties and seems to be unsettled in the literature [3,27,30]. The puzzle arises in connection with sentences such as

(2) Every linguist has studied every solution that some problem might have.

Sentence (2) noncontroversially allows the so-called narrow scope reading for *some problem* (for the ease of exposition, here and in the following examples we will use formal representations in typed predicate logic; the syntax of our system is different though — for the details, see Appendix I)

$$\forall_{l:Linguist}\forall_{s:Solution}[\underline{\exists_{p:Problem}}\ \ solves(s,p) \rightarrow studied(l,s)].$$

This narrow scope reading is (pragmatically) implausible, as it is logically equivalent to the reading saying that every linguist l studied every solution s to every problem p

$$\forall_{l:Linguist}\forall_{s:Solution}\forall_{p:Problem}[solves(s,p) \rightarrow studied(l,s)].$$

As discovered by a number of authors ([25,32] among others), sentence (2) also allows the so-called long-distance intermediate scope reading for *some problem* saying that for every linguist l there is a possibly different problem p such that l has studied every solution s to p

$$\forall_{l:Linguist}\underline{\exists_{p:Problem}}\forall_{s:Solution}[solves(s,p) \rightarrow studied(l,s)].$$

This long-distance reading is plausible and intuitively available to many authors but it is also problematic, for on this reading the indefinite *some problem* (unlike standard quantifiers) takes exceptional scope out of its scopal island (relative clause).

One prominent strategy for handling this puzzling scopal behavior of indefinites is Kratzer-Matthewson's proposal where they argue that there is a way of accounting for the problematic readings without assuming that indefinites scope out of their islands [15,21]. Chierchia, however, has pointed out to some examples that prove Kratzer-Matthewson's strategy crucially insufficient [3]. In the following, we will first discuss Kratzer-Matthewson's solution and then explain what has become to be known as Chierchia's puzzle [3,27]. Finally, we will show how one can use dependent types to tackle Chierchia's puzzle and thus answer our title question: whence long-distance indefinite readings?

4.1 Kratzer-Matthewson's Solution

Kratzer in [15] credits the problematic reading in (2) to the presence of a hidden pronoun/functional element, i.e. the apparent long-distance intermediate reading is in fact a functional reading and can be expressed by the following paraphrase

(2a) Every linguist has studied every solution that some problem that intrigued him/her most might have.

The proposed paraphrase can be represented as follows

$$\forall_{l:Linguist}\forall_{s:Solution}[solves(s, \underline{f_{problem}(l)}) \rightarrow studied(l, s)]$$

where $f_{problem}$ denotes a function that maps every linguist l into the problem that intrigues l the most. As can be seen from the above representation, functional readings can be obtained with leaving the functional term representing the indefinite in situ.[1]

One relatively minor problem for Kratzer's solution has to do with her assumption that functional readings only become available when there is a contextually salient function, e.g. the *most intriguing problem* function. Contra to this assumption, it has been pointed out that one can often get a long-distance reading for a paraphrase like (2b), even if one has no clue as to which function is involved

(2b) Every linguist has studied every solution that a certain problem that intrigued him/her might have.

To solve this problem, Matthewson in [21] proposes that the function variable should be closed existentially (at the topmost possible level)

$$\underline{\exists f_{problem}}\forall_{l:Linguist}\forall_{s:Solution}[solves(s, \underline{f_{problem}(l)}) \rightarrow studied(l, s)].$$

Matthewson's reading says that there is a way of associating problems to linguists such that every linguist has studied every solution to the problem he/she is associated with.

[1] More precisely, Kratzer claims that indefinites are ambiguous between a (skolemized) choice function interpretation and a generalized quantifier interpretation. Skolemized choice function indefinites yield functional readings (masquerading as long-distance readings). Quantificational indefinites have only local (clause-bounded) scopes.

4.2 Chierchia's Puzzle

Chierchia in [3] poses a challenge for Kratzer-Matthewson's solution by observing that there is another kind of long-distance readings that cannot be reduced to functional readings

(3) Not every linguist has studied every solution that some problem might have.

Sentence (3) is intuitively true in a situation where for some linguist there is no problem such that he/she has studied every solution that this problem might have. These are the truth-conditions for the negated long-distance intermediate reading, and NOT for Kratzer/Matthewson's readings

– long-distance intermediate reading

$$\forall_{l:Linguist}\underline{\exists_{p:Problem}}\forall_{s:Solution}[solves(s,p) \rightarrow studied(l,s)]$$

– negated version saying that for some linguist l and for every problem p there is a solution s such that s solves p and l did not study s

$$\exists_{l:Linguist}\forall_{p:Problem}\exists_{s:Solution}[solves(s,p) \wedge \neg studied(l,s)].$$

This last representation amounts to the more wieldy paraphrase already mentioned above saying that for some linguist there is NO problem such that he/she has studied every solution that this problem might have. For Kratzer's reading, sentence (3) would claim that some linguist didn't study some solution to the problem that intrigued him/her most

$$\exists_{l:Linguist}\exists_{s:Solution}[solves(s,\underline{f_{problem}(l)}) \wedge \neg studied(l,s)].$$

For Matthewson's reading, sentence (3) would claim that there is a way of pairing of linguists and problems such that some linguist did not study some solution to the problem he/she is paired with

$$\underline{\exists f_{problem}}\exists_{L:linguist}\exists_{s:Solution}[solves(s,\underline{f_{problem}(l)}) \wedge \neg studied(l,s)].$$

Chierchia conlcudes that we need to accommodate both kinds of long-distance readings: functional and long-distance intermediate. So the puzzle is that we seem to need heterogeneous mechanisms (e.g. skolemized choice functions, intermediate existential closure of the choice functions variables) to account for the behavior of long-distance indefinites. Moreover, the mechanisms proposed have been argued to be problematic on both theoretical and empirical grounds (see e.g. [23]).

4.3 Whence Long-Distance Indefinite Readings?

Our proposal distinguishes three kinds of long-distance readings

1. pragmatically induced dependent readings (corresponding to Chierchia's long-distance intermediate readings)
2. functional readings (corresponding to Kratzer's readings)
3. dependent referential readings (corresponding to Matthewson's readings).

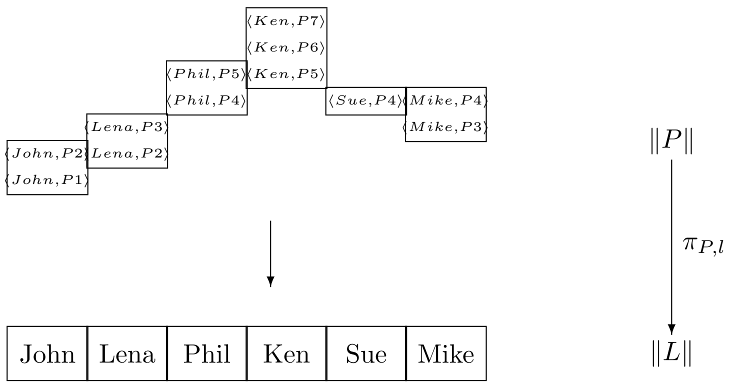
Pragmatically Induced Dependent Readings. We propose to credit Chierchia's long-distance readings to the presence of contextually salient dependencies. Consider an example in (2) again

(2) Every linguist has studied every solution that some problem might have.

On our analysis, Chierchia's long-distance intermediate readings are not semantically generated. As explained in Appendix I, in our semantic system a quantifier expression (indefinite) can never escape a modified common noun such as *solution that some problem might have* — thus our semantics with dependent types can only yield a narrow scope reading for the indefinite *some problem* in (2). The proposal is, however, that the problematic readings can be sometimes pragmatically induced by contextually available dependencies. Here we draw on the observation made by Chierchia himself [3]. As observed by Chierchia, special context is needed to get a long-distance intermediate reading for a sentence like (2). Chierchia gives an example of one such context 'You know, linguists are really systematic: Lee studied every solution to the problem of weak crossover, Kim every solution to the problem of donkey sentences, etc.'

Suppose now that some such context can make one posit certain dependencies, in that case that the type of problems depends on (the variable of) the type of linguists and the type of solutions depends on (the variables of) the type of linguists and the type of problems

$$l : L; p : P(l); s : S(l, p)$$



By quantifying over dependent types posited in this way, we get the pragmatic dependent reading (≃ Chierchia's reading) saying that for every linguist a in $\|L\|$ there is at least one problem b in the corresponding fiber $\|P\|(a)$ such that a has studied every solution c in $\|S\|(a, b)$

$$\forall_{l:L} \exists_{p:P(l)} \forall_{s:S(l,p)} l \textit{ has studied } s.$$

Notice that the dependencies available in the context force the indefinite *some problem* to take scope over *every solution* — in that case, the pragmatic logical form OVERRIDES the sentence's semantic logical form. The negative sentence would then claim that for some linguist a in $\|L\|$ there is NO problem b in the corresponding fiber $\|P\|(a)$ such that a has studied every solution c in $\|S\|(a,b)$

$$\exists_{l:L}\forall_{p:P(l)}\exists_{s:S(l,p)} l \textit{ did not studied } s.$$

Thus the strategy proposed allows us to account for Chierchia's truth-conditional intuitions regarding negative sentences such as (3).

Our proposal perhaps can be further strengthened by the examples of the following kind

(4) Every accomplished food critic reviewed every restaurant that some millionaire visited.

(5) Every accomplished food critic reviewed every restaurant that some millionaire owned.

Our intuitions are that the dependent reading ($\simeq$ Chierchia's reading) and its corresponding negative version are hard to get with the example in (4), as it is rather difficult to come up with a plausible context where the type of millionaires depends on (the variable of) the type of food critics and the type of the visited restaurants depends on (the variables of) the type of millionaires and the type of food critics. This contrasts with the example in (5) where the dependent reading and its corresponding negative version should be much more readily available, as it is more natural to assume that the type of the owned restaurants depends on (the variable of) the type of millionaires and it is also much easier to come up with a plausible scenario where the type of restaurants depends on (the variables of) the type of millionaires and the type of food critics (say, Food Critic 1 has been assigned the task of reviewing all of the restaurants owned by Millionaire A, Food Critic 2 has been assigned the task of reviewing all of the restaurants owned by Millionaires B and C, etc.). A number of researchers have observed that the availability of long-distance intermediate readings is a gradient effect rather than all or nothing ([3,25,32] among others). Since our proposal takes such readings to be dependent on the context, we have a ready explanation for this observation.

Functional and Dependent Referential Readings. Now, what about Kratzer-Matthewson's readings? Our proposal distinguishes functional and dependent referential readings. Consider first an example of Kratzer's functional reading in (2a)

(2a) Every linguist has studied every solution that some problem that intrigues him/her most might have.

On our view, sentence (2a) yields a functional reading saying that every linguist a has studied every solution to $f(a)$. For our analysis of function terms, see Appendix II. To explain our account of dependent referential qua Matthewson's readings, we need to say something more about indefinites.

Unlike standard quantifiers (e.g. *every student*, *most students*), indefinites (e.g. *a student*, *three students*) have been observed to exhibit ambiguity between the so-called general (quantificational) and specific (referential) reading. To give an example, I can make a general claim using a sentence *I have been friends with some student* and my use of the indefinite will not imply that I am thinking about any particular student. But I can also use the same sentence to make a specific claim and my use of the indefinite will introduce some particular one student that I have in mind. To handle this ambiguity in indefinites, we introduce into our system two kinds of type-assignment. An indefinite, e.g. *some student*, is ambiguous allowing a combination of the determiner *some* and either:

– the variable of the 'standard' type *Student*, interpreted as the set of all students (given in the context) — $\|Student\|$, or
– the variable of the 'referential' type *Student**, interpreted as a certain set containing a single student that the speaker has in mind — $\|Student^*\|$.[2]

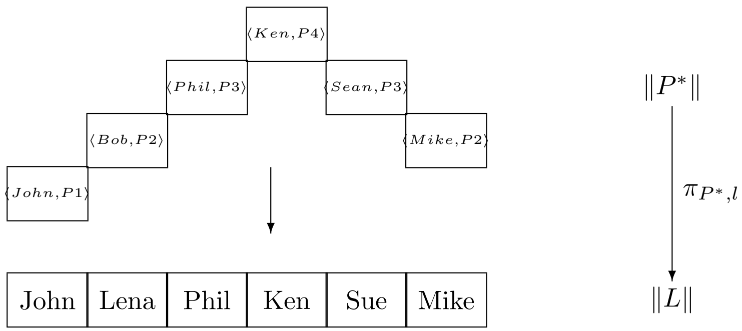
Correspondingly to 'referential' types, we can also have dependent referential types in our semantics.

Consider now an example in (2b)

(2b) Every linguist has studied every solution that a certain problem that intrigued him/her might have.

On our analysis, sentence (2b) quantifies over the dependent referential type (with one problem per fiber)

$$l : L; p : P^*(l)$$



yielding the desired dependent referential reading (Matthewson's reading). As already mentioned above (for the details, see Apprendix I), in our semantic system a quantifier expression (indefinite) can never escape a modified common noun such as *solution that a certain problem that intrigued him/her might have* — our semantics with dependent types can only yield a narrow scope reading

[2] A similar idea with thorough linguistic motivation can be found in [28].

for the dependent referential indefinite *a certain problem that intrigued him/her* in (2b). In that case, however, since there is just one problem per fiber, the narrow scope reading becomes indistinguishable from the long-distance one (in Matthewson's sense) — thus we can obtain the dependent referential reading (Matthewson's reading) with the dependent referential indefinite being interpreted in situ. Notice also that we can utter a sentence like (2b) without having any clue as to which pairing is involved.

5 Conclusion and Related Work

In this paper, we propose a new solution to Chierchia's puzzle of long-distance indefinites from the perspective of our semantics combining generalized quantifiers with dependent types [9,11]. To the best of our knowledge, our proposal is the first solution to Chierchia's puzzle that keeps the quantificational analysis of indefinites. The two recent quantificational accounts of indefinites and their exceptional scopal properties have been proposed in Schwarzschild [28] and Brasoveanu & Farkas [2]. Schwarzschild's approach assumes a unitary analysis of indefinites as existential quantifiers. On occasion the domain of an indefinite can be contextually narrowed down to one individual via pragmatic mechanism — this gives referential readings. Furthermore, the proposal is enriched with the assumption that the domain restriction of a 'singleton indefinite' can involve bound variables — this accounts for Matthewson's readings but Chierchia's intermediate readings are left unexplained. Brasoveanu & Farkas also assume a uniform analysis of indefinites as existential quantifiers but they develop a proposal that follows Independence-Friendly Logic. On this approach, an existential quantifier explicitly specifies which variables it will be dependent or independent of — this accounts for Chierchia's intermediate readings but it is not obvious how the proposal can be extended to include Matthewson's readings (for a similar worry, see [23]). At the core of our solution is quantification over fibers. By having quantification over fibers available, we offer a new pragmatic take on Chierchia's intermediate readings that credits the problematic readings to the presence of contextually induced dependencies. By introducing the device of dependent referential types, we provide a novel analysis of Matthewson's readings that does not require quantification over functions.

Acknowledgments. For valuable comments, the authors would like to thank the anonymous reviewers and the audience of three conferences: ESSLLI 2015 TYTLES Workshop on Type Theory and Lexical Semantics, Semantics and Philosophy in Europe (SPE8), and Tbilisi Symposium on Language, Logic and Computation (TbiLLC 2015). An early version of the proposal developed in this paper appeared electronically in *ESSLLI proceedings of the TYTLES workshop on Type Theory and Lexical Semantics* (Cooper, Retoré, eds) [10].

Appendix I

In the appendix, we only discuss the elements of our semantic system relevant for the linguistic purposes of this paper. For the full system, see [9,11].

Combining Quantifier Phrases - Chains of Quantifiers

To handle multi-quantifier sentences, the interpretation of quantifier phrases is extended in our system into the interpretation of (generalized) quantifier prefixes. (Generalized) quantifier prefixes can be built from quantifier phrases using the sequential composition ?|? constructor. The corresponding semantical operation is known as iteration. To illustrate with an example: *Every linguist has studied some problem* can be understood to mean that each of the linguists has studied a potentially different problem. To capture this reading:

- a sequential composition constructor ?|? is used to produce a multi-quantifier prefix: $\forall_{l:L}|\exists_{p:P}$;
- the corresponding semantical operation of iteration is defined as follows

$$\|\forall_{l:L}|\exists_{p:P}\| = \{R \subseteq \|L\| \times \|P\| : \{a \in \|L\| : \{b \in \|P\| : \langle a,b\rangle \in R\} \in \|\exists_{p:P}\|\} \in \|\forall_{l:L}\|\}.$$

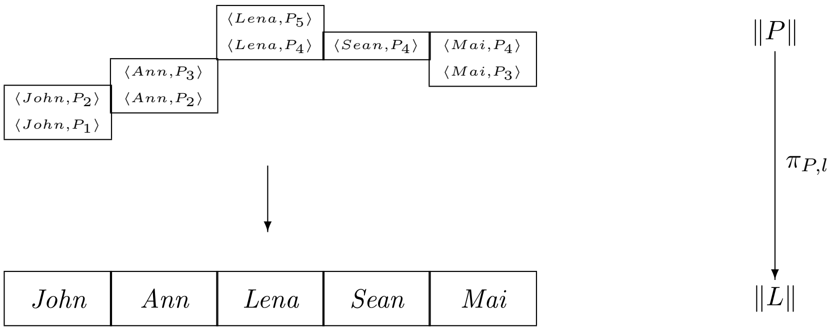
The multi-quantifier prefix $\forall_{l:L}|\exists_{p:P}$ denotes a set of relations such that the set of linguists such that each linguist is in this relation to **at least one problem** is the set of all linguists. Obviously, the iteration rule gives the same result as the standard nesting of quantifiers in first-order logic.

Combining Generalized Quantifiers with Dependent Types

The interpretation of generalized quantifier prefixes is further extended to dependent types

$$\|\forall_{l:L}|\exists_{p:P(l)}\| = \{R \subseteq \|P\| : \{a \in \|L\| : \{b \in \|P\|(a) : \langle a,b\rangle \in R\} \in \|\exists_{p:P(l)}\|(\|P\|(a))\} \in \|\forall_{l:L}\|\}.$$

The multi-quantifier prefix $\forall_{l:L}|\exists_{p:P(l)}$ denotes a set of relations such that the set of linguists such that each linguist is in this relation to **at least one problem in the corresponding fiber of problems** is the set of all linguists. By extending the interpretation of generalized quantifier prefixes to dependent types, our semantics introduces quantification over fibers, e.g. quantification over the fiber of the problems of John - $\|P\|(John)$



Our Analysis of Sentence (2)

Alphabet. The alphabet of the system consists of:
type variables: $X, Y, Z, \ldots$;
type constants: $Linguist, Problem, Solution, \ldots$;
type constructor: $\mathbb{T}$;
individual variables: $x, y, z, \ldots$;
predicates: $P^n, P_1^n, \ldots$;
quantifier symbols: $\exists, \forall, \ldots$;
prefix constructors: $?|?, \ldots$.

English-to-Formal Language Translation. Consider now an example in (2)

(2) Every linguist has studied every solution that some problem might have.

Our English-to-formal language translation process consists of two steps (i) *representation* and (ii) *disambiguation*. The syntax of the representation language - for the English fragment considered in this paper - is as follows

$S \rightarrow Prd^n(QP_1, \ldots, QP_n)$;
$MCN \rightarrow Prd^n(QP_1, \ldots,\ CN\ , \ldots, QP_n)$;
$MCN \rightarrow CN$;
$QP \rightarrow Det\ MCN$;
$Det \rightarrow every, some, \ldots$;
$CN \rightarrow linguist, problem, \ldots$;
$Prd^n \rightarrow study, have, \ldots$

Sentence (2) is accordingly represented as
Study^2 (*every linguist*, *every solution that some problem might have*).

Multi-quantifier sentences of English, contrary to sentences of our formal language, are often ambiguous. Hence one sentence representation can be associated with more than one sentence in our formal language. The second step thus involves disambiguation. We take quantifier phrases out of a given representation and organize them into possible prefixes of quantifiers. In the case of our example, the sentence translates as

$$\forall_{l:Linguist} | \forall_{t_s:\mathbb{T}_{Solution\ to\ some\ problem}} Study^2(l, t_s).$$

Interpretation. In the Montague-style semantics, common nouns are interpreted as predicates (expressions of type $e \rightarrow t$). In our type-theoretic setting, common nouns (CN), e.g. *linguist*, are treated as types. To handle modified common nouns (MCN), e.g. *solution that some problem might have*, we introduce into our system $*$-*sentences* (= *Have*^{2} (*some problem*, *solution*) determining some possibly dependent types (interpreted via the algorithm introduced in [9,11]).

In our example, the type *Linguist* is interpreted as a set of linguists (indicated in the context) $\|Linguist\|$ and the type $\mathbb{T}_{Solution\,to\,some\,problem}$, as

$\|\mathbb{T}_{Solution\,to\,some\,problem}\| = \{c \in \|Solution\| : \{b \in \|Problem\| : \langle b, c\rangle \in \|Have\|\} \in \|\exists_{p:Problem}\|\}$.

Thus, as can be seen from this analysis, our semantics with dependent types can only yield a narrow scope reading for the indefinite *some problem* in (2): every a in $\|Linguist\|$ has studied every c in $\|\mathbb{T}_{Solution\,to\,some\,problem}\|$.

Appendix II: Interpretation of Function Terms

To complete the semantic account given in [9,11], we need to show how function terms can be handled in our system. Consider two sentences involving, respectively, a reflexive pronoun (RP) and a function term (FT):

(RP) Every man loves himself.

$\forall_{m:Man} Love(m, m)$

(FT) Every man loves his partner.

$\forall_{m:Man} Love(m, f(m))$

Our method for interpreting reflexive pronouns uses a diagonal function:

$$\delta_X : X \to X \times X \;\; \textit{such that} \;\; \delta_X(x) = \langle x, x\rangle, \;\; \textit{for} \;\; x \in X.$$

Then for any subset $R \subseteq X \times X$ we have $\delta_X^{-1}(R) \subseteq X$, the inverse image of R under δ_X. Thus if *Love* is a binary predicate interpreted as a subset $\|Love\|$ of $X \times X$, the interpretation of the formula $Love(x, x)$ is a subset $\delta_X^{-1}(\|Love\|)$ of X, the inverse image of $\|Love\|$ under δ_X. It contains those $x \in X$ that $\langle x, x\rangle \in \|Love\|$. What (RP) says then is that the set of men such that they love themselves is the set of all men.

Reflexive pronouns, as in (RP), are just a special case (involving identity) of a broader class of function terms, as in (FT). In case of (FT), the function is no longer diagonal, but otherwise the interpretational procedure is the same. Now the function G_f is the embedding of X onto the graph of the function $\|f\|$ which is the interpretation of the function symbol f, i.e.

$$G_f : X \to X \times Y \;\; \textit{such that} \;\; G_f(x) = \langle x, \|f\|(x)\rangle, \;\; \textit{for} \;\; x \in X.$$

Then for any subset $R \subseteq X \times Y$ we have $G_f^{-1}(R) \subseteq X$, the inverse image of R under G_f. Thus if *Love* is a binary predicate interpreted as a subset $\|Love\|$ of $X \times Y$, the interpretation of the formula $Love(x, f(x))$ is a subset $G_f^{-1}(\|Love\|)$ of X, the inverse image of $\|Love\|$ under G_f. It contains those $x \in X$ that $\langle x, \|f\|(x)\rangle \in \|Love\|$. What (FT) says then is that the set of men such that they love their partners is the set of all men.

References

1. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguist. Philos. **4**, 159–219 (1981)
2. Brasoveanu, A., Farkas, D.F.: How indefinites choose their scope. Linguist. Philos. **34**, 1–55 (2011)
3. Chierchia, G.: A Puzzle about Indefinites. In: Cecchetto, C., Chierchia, G., Guasti, M.T. (eds.) Semantic Interfaces: Reference, Anaphora, and Aspect, pp. 51–89. CSLI, Stanford (2001)
4. Cooper, R.: Dynamic generalised quantifiers and hypothetical contexts. In: Ursus Philosophicus, a festschrift for Björn Haglund, Department of Philosophy, Göteborg University (2004)
5. Dekker, P.: The pragmatic dimension of indefinites. Res. Lang. Comput. **2**, 365–399 (2004)
6. Fernando, T.: Conservative generalized quantifiers and presupposition. In: Proceedings SALT XI, Cornell University, pp. 172–191 (2001)
7. Geurts, B.: Indefinites, presupposition, and scope. In: Buerle, R., Zimmermann, T.E. (eds.) Presuppositions and Discourse, pp. 125–158. Elsevier, Oxford (2010)
8. Groenendijk, J., Stokhof, M.: Dynamic predicate logic. Linguist. Philos. **14**, 39–100 (1991)
9. Grudzińska, J., Zawadowski, M.: System with generalized quantifiers on dependent types for anaphora. In: Cooper, R., Dobnik, S., Lappin, S., Larsson, S. (eds.) Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics, pp. 10–18 (2014)
10. Grudzińska, J., Zawadowski, M.: A puzzle about long-distance indefinites and dependent type semantics. In: Cooper, R., Retoré, C. (eds.) ESSLLI Proceedings of the TYTLES Workshop on Type Theory and Lexical Semantics. ESSLLI 2015, Barcelona (2015)
11. Grudzińska, J., Zawadowski, M.: Generalized quantifiers on dependent types: a system for anaphora. In: Chatzikyriakidis, S., Luo, Z. (eds.) Modern Perspectives in Type-Theoretical Semantics. Studies in Linguistics and Philosophy, pp. 95–131. Springer, Heidelberg (2017)
12. Jäger, G.: Partial variables and specificity. In: Sauerland, U., Penka, D. (eds.) Presupposition and implicature in compositional semantics, pp. 121–162. Palgrava-McMillan, Basingstoke (2007)
13. Kamp, H.: A theory of truth and semantic representation. In: Groenendijk, J., Janssen, T., Stokhof, M. (eds.) Truth, Interpretation and Information, pp. 1–41. Foris Dordrecht (1981)
14. Kamp, H., Reyle, U.: From Discourse to Logic. Kluwer Academic Publishers, Dordrecht (1993)
15. Kratzer, A.: Scope or Pseudoscope? Are there Wide-Scope Indefinites? In: Rothstein, S. (ed.) Events and Grammar, pp. 163–196. Kluwer Academic Publishers, Dordrecht (1998)
16. Lappin, S., Francez, N.: E-type pronouns, I-sums, and Donkey anaphora. Linguist. Philos. **17**, 391–428 (1994)
17. Lindström, P.: First-order predicate logic with generalized quantifiers. Theoria **32**, 186–95 (1966)
18. Luo, Z.: Formal semantics in modern type theories with coercive subtyping. Linguist. Philos. **35**, 491–513 (2012)

19. Makkai, M.: First Order Logic with Dependent Sorts, with Applications to Category Theory. Preprtint McGill University (1995)
20. Martin-Löf, P.: An intuitionstic theory of types. Technical Report, University of Stockholm (1972)
21. Matthewson, L.: On the interpretation of wide-scope indefinites. Nat. Lang. Semant. **7**, 79–134 (1999)
22. Mostowski, A.: On a generalization of quantifiers. Fundam. Math. **44**, 12–36 (1957)
23. Onea, E.: Why indefinites can escape scope Islands. Linguist. Philos. **38**, 237–267 (2015)
24. Ranta, A.: Type-Theoretical Grammar. Oxford University Press, Oxford (1994)
25. Reinhart, T.: Quantifier scope: how labour is divided between QR and choice functions. Linguist. Philos. **20**, 335–397 (1997)
26. Schlenker, P.: A note on Skolem functions and the scope of indefinites. Poster, NELS (1998)
27. Schwarz, B.: Two kinds of long distance indefinites. In: van Rooy, R., Stokhof, M. (eds.) Proceedings of the Thirteenth Amsterdam Colloquium, Amsterdam University, pp. 192–197 (2001)
28. Schwarzschild, R.: Singleton indefinites. J. Semant. **19**, 289–314 (2002)
29. Steedman, M.: Taking Scope. The Natural Semantics of Quantifiers. The MIT Press, Cambridge (2012)
30. Szabolcsi, A.: Quantification. Cambridge University Press, Cambridge (2010)
31. Van den Berg, M.H.: The Internal Structure of Discourse. Ph.D. thesis, Universiteit van Amsterdam, Amsterdam (1996)
32. Winter, Y.: Choice functions and the scopal semantics of indefinites. Linguist. Philos. **20**, 399–467 (1997)

First-Order Abduction as Enumeration of Stable Models

Petr Homola(✉)

Codesign Ltd., Prague, Czech Republic
phomola@codesign.cz

Abstract. We present a novel method for translating (a fragment of) first-order abduction into an answer-set program in the context of natural language understanding since the heretofore used algorithms do not allow for seamless integration of the process of abduction with deduction. Our method helps the solver confine the search space by ruling out logically impossible proofs (with respect to a background theory).

Keywords: Abduction · Natural language processing

1 Introduction

Natural language understanding (NLU), a field of natural language processing (NLP) which interprets sentences at the level of pragmatics, can be thought of as first-order abduction [10,13]. The sentence we want to interpret is what is "observed" and the "best" abductive proof tells us what the sentence actually means. Abductive reasoning is nonmonotonic and ampliative, that is, what we conclude cannot be proved deductively, but it extends our background knowledge in a coherent way.

Abduction can be combined with other forms of reasoning, such as deduction. Indeed, most ontologies consist of both deductive and abductive rules. Particularly useful is weighted abduction, a variant of probabilistic abduction that can be used to find optimal proofs. What "optimal" means depends on what is being reasoned about. In discourse analysis, for instance, the least specific proof maximizing redundancy is preferred [10], that is, we assume as little as possible while trying to unify as many assumptions as possible with what is already known and preferring more salient individuals. In the domain of automated planning, the shortest proof (i.e., the shortest plan leading to the desired goal state) is considered the best, and so forth.

There is no efficient (polynomial) algorithm for first-order abduction (cf. [1], "even in the case of propositional Horn clause theories, the problem of computing an explanation for an arbitrary proposition is NP hard"). In early experiments with interpretation as (weighted, i.e., cost-based) abduction, the only published algorithm for rule-based (i.e., nonstochastical) abduction proposed by Hobbs et al. [10], which is based on a Prolog-based abductive theorem prover, was used but it is too slow.[1] Later a more efficient algorithm based on inte-

[1] Most sentences of average length take up to 20 min to process.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 54–65, 2017.
DOI: 10.1007/978-3-662-54332-0_4

ger linear programming was developed [12,13].[2] But neither approach allows for seamless integration of abduction with full-fledged deduction. Although propositional abduction is known to be implementable as stable model enumeration [5], in the case of NLU the underlying theory is first-order. In this paper we present a method based on ideas first presented in [11] for comparatively efficient first-order abduction implemented as answer-set solving. In our experiments, we used the solver described in [3].

Our approach is somewhat similar to that of [14], though we additionally consider salience effects. Since the work described in [14] is very recent, a comparison of the two methods has yet to be done.

In Sect. 2 we give an overview of the problem illustrated with a simple example. Section 3 describes the translation of a first-order abductive problem into an answer-set problem. Section 4 briefly describes possible integrity constraints that significantly constrain the proof search space. Section 5 presents a method for selecting the best abductive proof in the domain of NLU. Finally, Sect. 6 concludes.

2 Preliminaries

Abduction is the reasoning process of finding explanations for observations. In NLU, the "observations" are (logical forms of) sentences and the interpretation of a sentence is the "best" (i.e., most coherent) abductive proof that explains it with respect to a background theory. Formally, I is an interpretation of observations O with respect to a background theory T if

$$T \cup I \vDash O \text{ and } T \cup I \nvDash \bot \tag{1}$$

that is, $T \cup I$ entails O and is consistent. The sentence *John is an elephant* may mean that there is an actual elephant whose name is John, i.e., I can be the literal meaning of the sentence. But if we know (from context) that John is a person, $T \cup I$ will be inconsistent, hence I cannot be the literal meaning of the sentence and we are forced to find some other (nonliteral) interpretation that makes sense in the given context.

We use the logical representation proposed by Hobbs [6], which is a "conjunctivist" scope-free first-order approach to linguistic meaning. Consider the sentence *John sees Mary.* Irrelevant details aside, its canonical logical representation is

$$(\exists e, x, y) see'(e, x, y) \wedge John(x) \wedge Mary(y) \tag{2}$$

that is, there is an eventuality e which is a seeing, John does it and Mary undergoes it. Similarly, the logical representation of *John doesn't see Mary* would be

$$(\exists e_1, e_2, x, y) not'(e_1, e_2) \wedge see'(e_2, x, y) \wedge John(x) \wedge Mary(y) \tag{3}$$

[2] There is also a stochastical algorithm [2] but we are concerned only with symbolic computation in this paper.

that is, the negation of the eventuality expressed in (2) is asserted. Hobbs [9] shows that an appropriately rich and precise theory of commonsense reasoning can be expressed in this way.

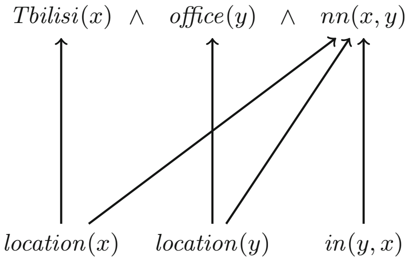


Fig. 1. Proof of *the Tbilisi office*

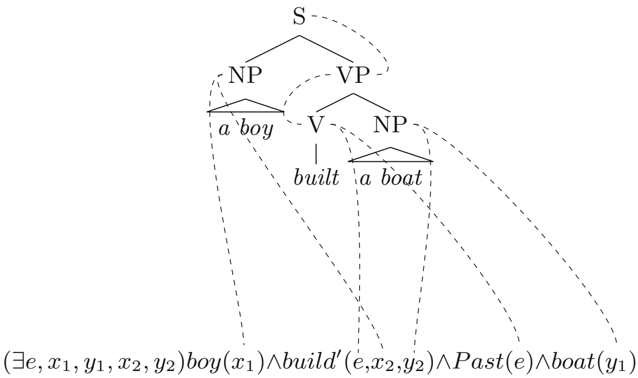


Fig. 2. Parse tree and variable bindings for *a boy built a boat*

The intended real-world meaning of (2) and (3) is their literal meaning. But the logical representation can contain linguistic predications that need be interpreted with respect to a background theory in order to be given a real-world meaning. The noun phrase (NP) *the Tbilisi office* is parsed as

$$(\exists x, y)\, \mathit{Tbilisi}(x) \wedge \mathit{office}(y) \wedge \mathit{nn}(x, y) \tag{4}$$

that is, there are two entities and an *nn* (i.e., N-N compound) relation between them, but the predicate *nn* only tells us that x and y are adjacent NPs and x precedes y. If we know from context that there is an office in Tbilisi, we want to interpret (4) as *in*(*office*, *Tbilisi*). If we have no such information but our background knowledge contains the facts that Tbilisi is a city and that cities and offices are locations, we can still draw the (defeasible) conclusion that there is an *in* relation between the entities. Whichever the case, though, we need the following lexical axiom (we use $\rightharpoonup$ to denote defeasible implications in abductive rules and $\supset$ to denote implications in "hard" rules):

$$(\forall x, y) in(x, y) \rightharpoonup nn(y, x) \tag{5}$$

that is, the fact that x is located in y can be expressed by a compound nominal at the lexical level. Of course, axiom (5) is only defeasible, for an *in* relation can also be expressed by other linguistic constructions. Thus, to interpret (4) we have to backchain on axiom (5) and unify both variables. The proof of (4) is depicted in Fig. 1. As can be seen, the abduction process is first-order even for simple examples (Fig. 2).

Having exemplified how abduction works in Hobbs' [10] framework, let us now turn to stable models used in ASP. We take the definition of a stable model from [4]. Let us assume that we have a set of rules of the form

$$A \leftarrow B_1, \ldots, B_m, \sim C_1, \ldots, \sim C_n$$

A model of such a set of rules is the set of atoms that are true. For any set of atoms I, the reduct of a set of rules R relative to I is the set of rules without negation obtained from R by dropping the rules that contain '∼C' for an atom C from I and then dropping all the '$\sim C_i$' from the remaining rules. I is a stable model of R if I is a model of the reduct of R relative to I.

3 Translation of Deductive and Abductive Rules

Classical deduction poses no problem to the solver. We can have strict rules such as[3]

$$\begin{array}{l} elephant_1(x) \supset mammal_1(x) \\ mammal_1(x) \supset animal_1(x) \end{array} \tag{6}$$

and we can use strong negation and disjunction, as in

$$\begin{array}{l} person_1(x) \supset \neg animal_1(x) \\ person_1(x) \supset man_1(x) \vee woman_1(x) \end{array} \tag{7}$$

The main result reported in this paper is a method for converting abduction with observations that contain variables into an answer-set problem. We represent

[3] We use subscripted predicates to represent real-world meaning and unsubscripted predicates to represent lexical meaning. This distinction is necessary in order to accommodate lexical ambiguity and figurative speech such as metaphors and metonymy. For example, the literal real-world meaning of *John is an elephant* (whose logical form is $John(x) \wedge elephant(x)$) is $John_1(x) \wedge elephant_1(x)$, but if we already know that John is a person ($person_1(John)$) we are forced into a figurative meaning such as $John_1(x) \wedge clumsy_1(x)$. Thus in this sense, to interpret a sentence is to steer clear of contradictions in the knowledge base.

observations and assumptions as follows ($\Rightarrow$ denotes translations from FOL into ASP):

$$\begin{array}{ll} \text{observations: } & elephant(x) \;\; \Rightarrow pred(elephant, obsrv, var_x) \\ \text{assumptions: } & elephant_1(x) \Rightarrow pred(elephant1, asmpt, var_x) \end{array} \tag{8}$$

that is, variables are encoded as individuals. An abductive rule is encoded as follows ($p \supset_0^1 q$ means `0{q}1 :- p`, that is, the consequent may or may not be included in the stable model; this rule encodes the fact that assumptions are optional):[4]

$$\begin{array}{c} elephant_1(x) \rightharpoonup elephant(x) \\ \Rightarrow \\ pred(elephant, x, y) \wedge \\ x \in \{obsrv, asmpt\} \supset_0^1 pred(elephant_1, asmpt, y) \wedge \\ explainedBy(elephant, y, rule_1(y)) \wedge assumedBy(elephant_1, y, rule_1(y)) \end{array} \tag{9}$$

that is, if there is a predication we want to explain that can be unified with the consequent of an abductive rule, we may assume the antecedent (but we may ignore the rule because the cost of assuming the antecedent may be higher than the cost of assuming the consequent, thus we have to always consider both cases). The auxiliary predications are used in the following rules, which help us guarantee that the computed stable model is a correct abductive proof (in the sense of [10]; $\sim$denotes default negation):

$$\begin{array}{c} assumedBy(p, x, r) \supset assumed(p, x) \\ pred(p, asmpt, x) \wedge \sim assumed(p, x) \supset \bot \\ explainedBy(p, x, r_1) \wedge explainedBy(p, x, r_2) \wedge r_1 \neq r_2 \supset \bot \end{array} \tag{10}$$

that is, a predication can be assumed only if it can be derived by backchaining on an abductive rule from predications that are already assumed by some other rule(s) and a predication cannot be explained by more than one rule.

The most important aspect is how variables in observations are handled. Since an answer-set program (ASP) has to be effectively propositional, we have to "emulate" equality. If a predication containing a "reified" variable (e.g., var_x) can be unified with another predication, we can bind the variable. For example, if the knowledge base contains $person_1(John)$ and we observe or assume $person_1(var_x)$, we may add $eq(John, var_x)$ to the stable model. We may also decide not to bind a variable, in which case a new individual has to be added to the knowledge base. Of course, we need axioms that guarantee that eq is an equivalence relation. We will not list all of them here but let us mention the most important axiom schema. If P is a predicate, we need $P(x) \wedge eq(x, y) \supset P(y)$ in order for deduction to work. Thus whenever we bind a variable, the knowledge

[4] An abductive rule as defined in [10] is an implication whose antecedent and consequent are conjunctions of positive literals, whereby all the literals in the antecedent are abducible (assumable).

base grows (in the worst case exponentially). There is no way around this problem since answer-set solving is propositional. Luckily for us, the logical form of a sentence contains only few variables (around a dozen), hence the presented method is viable for NLU. In actual fact, we are sacrificing space for time, since we need new literals for each variable assignment.

We use defeasible rules to infer what might be true (thus extending our knowledge) based on what we already now. We process one sentence at a time in order to keep the proof search space as small as possible. Of course, in a connected discourse this may lead to a contradiction. Generally we try to assume as little as possible (see Sect. 5), but if we arrive at a contradiction, we have to backtrack and reprocess part of the discourse. We plan on using a truth maintenance system in the future, but for the time being, we simply reinterpret the sentences which might be affected by the false assumption.

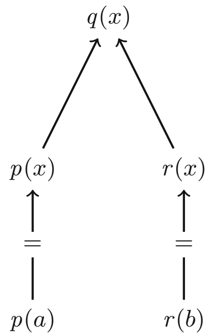


Fig. 3. Complete proof graph for (11)

We will now illustrate with a simple example how the solver finds abductive explanations by generating a proof graph and yielding subgraphs corresponding to well-formed proofs. Consider the following observation, abductive rules, and known facts (as usual, x is a variable and a, b are constants):

$$\begin{array}{ll} \text{observation:} & q(x) \\ \text{rule:} & p(x) \rightharpoonup q(x) \\ \text{rule:} & r(x) \rightharpoonup q(x) \\ \text{facts:} & p(a), r(b) \end{array} \tag{11}$$

The complete proof graph for (11) is given in Fig. 3. The labelled edges are possible "merges" (syntactically unified predications) and the unlabelled edges are licensed by backchaining on abductive rules. An abductive proof is equivalent to a subgraph of the complete proof graph complying with the following conditions:

1. An observation or assumption is explained by no more than one defeasible rule.[5]

[5] This condition does not mean that a literal cannot be implied by more rules, it only says that only one (defeasible) rule is taken to be its explanation.

2. There is a path from any assumption to an observation (that is, we eliminate assumptions that do not, directly or indirectly, contribute to the explanation of an observation).
3. Variable assignments conform to the usual constraints on equivalence, i.e., reflexivity, symmetry, and transitivity.

These constraints on the subgraph guarantee that the corresponding proof be well-formed.

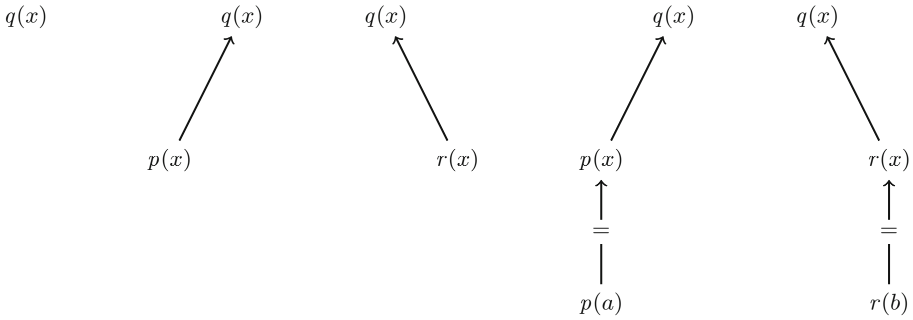


Fig. 4. Proofs of (11)

All the proofs of (11) are depicted in Fig. 4. The corresponding hypotheses (including the observation) are:[6]

$$\begin{array}{c} q(x) \\ p(x) \wedge q(x) \\ r(x) \wedge q(x) \\ q(a) \\ q(b) \end{array} \tag{12}$$

It is not hard to see that each stable model of the translated theory represents an abductive proof in the original theory. The proof can be sketched by induction on the size of abductive proofs as follows: If we add no new assumptions (no *assumed* literals), the stable model corresponds to the trivial proof where all the observations are assumed. Let us now have a stable model that corresponds to an abductive proof, and a superset of it that contains one more assumption. Since it could only be derived through a rule like that in (9), we know that the bigger model also represents a correct abductive proof, because the additional assumption can be obtained by backchaining on an abductive rule. Since abductive proofs can be constructed iteratively, each stable model encodes a well-formed proof. On the other hand, each abductive proof—provided the

[6] In the domain of NLU, the logical form of a sentence is an existential closure so we would have to introduce a new individual for every free variable.

added assumptions are consistent—corresponds to a stable model, because the form of the rules like that in (9), namely

$$0\{\ldots\}1 \text{ :- } \ldots$$

ensures that all the possible assumptions obtainable by backchaining on a defeasible rule in the theory are generated. There is therefore a one-to-one correspondence between abductive proofs and the stable models of the translated ASP program.

4 Constraining the Proof Search Space

In modern answer-set solvers, aggregate functions such as *count*, *sum*, or *max* can be used. We can thus define a predicate whose argument tells us the size of (that is, the number of assumptions in) a proof and rule out any proof that is too big:

$$\begin{gathered} numberOfAssumptions(x) = |\{p : assumed(p)\}| \\ numberOfAssumptions(x) \wedge \\ \wedge x > maxNumberOfAssumptions \supset \bot \end{gathered} \tag{13}$$

We can also have a predicate that tells us the length of the proof path from p_1 to p_2 and an integrity constraint that rules out any proof whose length is greater than an integer constant, *maxProofLength*:

$$proofLength(p_1, p_2, l) \wedge l > maxProofLength \supset \bot \tag{14}$$

These two simple integrity constraints can significantly constrain the proof search space, thus speeding up the enumeration of proofs.

5 Ranking Abductive Proofs

Even a relatively simple background theory, as in our experiments, will have at least hundreds of abductive and deductive rules, which means that the logical form of an average sentence can have many different proofs (since both deduction and abduction are explosive). Moreover in a long discourse consisting of many sentences, the knowledge base will contain many individuals available for unification, which can multiply the number of proofs. The method of weighted-abduction [7,8,10] seems to yield good results, but it cannot be used in an answer-set program because it works with real numbers. While it is possible to enumerate all the abductive proofs and evaluate them later, we decided to try to solve the ranking problem within ASP. In this section, we describe how to rank proofs using the answer-set solver.

The basic idea, coming from [10], is that one should prefer proofs that unify assumed predications with what is already known and assume as little as possible. In other words, maximally coherent (with respect to the context) and minimally ampliative proofs are preferred. We add one more criterion: salience. Informally, individuals that occurred recently in the discourse have higher salience.

If the pronoun *he* is used in a sentence, it is interpreted by backchaining on the following abductive rule:

$$person_1(x) \wedge male_1(x) \rightharpoonup he(x) \tag{15}$$

that is, *he* refers to a male person. To bind the variable, we have to find an individual that conforms to the selectional constraints. But there can be many such individuals in the knowledge base. Thus we rank the proofs with respect to their salience, which is the sum of the saliences of all individuals unified with a variable. Newly introduced constants are assigned the highest salience, as in the case of c for the indefinite NP in

$$\begin{array}{l}\text{He bought a book.}\\ he(x) \wedge buy'(e, x, c) \wedge book(c)\end{array} \tag{16}$$

For example, if there is $elephant(x)$ among the observations and the knowledge base contains $elephant_1(E_1^1), elephant_1(E_2^2), elephant_1(E_3^3)$,[7] there are three literal interpretations in which the variable x is unified, as illustrated in Fig. 5. Based on a linguistic insight, we want to prefer the proof which unifies the variable x with the most salient individual. In this simple example there is only one unified individual, so the salience of the proof equates to this individual's salience.

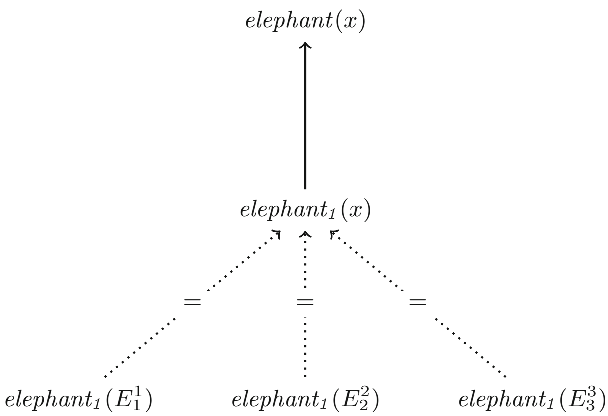


Fig. 5. Literal interpretations of *elephant(x)*

Unification can help us resolve lexical ambiguity even when there are no individuals in the knowledge base that could be unified with the variables. Consider the compound nominal

$$\begin{array}{l}\text{bank account}\\ bank(x) \wedge account(y) \wedge nn(x, y)\end{array} \tag{17}$$

[7] The upper index expresses the salience of the individual.

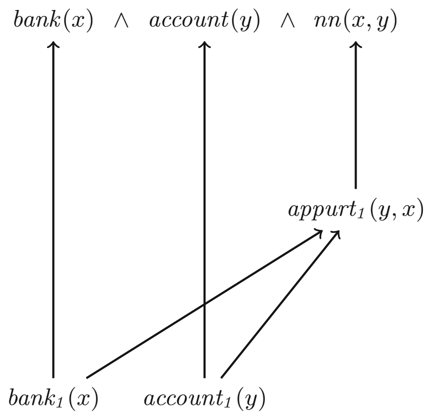


Fig. 6. Interpretation of *bank account*

and assume that we have the following background theory (with five lexical rules and one commonsense rule):

$$\begin{gathered} bank_1(x) \rightharpoonup bank(x) \\ bank_2(x) \rightharpoonup bank(x) \\ account_1(x) \rightharpoonup account(x) \\ account_2(x) \rightharpoonup account(x) \\ appurt_1(x,y) \rightharpoonup nn(y,x) \\ account_1(x) \wedge bank_1(y) \rightharpoonup appurt_1(x,y) \end{gathered} \tag{18}$$

that is, accounts (defeasibly) appertain to banks and the relation *appurt(enance)* can be expressed by a compound nominal (*nn*).[8] We see in Fig. 6 that two predications are unified. If we interpreted $bank(x)$ as $bank_2(x)$ and/or $account(y)$ as $account_2(y)$, there would be no unification of predications and hence the proof would be less coherent.

As we have just shown, abduction can be used to lexically disambiguate phrases even if there is no additional context or previous discourse. Of course, in order for this method to work background theories are needed that capture relations between entities that occur in the sentence. Creating commonsense knowledge bases is generally a very complex task but it is feasible at least for smaller closed domains.

There can be many proofs with the same number of unified predications, thus we need a criterion that will help us distinguish them. The simplest criterion is the size of the proof, i.e., the number of assumptions made. Intuitively, we do not want to assume more than is necessary; if an assumption does not help us arrive at a more coherent proof (that is, a proof with more unifications), is should be omitted. This simple idea seems to be good enough to rule out most undesired proofs.

[8] $bank_x$ and $account_x$ are different lexical meanings of *bank* and *account*, respectively.

We can use the predicate *numberOfAssumptions* defined in Sect. 4 and an analogously defined predicate *numberOfUnifications* to select the best proof. Our method for ranking abductive proofs yields slightly better results than that proposed by Hobbs el at. [10] in our evaluation,[9] but this does not mean that it is better because the difference is not statistically significant and because Hobbs' method relies on probabilistic weights which are hard to "get right" empirically. Nevertheless our method is relatively simple and can be implemented in ASP (for it uses only natural numbers).

6 Conclusions

We have presented a method for translating (a fragment of) first-order abduction into an answer-set program in the context of NLU. The heretofore used algorithms do not allow for seamless integration of the process of abduction with deduction. Our method helps the solver confine the search space by ruling out logically impossible proofs (with respect to a background theory). We have also suggested how to rank proofs within the answer-set program, since the original framework of weighted abduction would require an additional step to evaluate the proofs. An evaluation has shown that in conjunction with a state-of-the-art answer-set solver, our method is an order of magnitude faster than the approach based on a general automated theorem prover. On the other hand, it currently does not provide significantly better results in terms of precision. The fact that we could not use weighted abduction as proposed in its original form, for one cannot use real numbers in ASP, means that our approach of ranking abductive proofs is not directly comparable to weighted abduction and it remains yet to be explored whether it can yield comparable results when applied to very large theories.

There is no doubt that first-order Horn abduction is useful in many areas of artificial intelligence. Our future work will investigate how the proposed method for ranking proofs can be applied to automated goal-driven planning with incomplete information.

References

1. Appelt, D.E., Pollack, M.E.: Weighted abduction for plan ascription. User Model. User-Adap. Interact. **2**(1–2), 1–25 (1992)
2. Blythe, J., Hobbs, J.R., Domingos, P., Kate, R.J., Mooney, R.J.: Implementing weighted abduction in markov logic. In: Proceedings of Ninth International Conference on Computational Semantics, pp. 55–64 (2011)
3. Gebser, M., Kaufmann, B., Schaub, T.: Conflict-driven answer set solving: from theory to practice. Artif. Intell. **187–188**, 52–89 (2012)

[9] We evaluated our approach against a corpus of 1,000 sentences from news articles using a knowledge base obtained semi-automatically from WordNet. There were approx. 60,000 abductive rules. The vast majority of the rules have 2–3 literals in the antecedent and 1–2 literals in the consequent.

4. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programs. In: Kowalski, R., Bowen, K. (eds.) Proceedings of the 5th International Symposium on Logic Programming, pp. 1070–1080 (1988)
5. Gelfond, M., Kahl, Y.: Knowledge Representation, Reasoning, and the Design of Intelligent Agents. The Answer-Set Programming Approach. Cambridge University Press, New York (2014)
6. Hobbs, J.R.: Ontological promiscuity. In: Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics, pp. 61–69 (1985)
7. Hobbs, J.R.: Syntax and metonymy. In: Bouillon, P., Busa, F. (eds.) The Language of Word Meaning, pp. 290–311. Cambridge University Press, New York (2001)
8. Hobbs, J.R.: Discourse and Inference, ms, Information Sciences Institute. University of Southern California, Marina del Rey (2003)
9. Hobbs, J.R.: Encoding Commonsense Knowledge, ms, Information Sciences Institute. University of Southern California, Marina del Rey (2005)
10. Hobbs, J.R., Stickel, M., Appelt, D., Martin, P.: Interpretation as abduction. Artif. Intell. **63**, 69–142 (1993)
11. Homola, P.: Abduction as satisfiability. In: Proceedings of the Workshop on Defeasible and Ampliative Reasoning (DARe 14) (2014)
12. Inoue, N., Ovchinnikova, E., Inui, K., Hobbs, J.: Coreference resolution with ILP-based weighted abduction. In: Proceedings of COLING 2012, pp. 1291–1308 (2012)
13. Ovchinnikova, E., Israel, R., Wertheim, S., Zaytsev, V., Montazeri, N., Hobbs, J.: Abductive inference for interpretation of metaphors. In: Proceedings of the Second Workshop on Metaphor in NLP, pp. 33–41 (2014)
14. Schüller, P.: Modeling variations of first-order horn abduction in answer set programming. Computer Engineering Department, Faculty of Engineering, Marmara University, Turkey, Technical report (2016)

A Semantic Account of the Intervention Effects in Chinese Why-Questions

Dawei Jin(✉)

Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, China
daweijin@buffalo.edu

Abstract. This paper revisits intervention effects in Mandarin Chinese *why*-questions. I present new data showing that the ability for quantifiers to induce intervention hinges upon their monotonicity and their ability to be interpreted as topics. I then develop a semantic account that correlates topicality with monotone properties. Furthermore, I propose that *why*-questions in Chinese are idiosyncratic, in that the Chinese equivalent of *why* directly merges at a high scope position that stays above a propositional argument. Combining the semantic idiosyncrasies of *why*-questions with the theory of topicality, I conclude that a wide range of intervention phenomena can be accounted for in terms of interpretation failure.

Keywords: Intervention effects · *Why*-questions · Illocutionary acts · Wide-scope indefinites · Mandarin Chinese

1 Data

This paper presents a semantic account of the quantifier-induced intervention effects in Chinese *why*-questions, schematized as follows.

(1) $\#[_Q\ [\textit{Quant why}]]$

That is, unacceptability arises when a quantifier c-commands the interrogative phrase *why*. Using Chinese data, this paper argues that the intervention induced by *why*-questions is distinct from other intervention effects that arise in non-*why* interrogative questions, which have received detailed investigations in the literature.[1] Specifically, I present new data showing that intervention effects in

This paper benefits from discussions with Jun Chen and Lihua Xu. I also thank one anonymous reviewer for the abstract of the TbiLLC 2015 conference and conference attendees for their feedback. I am particularly indebted to the two anonymous reviewers for the TbiLLC 2015 Post-Proceedings for their detailed and insightful comments and suggestions for improvement. All the remaining errors are my own.

[1] See Beck [5] for a semantic account of intervention in non-*why* *wh*-questions, Beck and Kim [6] for a similar account of intervention in alternative questions, and Tomioka [59] for a pragmatic, information structure-based account of intervention effects in non-*why* constituent questions.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 66–88, 2017.
DOI: 10.1007/978-3-662-54332-0_5

Chinese *why*-questions are sensitive to the type of quantifier. Since the Mandarin Chinese-speaking community is huge by population size and internal linguistic/social diversity, there is an important issue as to the extent of variation in how an exhaustive list of quantifiers is accepted. The previous literature has (understandably) tended to abstract away from any such variation. While I won't be able to offer any characterization of the nature of variation here, to the degree possible I have tried to minimize variation by focusing on a specific dialect group: the Mandarin spoken in Beijing and the adjacent *Dongbei* 'Northeast' provinces. My primary consultants are three female speakers in their twenties. Two additional male speakers in their thirties are recruited for a subset of the elicited data. All of them come from the above two regions.

As (2) shows, when *weishenme* 'why' is c-commanded by a monotone decreasing quantificational DP, oddness ensues.[2]

(2) #{Meiyou ren /Henshao ren/Budao san-ge ren}
{No person /few person/Less.than three-CLF person}
weishenme cizhi?
why resign
#'{For nobody/For few people/For less than three people}, why did they resign?'

In contrast, a quantificational DP with a simplex monotone increasing determiner, such as *most people*, or *a few people*, does not induce intervention effects.[3]

(3) {Daduoshu ren /Shaoshu ren} weishenme cizhi?
{Most person /A.few person} why resign
'{For a majority group of people/for a minority group of people}, why did they resign?'

To make things more complex, one class of monotone increasing quantificational DPs with morphosyntactically complex determiners induce weak intervention. This class includes modified numerals such as *at least three people*, *more than three people*, etc. Non-monotonic bare numerals, such as *three people*, also induce weak intervention. An example is given in (4).

---

[2] The glossing in this paper follows the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). A list of the abbreviations in this paper is given as follows:
ACC: accusative; CLF: classifier; COP: copula; DEM: demonstrative; NEG: negative, negation; NOM: nominative; LOC: locative; PASS: passive; PL: plural; POSS: possessive; PRF: perfect; PRS: present; PRT: particle; PST: past; Q: question particle; REL: relativizer; RES: resultative; TOP: topic marker.

[3] Based on monotonicity, I treat the Chinese quantifier *henshao ren* as an equivalent of *few people*, since both require a less-than-half cardinality reading and are monotone decreasing. Furthermore, I treat *shaoshu ren* as an equivalent of *a few people*, as they pattern together as non-monotonic quantifiers with a less-than-half reading. It is also worth noting that *a few people*/*shaoshu ren* generally give rise to a non-empty scalar implicature (see Horn [28]), whereas *few people*/*henshao ren* generally do not.

(4) ??{San-ge ren/ zhishao san-ge ren/ chaoguo
{Three-CLF person/ at.least three-CLF person/ more.than
san-ge ren} weishenme cizhi?
three-CLF person} why resign
'{For three people/at least three people/more than three people}, why did they resign?'

My notational choice here, using ?? in (4) to contrast with the use of # in (2), will be justified in my coming argument that the unacceptability found in examples in (2) results from interpretation failure, whereas the unacceptability in (4) is a case of contextual infelicity. The choice also reflects the intuition of my consulted speakers. When uttered out of the blue, (4) triggers rather low judgments for some speakers, while for other speakers the oddness is less severe than that which is induced in monotone decreasing contexts. So far, I have only discussed matrix *why*-questions. In an embedded *why*-question, morphosyntactically simplex monotone increasing quantifiers still induce no intervention, as shown by the perfectly acceptable sentence as follows:

(5) Wo yijing zhidao-le {daduoshu ren/shaoshu ren} weishenme
I already know-PRF {most person/a.few person} why
cizhi.
resign
'I already knew for {a majority/a minority group of people}, why they resigned.'

More noteworthy is the fact that the *weak* intervention we witness in (4) disappears in embedded *why*-questions. This is demonstrated by the acceptability of (6).

(6) Wo yijing zhidao-le {san-ge ren/zhishao san-ge
I already know-PRF {three-CLF person/at.least three-CLF
ren/chaoguo san-ge ren} weishenme cizhi.
person/more.than three-CLF person} why resign
'I already knew for a group of (at least/more than) three people, why they resigned.'

By comparison, intervention cannot be circumvented in embedded contexts for monotone decreasing quantifiers. As (7) illustrates, the unacceptability in an embedded *why*-question is as strong as it is in a matrix one.

(7) #Wo yijing zhidao-le {meiyou ren/henshao ren/budao
I already know-PRF {no person/few person/less.than
san-ge ren} weishenme cizhi.
three-CLF person} why resign
#'I already knew for {nobody/few people/less than three people}, why they resigned.'

In sum, intervention effects in Chinese *why*-questions are sensitive to quantifier monotonicity. In addition, they are sensitive to whether *why*-questions occur in matrix or embedded contexts. The overall pattern is summarized in (8):

(8) Matrix and embedded *why*-questions:
1. Monotone decreasing quantifiers consistently induce intervention effects;
2. Non-monotone increasing, non-numeral quantifiers do not induce intervention effects;
3. (Monotone increasing) modified numerals and (non-monotonic) bare numerals induce weak intervention in matrix *why*-questions, which is ameliorated under embedded contexts.

Apart from quantificational DPs, adverbs of quantification exhibit similar patterns. (9) illustrates the ban for monotone decreasing quantificational adverbs to c-command *weishenme* 'why'.

(9) a. #Ta congbu weishenme cizhi?
He never why resign
#'On no occasions, why did he resign?'
b. #Ta henshao weishenme cizhi?
He seldom why resign
#'On few occasions, why did he resign?'

Furthermore, this ban on c-commanding quantificational adverbs is lifted if the adverbs are monotone increasing or non-monotonic:

(10) a. Ni dabufen shijian weishenme juede kun?
You most time why feel be.drowsy
'For most of the occasions, why did you feel drowsy?'
b. Wo yijing zhidao-le ta zhishao liang-ci weishenme
I already know-PRF he at.least two-token why
bu-gan zuo zhei-jian shi.
NEG-dare do DEM-CLF affair
'I already knew, for at least two occasions, why he wouldn't dare to do that.'

In this paper, I propose to account for this complex array of data in terms of the idiosyncratic semantics of *weishenme* 'why'. In a nutshell, I argue that Chinese *weishenme* must be initially merged at the high scope position of [Spec, CP]. When quantifiers are interpreted as taking wide scope over [Spec, CP], we obtain coherent interpretations. On the other hand, intervention arises when certain quantifiers are unable to be interpreted at such high scope. Hence, this account of intervention effects in *why*-questions does not involve 'real' intervention, in the sense that no mechanism of covert movement is assumed. Rather, my central claim in this paper is that the unacceptability we are dealing with here is not syntactic ill-formedness, but interpretational failure, i.e., a native speaker cannot assign an interpretation to a *why*-question in certain scopal relations.[4]

The rest of this paper is structured as follows. Section 2 reviews previous syntactic theories of the Chinese intervention effects in *why*-questions. Section 3 develops a semantic account with reference to *why*'s syntactic and semantic idiosyncrasies. Afterwards, I provide evidence that the intervention patterns of quantifiers correlate with quantifier monotonicity. Section 4 concludes the paper.

2 Past Accounts of the Quantifier-Induced Intervention Effects

In this section, I review several recent approaches to the Chinese intervention effects in *why*-questions that resort to covert LF movement. I then show that this line of research holds out little promise in accommodating the full range of data as discussed in the previous section. In the next section, I develop a semantic account that achieves the desired empirical coverage.

Building upon Beck [5] and Pesetsky [44], Soh [54] proposes that *in situ weishenme* 'why' undergoes covert feature movement at LF. According to Soh, intervention effects detect the movement of *wh*-feature, such that the feature cannot be separated from what's left behind on the *wh*-phrase by a scope-bearing element. Cheng [11] echoes Soh's solution, taking intervention effects as one crucial piece of evidence for the existence of covert feature movement.

Yang [63,64] reformulates the covert feature movement approach in terms of the framework of Relativized Minimality [47–49]. In a nutshell, intervention is a minimality effect, in which the quantificational 'likeness' between a quantifier and the interrogative phrase *weishenme* 'why' means that the feature of *weishenme* is attracted to the left periphery scope position only if it is closer to the scope position than the quantifier is. Yang borrows from recent works of Starke [55] and Rizzi [49] on Relativized Minimality and provides the following condition, in which the minimality effect is captured in terms of a filter:

(11) Maximal Matching Filter (Yang 2011, 63)
Let X and Y be bundles of features in a sequence of [...X...Y...]; Y cannot cross X when Y is maximally matched by X.

If a scopal element A bears feature [F1] and moves to its left periphery scope position, and if another scopal element B has the feature geometry that includes the bundle [F1 F2], then the movement of A from its initial merge position to its scope position is blocked because the bundle [F1 F2] *maximally matches* [F1].

[4] Consequently, I choose to put a # sign before unacceptable Chinese *why*-question sentences as well as their English translations to indicate that the examples are odd because the readings they generate are semantically anomalous. However, I still consistently use the term 'intervention effects' to refer to the types of phenomena that are already well established in the tradition, without taking this term in its literal sense.

In other words, the filter condition rules out the scope-taking of an operator at the left periphery when a 'like' operator is closer to the scope position of said operator.

The criteria of operator type matching are determined as follows (Rizzi [49]: 19):

(12) a. Argumental: person, number, gender, case
b. Quantificational: Wh, quantifier, measure, focus...
c. Modifier: evaluative, epistemic, Neg, frequentative, celerative, measure, manner, ...
d. Topic

Based on this classification, quantifiers as well as focus-sensitive phrases (focus) possess the same quantificational feature as the interrogative operator (Wh). Apart from the quantificational feature, quantifier/focus also bear other features. In a [quantifier $<$ Wh] configuration, the maximal matching filter is violated during the covert feature movement, because Wh's quantificational feature is maximally matched by the intervening quantifier.

Other *wh*-phrases such as *shenme* 'what' and *zenme* 'how' do not cause intervention in the same way as *weishenme* 'why' [56]. For Soh [54], the absence of intervention is because these *wh*-phrases undergo covert phrasal movement, rather than feature movement. In phrasal movement, entire *wh*-phrases are pied-piped across quantificational interveners. As such, there is no separation between *wh*-feature and the restriction on *wh*-phrases [44]. Yang [64] accounts for the absence of intervention by resorting to the mechanism of unselective binding [43]. For instance, Yang cites Cheng and Rooryck [12] and endorses the view that *wh*-phrases have the option of being licensed at a distance by a Q operator that merges directly at [Spec, CP]. According to this view, in *weishenme* 'why'-questions, intervention arises because the *weishenme*-adjunct does not possess this option, and ergo must be licensed via covert feature movement. In contrast, other *wh*-phrases can be licensed by unselective binding and undergo no movement, in which case the maximal matching filter is vacuously satisfied and no intervention arises. Note in addition that Yang's framework is also compatible with a covert phrasal movement solution: Pied-piped *wh*-phrases may be argued to bear more features than intervening quantifiers, therefore the maximal matching filter is not violated, unlike in feature movement.

The minimality-based approach as specified above is problematic upon closer scrutiny. This is because the minimality approach treats all quantifiers (both quantificational nominal phrases and adverbs of quantification) as legitimate interveners that block the LF movement of an interrogative operator. Quantifiers are interveners, simply because they bear a quantificational feature. Therefore, this approach would not predict the Chinese intervention pattern, where the intervention is sensitive to the types of quantifiers. Instead, the approach as it stands should predict that a finer distinction within quantifier types will not make any difference in intervention. If quantifiers in general possess enough features to maximally match the interrogative operator, then by including monotonicity as a further dimension in the feature geometry we only increase the inventory of the feature set for the quantifiers. Therefore, both monotone increasing and decreasing quantifiers are supposed to maximally match the interrogative operator and block its covert movement. Furthermore, it is rather stipulative if we bring monotonicity into our feature geometry, especially given that we find no independent evidence that monotonicity plays a role in other intervention environments (i.e., those involving non-*why* interrogative questions). Given the lack of apparatus to allow only a subset of quantifiers to block covert LF movement, it seems that the validity of a minimality account is in question. Finally, in embedded questions, a minimality account predicts that the covert interrogative operator still moves to take the embedded [Spec, CP] scope position (crossing the quantificational interveners along the way). Hence, even assuming that quantifier types can be fine-tuned to accommodate the intervention data in matrix questions that we have seen in (2)–(4), it is mysterious how a minimality account handles the selective amelioration phenomenon in the embedded questions of (5)–(7) in a principled manner.

The restricted set of quantificational interveners, i.e., downward quantifiers only, is reminiscent of another intervention environment that has received rich treatment, namely negative islands. It thus evokes the possibility that the intervention phenomenon in Chinese is subsumed under negative island sensitivity. A full survey of this connection is not available in the literature, in part due to the lack of dedicated literature of negative islands in Chinese. At present, I would like to point out that *why* is generally excluded from discussions of negative islands for being rather 'atypical'. Both Szabolcsi and Zwarts [58] and Abrusán [1] explicitly rule out *why*-questions in their theories of negative islands, noticing that *why* differs from other *wh*-adjuncts in that its extraction is blocked in a wider range of environments than others, suggesting that *why* independently favors late insertion/high attachment in the structure. The idiosyncratic structural property of *why* will be discussed in the following.[5]

[5] On a separate note, the modal obviation effect that is associated with negative islands (cf. Abrusán [1]) is absent in Chinese *why*-questions. In (ia), I show that adding the modal *keyi* 'can/might' circumvents the negative islands in a *how many*-question. In (ib), in contrast, I show that adding the same modal fails to improve a *why*-question.

(i) a. Zai zhongguo, meiyou ren keyi sheng duoshaoge haizi?
At China, no person can give.birth.to how.many children
'In China, how many children$_i$ can nobody give birth to t$_i$?'
b. #Zai zhongguo, meiyou ren weishenme keyi mianshui?
At China no personwhy can exempt.taxation
#'In China, why$_i$ can nobody be exempt from taxes t$_i$?'

If the modal obviation effects, as the majority of accounts of negative islands assume, serve as a diagnostic for islandhood in negative contexts, then the contrast in (ia-b) provides additional evidence that the intervention pattern witnessed in *why*-questions is a different beast.

3 A Semantic Account of the Intervention Effects in Chinese Why-Questions

3.1 The Syntax and Semantics of *weishenme* 'Why'

In this section, I build on previous observations that the reason/cause *wh*-adjunct *why* behaves in a different way from other *wh*-phrases. Following Ko [33], I assume that, crosslinguistically, *why*-adverbs favor high merge. Specifically, the East Asian (Chinese, Japanese, Korean, etc.) counterparts of *why* are directly merged at [Spec, CP], as opposed to other *wh*-phrases that are moved to [Spec, CP] from a lower initial merge position. In what follows, I cite a few published data that motivate the above treatment. As early as Lawler [40], it has been proposed that, in a *why*-question, *why* does not associate with any variables in the clause that it attaches to. For example, in the following mono-clausal sentence, it has been proposed that *why* does not bind a trace that links to the VP *leave* [48].

(13) Why did John leave early?

The no-trace property of *why* is seen more clearly in (14). As Lawler [40] points out, only one reading is available in the following quantificational environment:

(14) Why did three men leave?
Reading A: 'Why is it the case that three men left?'
Reading B: #'What reason$_i$ did three men have t_i for leaving?'

In reading A, an event, three men left, is presupposed. By wondering why this event occurs, we are committed to a situation in which the total number of people that left has to be three. In reading B, it is also the case that a group of three individuals left. Yet there is no requirement that, in this situation, altogether three people left. There could be other individuals who left, but for some reason the speaker is only concerned with a specific group of three people. When it happens that only three people left in the context, the two readings are not distinguishable. Crucially, however, when the context contains more than three individuals having left, the *why*-question in (14) cannot be uttered, at least according to the speakers Lawler [40] consulted.

Furthermore, it has been observed that *why* cannot be associated with the embedded clause (or the long-distance construal), and can only be associated with the matrix clause (or the short-distance construal). This can be exemplified by the examples in (15) [10,41].

(15) Why did you regret that Dr. Graff left the academia?
Reading A: 'What reason caused you to regret the fact that Dr. Graff left the academia?'
Reading B: #'What reason$_i$ did you regret that Dr. Graff have t_i for leaving the academia?'

Bromberger [9] argues that the above data would again follow if *why* merges directly to its scope position, and cannot be incorporated into the rest of the sentence by means of a trace. Bromberger points out a further piece of evidence, in which *why* and other *wh*-phrases interact with scopal elements such as focus operators in different ways.

(16) a. Why did ADAM eat the apples?
b. When did ADAM eat the apples?

Here I use small caps to mark that *Adam* is a focussed constituent. While (16a) presupposes that only Adam ate the apples, (16b) is compatible with the reading in which every individual ate the apples at different times, and the speaker is simply concerned with the time of Adam's eating event. Bromberger [9] argues that we can account for the reading in (16b) if we assume that *when* is base-generated in a position below the focus operator and that it binds a trace after it undergoes movement. Let's assume that the focus operator provides a focus value against a set of alternatives. That is, we first have a set of alternatives in the form of {*x eat the apples when* | x ranges over contextually relevant individuals}. The focus operator then applies to the set of alternatives, setting the value of x to Adam (In Bromberger's representation: (When$_i$) {(∃x: x = Adam){x ate the apples at t$_i$}}). On the other hand, if *why* leaves behind no trace and directly merges above the scope of the focus operator, then the focus value will be set to Adam first, before we use *why* to ask for the reason (In Bromberger's representation: (Why) {(∃x: x = Adam){x ate the apples}}). As a result, a *why*-question presupposes that only Adam, out of all individuals, ate the apples.

Related to the above observations, Tomioka (2009) demonstrates that, in downward entailing environments, *why* triggers different presuppositions from other *wh*-phrases. Compare (17a) with (17b), taken from Japanese.

(17) a. Daremo naze ko-nak-atta-no?
Anyone why come-NEG-PAST-Q
'Why did no one come?'
Presuppose: No one left.
Not Presuppose: There is a reason that no one left for.
b. Daremo nani-o yom-ana-katta-no?
Anyone what-ACC read-NEG-PAST-Q
'What$_i$ did no one read t$_i$?'
Presuppose: There is something such that no one read it.
Not Presuppose: No one read anything.

In line with the above observations, Tomioka formulates the following semantic constraint for *why*:

(18) Tomioka's constraint:
In a *why*-question and only in a *why*-question, the proposition that corresponds to the non-*wh* portion of the question must be presupposed.

This constraint calls for a high merge position of *why*, which Ko [33] assumes to be [Spec, CP]. Ko's proposal is exclusively about counterparts of *why* in East Asian languages such as Chinese, Japanese and Korean. Independently, Rizzi [48] argues that *perché* 'why' in Italian merges directly at [Spec, IntP]. Rizzi assumes that the head of IntP carries a [+wh] feature inherently, therefore this direct high merge explains why *perché* does not trigger auxiliary inversion. Given that there is no motivation for a structural distinction between [Spec, CP] and [Spec, IntP] in East Asian languages, we can essentially consider Rizzi's high attachment analysis of *perché* the same as Ko's proposal for East Asian *why*s. What is important for our current purpose is that both [Spec, CP] and [Spec, IntP] are higher than the scope positions of the focus operator and quantifiers at the left periphery (according to Rizzi), thus capturing the readings such as in (16) and (17).

3.2 Quantifiers as Plural Indefinites

If Chinese *weishenme* 'why' directly merges at [Spec, CP], it does not take part in quantifier scope interactions, because it is directly interpreted at a scope position above quantifier scope. Moreover, Chinese is known to observe a scope isomorphism at the left periphery, such that scopal relations at LF are preserved at surface syntax [2,21]. Unlike Japanese or Korean, Chinese quantifiers cannot scramble across outscoping operators to create a mismatch between word order and scope order [33]. Therefore, we would expect that quantificational elements, when taking scope as a generalized quantifier, be c-commanded by *weishenme*. However, in (19a-b), we see that *weishenme* and quantifiers may occur in two relative orderings.

(19) a. Weishenme daduoshu ren cizhi?
Why most person resign
'Why (is it the case that) most people resigned?'

b. Daduoshu ren weishenme cizhi?
Most person why resign
'For a certain plurality of individuals that is the majority of all the context-relevant individuals, why did they resign?'

In (19a), where *weishenme* c-commands the quantifier *duoshu ren* 'most people', we obtain an expected reading in which the latter denotes a standard GQ meaning, and *weishenme* takes the entire quantified proposition as its argument. Importantly, the question in (19b) does not seem to involve a generalized quantifier that scopes below *weishenme*. What (19b) asks is the reason that causes one particular plurality of individuals to resign, and this plurality has to be a majority subset of all the context-relevant individuals. For (19a), an answer can be given in the form of (20):

(20) Yinwei zhiyou shaoshu ren manyi gongsi de
Because only minority person be.satisfied.with company REL
xinchou daiyu.
pay treatment
'Since only a minority (of employees) were satisfied with the payroll of the company (and hence didn't resign).'

Meanwhile, (20) cannot be an answer for (19b). A felicitous answer must provide a reason of resignation for a *particular* plurality of individuals. Therefore, the reading of (19b) suggests that *most people* receives the interpretation of a plural indefinite and exhibits exceptional wide scope, above the scope of *why*, which is characteristic of plural indefinites. Both Reinhart and Winter have proposed that quantifier phrases such as *some people* or *many people* can be interpreted as plural indefinites, in which they do not denote a relation between predicates, in the traditional sense of Barwise and Cooper [4]. Rather, they denote individuals, by being coerced into a minimal witness set [19].[6]

In this paper, I propose that *most people* may also denote a plural indefinite. To go one step further, I argue that the plural indefinite *most people* is a topic when it takes wide scope over *weishenme* 'why'. That is, I believe that exceptional wide scope is a topic phenomenon [19]. A topical reading is possible for quantifiers interpreted as plural indefinites, because all referring expressions that are individual-denoting may serve as topics under the right contextual conditions. Importantly, I argue that topics are able to take scope outside of a speech act (that is, they may scope above the illocutionary operator of a sentence). As such, topics scope above the high initial merge position of *weishenme* in a *weishenme*-question. This accounts for the exceptional wide scope position of plural indefinites.

3.3 The Wide Scope Behavior of Topical Quantifiers: Some Evidence

Below I present evidence that topics are able to take scope outside speech acts. In the next section, I show that the ability for quantifiers to be topics depends on their monotonicity. Various authors have pointed out that if any part of a proposition is capable of scoping out of a speech act, it will have to be a topic [18,36,45]. This is because topic establishment is a separate speech act by itself. The idea that topics are assigned illocutionary operators of their own is first raised in Jacobs [31]. Jacobs points out that introducing a topic is an act of frame setting. In the following, I follow Krifka's recent position that natural language allows speech acts to conjoin. A topic-comment structure expresses two sequential, conjoined speech acts, comprising the topic's referring act, to be followed by a basic speech act (assertion, request, command, etc.) that is performed as an update on the referent established by the topic. Krifka [36] notes that, in English, overt devices are used to mark topics as scoping out of questions, commands and curses, such as the following:

(21) a. As for Al, Bill and Carl, which dishes did they make?
b. The hamburger, please hand it to me.
c. This guy, he should go to hell!

[6] Witness set refers to the plurality determined by the intersection of the restrictor and the nuclear scope. That is, given a quantificational determiner D, one predicate P and another predicate Q, D(P)(Q) gives rise to the witness set $W = P \cap Q$ [4,57].

According to Krifka, topics even have to scope out of speech acts, given that they function as a separate speech act. In Chinese, if we assume that the topic act conjoins with a subsequent request speech act performed by a *weishenme*-question, we would predict that all the expressions that may serve as topics may occur outside the scope of *weishenme* without causing intervention. This prediction is borne out. As (22) demonstrates, proper names, pronouns and temporal/locative adverbs can legitimately c-command *weishenme*. These are expressions that have long been known to allow for a topic reading [21,39].

(22) a. Lisi weishenme mei qu paobu?
Lisi why NEG go jogging
'As for Lisi, why didn't he go jogging?'

b. Zuotian/Zai na'er weishenme dajia xihuan chi kaorou?
Yesterday/LOC there why folks enjoy eat barbecued.meat
'As for {yesterday/there}, why do folks enjoy eating barbecued meat?'

Example (22) additionally shows that when multiple topics are co-occurring, they can all c-command *weishenme*. There seems to be a functionally based cognitive constraint preventing more than three topics from co-occurring in the same sentence in Chinese. But a sentence with three topics is marginally acceptable [62]. In such case, we also find a *weishenme*-question with three c-commanding topics acceptable:

(23) ?Zhe-chang yinyuehui ni mingtian weishenme yao qu?
This-CLF concert you tomorrow why will go
'(As for) This concert, (talking about) tomorrow, why will you go?'

Furthermore, in biscuit conditionals, an *if*-antecedent may co-occur with a *weishenme*-question as its consequent, illustrated in (24):

(24) Ruguo ni bu-jieyi wo wen dehua, ni weishenme cizhi?
If you NEG-mind I ask PRT, you why resign
'If you wouldn't mind me asking you, why did you resign?'

Various proposals have suggested that the antecedents of biscuit conditionals are topics [18,20], such that they scope out of the speech act performed by the consequents of the conditionals. If this is valid, then it is readily predicted by our proposal of topic act that the antecedent in (24) is able to scope above a *weishenme*-consequent.

Another prediction is that if an element is by nature not topical, it will never c-command *weishenme*. This would readily explain the fact that focus-sensitive expressions also induce intervention in *weishenme*-questions, since they are known to be strongly anti-topical [59]. The following example demonstrates that focus-sensitive phrases also induce intervention effects in *weishenme*-questions. Sentence (25a) is unacceptable, because *weishenme* is c-commanded by the focus sensitive *only*-NP. (25b) and (25c) are similarly unacceptable, when *weishenme* is c-commanded by the focus adverbial *zhi* 'only' and the focus particle *lian... ye/dou* 'even'.[7]

(25) a. #Zhiyou Lisi weishenme cizhi?
Only Lisi why resign
#'For only Lisi, why did he resign?'
b. #Lisi zhi weishenme cizhi?
Lisi only why resign
#'It is only the case that why Lisi resigned?'
c. #Lian Lisi ye/dou weishenme cizhi?
LIAN Lisi YE/DOU why resign
#'For even Lisi, why did he resign?'

Apart from topics, the second class of subsentential expressions that scopes out of illocution are the epistemic attitude adverbs such as *daodi* 'on earth' and *jiujing* 'frankly/honestly'. Importantly, this class of adverbs express epistemic attitude towards speech acts [21,22,30]. As such, they are speech act-level modifiers and take the illocutionary operator as their argument. Hence, they fall outside the scope of illocution. In (26), I show that both a speech-act adverb and a topic may precede *weishenme*:

(26) Ta jiujing/daodi weishenme cizhi?
He in.the.hell/honestly why resign
'As for him, why the hell did he resign?'/ 'As for him, honestly, why did he resign?'

A contrast exists between this class of speech act-level adverbs and proposition level attitude adverbs such as *yiding* 'definitely' and *kongpa* 'probably/most likely', as we can see below:

(27) #Ta yiding/kongpa weishenme cizhi?
He definitely/probably why resign
#'Definitely/Probably, why did he resign?'

Unlike *daodi/jiujing*, adverbs such as *yiding* 'definitely' indicate the speaker's attitude towards the propositional content or contents of smaller units, rather than the speaker's attitude towards the speech act. Interpreting the question operator within the scope of *yiding* creates a semantic anomaly, because such adverb is not compatible with taking question operators as arguments. In other

[7] In (25a), *zhiyou* 'only' forms a constituent with an NP and assigns focus value to the NP. In (25b), *zhi* 'only' is a focus adverb. The *lian* + NP + *ye/dou* construction in (25c) is often assumed to be the Chinese counterpart of the English focus-sensitive *even*-NP [27,42,52]. It seems that *lian* and *ye/dou* together contribute to the semantics of the English focus particle *even*, although the exact nature of the division of labor is still not clear. According to some analyses, *lian* assigns focus accent to the NP it combines with, and *ye/dou* is a maximality operator that overtly expresses the alternatives in the focus value [24].

words, an expression is able to precede *weishenme* if and only if it is able to take the *weishenme*-question's illocutionary operator in its scope. A speech-act level adverb does so by modifying the speech act itself. As such, it patterns with topics and does not cause intervention. Note that I have assumed all along that c-command relation mirrors scopal relation in the Chinese left periphery. This is because long-distance scrambling is impossible in Chinese [21,29,33]. Importantly, scrambled operators reconstruct their scopes at LF. In Japanese and Korean, when generalized quantifiers scramble across the *why*-adjunct at surface syntax, they reconstruct their scope at the trace position [32]. Because reconstruction is not available in Chinese, when quantifiers such as *meiyou ren* 'no one' c-commands *weishenme*, we cannot receive an interpretation in which *meiyou ren* is reconstructed below the scope of *weishenme*.[8]

3.4 Intervention as a Speech Act Constraint

In the above, I present evidence that topics (together with speech act-modifying epistemic attitude adverbs) are able to scope above speech act. In this section, I show that an exceptional wide scope theory of topics renders a straightforward explanation of the intervention in Chinese *why*-questions.

First, I briefly discuss how a scope theory of topics can be couched in a formally precise framework of speech act establishment and conjoining. Here I follow the Wittgensteinian view that the speech act of a sentence corresponds to a component of the sentence that combines with the sentence radical. The sentence radical can be seen as unsaturated unless attached to the speech act operator [3,7,13,38,60]. According to Krifka [36], we can define speech act as a semantic object with the basic type a. A speech act operator thus can be seen as taking as input a sentence radical and returning a speech act. For example, the assertion operator ASSERT is of type <<s,t>, a> (taking as input a proposition, and returning a speech act). The question operator REQUEST is of type <<<s,t>, t>, a> (taking as input a set of propositions, and returning a speech act). We further assume that natural language allows speech acts to conjoin. A topic-comment structure expresses two sequential, conjoined speech acts, comprising the referring act of a topic, to be followed by a basic speech act (assertion, request, command, etc.) that is performed against the referent as established by the topic. To capture a topic's referring act, Krifka also posits a referring speech act operator REF of type <e,a>. Finally, & is a conjunction operator that conjoins speech acts (type <a, <a, a>>). In the case where a question is structured into a topic and a comment question, the sentence performs a conjunction of topic establishment and request, represented as the following:

(28) $\text{REQUEST } (<\phi_{topic}, \psi_{comment}>) \rightarrow \text{REF}_x\ (\phi_{topic})\ \&\ \text{REQUEST}(\psi_{comment}(\text{X}))$

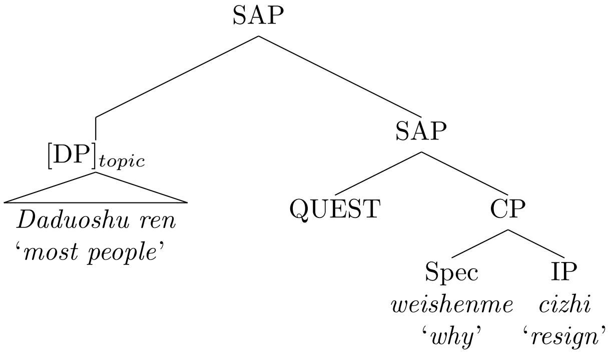
We can further incorporate speech act, as semantic objects with basic types, within the sentence grammar. Krifka [36] proposes that the speech act operator

[8] For further discussions on Japanese and Korean scrambling and reconstruction, see [14,26,51]. For the argument that Chinese does not allow scrambling, see Soh [53].

heads a Speech Act Phrase (SAP) projection that takes the sentence core (CP) as its complement. In the case of topicalization, Krifka proposes that SAPs can be recursively defined. The topic merges to the specifier of the first SAP, the head of which is occupied by another SAP, which is in turn headed by a basic speech act operator taking a CP complement. For instance, in the *why*-question (29a), I analyze the DP *daduoshu ren* 'most people' as a topical quantifier. Under this analysis, this sentence can be represented as (29b) (QUEST being the label used by Krifka for a request operator).

(29) a. Daduoshu ren weishenme cizhi?
Most person why resign
'For most people, why did they resign?'

b.



Finally, I provide a simplified semantics of topical quantifiers used as individual-denoting plural indefinites. To start with, I define a quantifier as witnessable if and only if the quantifier receives a plural indefinite reading, denoting its witness set [16,19,46].

(30) A quantifier is *witnessable* iff it entails the existence of a plurality that satisfies both the quantifier's restrictor and its nuclear scope, *i.e.* it entails the existence of its witness set.

Following Reinhart [46] and Winter [61], witnessable quantifiers denote type-e meaning via a covert choice function variable of type $<<e,t>,e>$ that, given a property (type $<e,t>$) as input, returns some plurality (type e) that has such property. The quantifiers in individual-denoting DPs are choice function modifiers that add a presuppositional restriction on the cardinality of the entity returned by the function. For example, *most* is represented as (31).

(31) $$[\![most]\!] = \lambda f_{<<e,t>,e>} \lambda P_{<e,t>}[f(P) \text{iff} |SUM(f(P))| > {}^{1}/_{2} |y : atom(y) \wedge P(y)|]$$

Here SUM is defined over pluralities that consist of atom individuals. Given a plurality, it outputs the set of all the atoms in the plurality. The witnessable quantifier *most people* denotes the plurality returned by the choice function f when applied to the property of being a majority of all the context-relevant individuals, represented as follows:

(32) $[\![most\ people]\!] = [\![[\mathrm{f}\ most]\ people]\!]$
$= f(\lambda x_{\mathrm{e}}\,[people(\mathrm{x}) \wedge |\mathrm{SUM(x)}| > {}^{1}/_{2}|\mathrm{y} : atom(y) \wedge person(\mathrm{y})|])$

The alternatives generated by '[f *most*] *people*' are computed by substituting different choice function variable values in the position of [f *most*]. Combining these alternatives with the restrictor *people*, we produce contrasting pluralities of individuals, each of them contain a majority of all the context-relevant individuals. Crucially, I claim that whereas monotone increasing and non-monotonic quantifiers are witnessable, monotone decreasing quantifiers are not witness-able. A non-witnessable quantifier, such as *few people*, may have a verifiable, non-empty witness set. However, it does not make reference to its witness set by denoting any choice-function selected pluralities.[9]

Now we can derive intervention effects from the interaction of topicalization, conjoined speech acts and witnessability. In a nutshell, if a quantifier is witnessable and hence is able to be construed as topical, it may scope above *weishenme*. On the other hand, if a quantifier cannot be construed as topical, outscoping would be impossible, due to *why*'s high scope. Intervention effects would arise in such cases, because for the non-topicalizable quantifier, the ordering of the quantifier preceding *weishenme* is impossible, hence semantically anomalous. The so-called intervention effects arise when an expression that cannot scope above *why* nevertheless occupies a wide scope position. In other words, there is no 'real' intervention involved here. Rather, the intervention in *why*-questions should be better characterized as a scope effect. In (33a), the *why*-question with the quantifier *daduoshu ren* 'most people' is acceptable as it is interpreted with the semantics in (33b). I also provide a less formal paraphrase of the question's meaning in (33c):

(33) a. Daduoshu ren weishenme qu?
Most person why go

b. Semantics:
REF_{y} (y = f($\lambda \mathrm{x}_e$[people(x) ∧ |SUM(x)| > ${}^{1}/_{2}$|y: atom(y)∧person(y)|])) & REQUEST (λq ∃r [q = λw [r CAUSE p in w ∧ p = λw' go (y)(w')]])

c. Paraphrase:
'(Speaking of/As for) the plurality returned by the choice function f when applied to the property of being a majority of all the context-relevant individuals, why are they going?'

[9] Independently, experimental results show that the monotonicity of a quantifier affects its ability to entail a witness set due to processing reasons [8,23]. To verify a quantified sentence containing *most* or *more than two*, one needs to find positive instances that members within the restrictor set satisfy the *most*-relation, the *more-than-two*-relation, etc. In other words, one needs to verify the existence of a witness set. In contrast, for quantified sentences with *no, few, or less than two*, the verification procedure more often requires drawing a negative inference based on the absence of positive instances (in which case the witness set is empty). Although there is still a paucity of relevant work on this topic, the intuition is that monotone decreasing quantifiers are not an informative way to denote a witness set.

On the contrary, the *why*-question with the quantifier *henshao ren* 'few people' is unacceptable because *henshao ren* cannot be a topic. That is, (34a) does *not* have the interpretation in (34b). Also, the paraphrase in (34c) is an impossible one:

(34) a. #Henshao ren weishenme qu?
Few person why go
b. Not compatible with the semantics:
REF_y (y = f(λx_e[people(x) $\wedge$ |SUM(x)| < $^1/_2$|y: atom(y) $\wedge$ person(y)|]))
& REQUEST (λq $\exists$r [q = λw [r CAUSE p in w $\wedge$ p = λw' go (y)(w')]])
c. Paraphrase:
#'(Speaking of/As for) the plurality returned by the choice function f when applied to the property of being few of all the context-relevant individuals, why are they going?'

In sum, when we consider quantifiers in terms of topicality, we immediately explain why monotone decreasing quantifiers induce intervention effects in *weishenme*-questions: they cannot be topical, hence they cannot give rise to coherent readings in *weishenme*-questions. Non-decreasing quantifiers are unproblematic, because they denote individuals that serve as topics.[10]

Furthermore, this theory claims that bare numerals and monotone increasing modified numerals can be topics. We still need to explain why these numeral quantifiers induce weak intervention, as seen in (35) (repeated from 4):

(35) ??{San-ge ren/ zhishao san-ge ren/ chaoguo san-ge
{Three-CLF person/ at.least three-CLF person/ more.than three-CLF
ren} weishenme cizhi?
person} why resign
'For three people/at least three people/more than three people, tell me why they resigned?'

[10] We should expect that the topicality constraint thus formulated applies even in the absence of *weishenme* 'why', since the topic position is generally available. This prediction is borne out. As mentioned above, the class of epistemic attitude adverbs such as *daodi* 'on earth' and *jiujing* 'frankly/honestly' take scope above speech act operators. This class of adverbs can be used to identify topic positions, in the absence of *weishenme* 'why', because when a quantified expression precedes this class of adverbs, the quantified expression has to reside outside the speech act of the sentence it occurs with and thus must receive a topical reading rather than a GQ reading. Importantly, as (i) shows, monotone decreasing quantifiers induce intervention when they precede epistemic adverbs even in non-*why* questions. Intervention is absent for non-decreasing quantifiers.

(i) a. *Budao san-ge ren daodi/jiujing qu na'er le?
Less.than three-CLF person on.earth/honestly go where PRT
'For less than three people, where on earth did they go?'
b. Daduoshu ren daodi/jiujing qu na'er le?
Most person on.earth/honestly go where PRT
'For most people, where on earth did they go?'

It thus seems that we can indeed reduce the 'intervention' in *why*-questions to a broad phenomenon of topicalizability.

I believe the weak acceptability in (35) has a pragmatic reason. Following Kratzer [34,35], I assume that choice function variables receive their values directly from the context of utterance. If context does not readily offer a particular plurality as the value for a choice function variable, the speaker won't know which plurality to pick out with the quantifier, and oddness arises. In the case of numeral quantifiers, we are required to pick out a particular plurality bearing a specific cardinal number, which would leave the hearers with no clues if there is no further information from the context. Krifka [36] observes the same problem for the English example in (36):

(36) ??Which dishes did two boys make?
'For two boys that you select: Which dishes did each of these boys make?'

The acceptability is claimed by Krifka to be marginal. This low acceptability of *two boys*, compared to phrases such as *most boys*, follows from the fact that it places a higher requirement on the discourse structure and on hearers' efforts to infer which particular set of two boys are under discussion. Similarly, we can explain why the topical use of quantifiers containing a numeral component is harder. Without explicit context providing supporting information, it is not plausible for a naive hearer to make a partition of the relevant individuals such that one particular plurality of a given cardinality should be distinguished against other individuals.

The context-based claim I have argued above predicts that *why*-questions with witnessable numeral quantifiers should be acceptable in a plausible scenario. This seems to be indeed the case, as the following example demonstrates.[11]

(37) (A soccer coach needed a minimum of three more healthy players to fill up his squad for a match. He felt frustrated that the scheduled operations on his injured players were two months away.)

Shangyuan	li	de	zhishao	san-ge	weishenme	bu	neng
Injured.players	inside	POSS	at.least	three-CLF	why	NEG	can

xian	shoushu?
first	operate

'For at least three of the injured players, why can't they be operated on first?'

Finally, embedded questions may offer the contextual information to anchor a particular plurality [57]. I will illustrate with the example in (38) (repeated from example (6)):

[11] According to my consultants, if we use a non-partitive form *zhishao san-ge shangyuan* 'at least three injured players', the sentence is still mildly acceptable, but nowhere close to the fine judgments we are getting with the partitive quantified expression in (37). Note that Constant [15,17] also notices (without suggesting an explanation) that partitive forms of quantifiers more readily license a referential reading than non-partitive forms. At present, I do not know how to account for this, and have to leave an answer to future work.

(38) (In a report investigating employees' resignation)
Wo yijing zhidaole {chaoguo san-ge ren/zhishao san-ge
I already know {more.than three-CLF person/at.least three-CLF
ren/san-ge ren} weishenme cizhi.
person/three-CLF person} why resign

'I already found out for more than three people/at least three people/three people, why they resigned.'

The indirect question that serves as the complement of *found out* does not denote a question type, but rather a fact derived from a question [25,37]. Specifically, the indirect question is construed as a true answer (true resolution) to the corresponding direct question. Thus, (38) is paraphrased as follows: 'I already found out (the answer to the question of) for three people, why they resigned.' Following Rooth [50], this indirect question intuitively answers one subquestion of the overall question: 'Why did a contextually-salient set of individuals resign?' In order to answer this overall question based on the knowledge of the speaker, the question is partitioned into two contrasting subquestions. The first asks about a plurality consisting of three people, of whom the speaker has knowledge about. The other asks about 'the rest of the individuals' of whom the speaker does not provide an answer due to lack of knowledge.

3.5 Further Evidence for the Type-e Meaning of Topical Quantifiers

In this section, I present evidence that the topicality of quantifiers correlates with their monotonicity. My diagnostics are based on Constant [15,17]. First, Constant notices that only witnessable quantifiers (monotone increasing and non-monotonic) may serve as contrastive topics. In (39), I put forward Chinese data in support of Constant's claim (CT for contrastive topic, F for focus):

(39) A: Yanjiusheng-men zhu zai na'er?
Graduate.student-PL live LOC where?
'Where do the grads live?'

B: [{Daduoshu/Wu-ge/#Henshao yanjiusheng}]^{CT} zhu zai [anhesite]^{F}.
Most/Five-CLF/#Few graduate.student live LOC Amherst
'[{Most of/Five of/#Few of the graduate students}]^{CT} live at [Amherst]^{F}.'

In (39), monotone increasing quantifiers serve as contrastive topics, but monotone decreasing quantifiers cannot. If CT-marked quantifiers such as *most* only have a standard GQ reading, they would be construed as answering one of the subquestions of question A. These subquestions would be the alternatives in {*Where did most grads live? Where did a few grads live? Where did no grads live?...*}[12] This does not accord with our intuition, in which B's answer means that B has information about where a majority subset of individuals live, as opposed to the rest

[12] See Rooth [50] for a discussion of how contrastive topic-marked answer is answering a subquestion of a preceding overall question.

of the individuals about whom B has no information. If *most grads* denotes a specific plurality of individuals, then the contrasting alternatives will be between different individual grads. This seems to be exactly what (39) does. Furthermore, if CT-marked quantifiers are standard GQs, it would be mysterious why quantifiers such as *few* cannot form an answer. If we subscribe to a choice functional approach, on the other hand, the reason is obvious, since *few* cannot denote a choice-function-selected plurality. If quantifiers such as *few* lack choice-functional interpretations, then an answer in (39B) with *few* only has the standard GQ reading. If we assume that CT is simply unable to contrast quantifiers of this type, then the sentence will be ruled out.

One further piece of evidence given by Constant is that quantifiers differ in their ability to appear in equative copular constructions: In an equative construction, the two-place copula *be* equates two individual-denoting expressions. On the left side, the first argument of the copula is a type-e plurality DP. For the equative construction to be well-formed, the right argument needs also to be an individual-denoting plurality DP. Therefore, the equative construction provides yet another diagnostic on which quantifier qualifies as type-e denoting. As it turns out, the judgment patterns in (40) match well with the patterns we have seen in the contrastive topic diagnostic.

(40) [Zhan zai na'er de ren] shi [wo de xuesheng li de
Stand LOC there REL person COP I REL student inside REL
{daduoshu/wu-ge/#henshao}].
{most/five-CLF/#few}
'[Those standing over there] are [most/five/#few of my students].'

4 Conclusion

This paper develops an account of intervention effects with Chinese *weishenme* 'why' and monotone decreasing quantifiers. The empirical generalization is that monotone decreasing quantifiers cannot scope above *weishenme* at surface, with *weishenme* 'intervening' between those quantifiers and the rest of the sentence. My take on this issue is to propose a new way of looking at things. *Weishenme* is not only *in situ*, but also at the position where it, syntactically speaking, checks off the *wh*-feature, and where it, semantically speaking, is interpreted. Materials to the left can only be interpreted as topics, giving rise to a secondary speech act in the sense of Krifka [36]. Using a notion of topicality involving witnessability (in the sense of Reinhart [46]), I then derive the quantifier restriction for this position based on which determiners can lead to witnessability, thus excluding monotone decreasing quantifiers. Quantificational expressions with monotone increasing numerals, as well as bare numerals, are also not acceptable in apparent intervention configurations, unless these sentences are embedded. I argue that this is due to the lack of context in root sentences, thus leaving the choice function variables without a value. In sum, the current analysis combines relatively independently but under a theoretical perspective disparate ideas, and arrives at a novel and simple solution to a rich array of empirical facts.

References

1. Abrusán, M.: Presuppositional and negative islands: a semantic account. Nat. Lang. Semant. **3**(19), 257–321 (2011)
2. Aoun, J., Li, A.: On some differences between Chinese and Japanese wh-elements. Nat. Lang. Semant. **24**(2), 365–372 (1993)
3. Åqvist, L.: A new approach to the logical theory of actions and causality. In: Stenlund, S., Henschen-Dahlquist, A.-M., Lindahl, L., Nordenfelt, L., Odelstad, J. (eds.) Logical Theory and Semantic Analysis, pp. 73–91. Springer, Dordrecht (1974)
4. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguist. Philos. **4**(2), 159–219 (1981)
5. Beck, S.: Intervention effects follow from focus interpretation. Nat. Lang. Semant. **14**(1), 1–56 (2006)
6. Beck, S., Kim, S.-S.: Intervention effects in alternative questions. Nat. Lang. Semant. **9**(3), 165–208 (2006)
7. Belnap, N.: Questions: their presuppositions, and how they can fail to arise. In: Lambert, K. (ed.) The Logical Way of Doing Things, pp. 23–37. Yale University Press, New Haven (1969)
8. Bott, O., Klein, U., Schlotterbeck, F.: Witness sets, polarity reversal and the processing of quantified sentences. In: Aloni, M., Franke, M., Roelofson, F. (eds.) Proceedings of the 19th Amsterdam Colloquium, pp. 59–66 (2013)
9. Bromberger, S.: On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them. University of Chicago Press, Chicago (1992)
10. Cattell, R.: On the source of interrogative adverbs. Nat. Lang. Semant. **54**(1), 61–77 (1978)
11. Cheng, L.: Wh-in-situ, from the 1980s to now. Nat. Lang. Semant. **3**(3), 767–791 (2009)
12. Cheng, L., Rooryck, J.: Licensing wh-in-situ. Nat. Lang. Semant. **3**(1), 1–19 (2000)
13. Chierchia, G.: Questions with quantifiers. Nat. Lang. Semant. **1**(2), 181–234 (1993)
14. Choe, H.S.: Syntactic wh-movement in Korean and licensing. In: Theoretical Issues in Korean Linguistics, pp. 275–302 (1994)
15. Constant, N.: On the positioning of Mandarin contrastive topic -ne. In: (Workshop) Information Structure, Word Order: Focusing on Asian Languages (2013)
16. Constant, N.: Witnessable quantiers license type-e meaning: Evidence from contrastive topic, equatives and supplements. In: Proceedings of Semantics and Linguistic Theory, pp. 286–306 (2013)
17. Constant, N.: Contrastive topic: meanings and realizations. Ph.D. thesis (2014)
18. Ebert, C., Ebert, C., Hinterwimmer, S.: A unified analysis of conditionals as topics. Linguist. Philos. **37**(5), 353–408 (2014)
19. Endriss, C.: Exceptional wide scope. In: Endriss, C. (ed.) Quantificational Topics. Springer, Dordrecht (2009)
20. Endriss, C., Hinterwimmer, S.: Direct and indirect aboutness topics. Linguist. Philos. **55**(3–4), 297–307 (2008)
21. Ernst, T.: Conditions on Chinese A-not-A questions. J. East Asian Linguist. **3**(3), 241–264 (1994)
22. Ernst, T.: The Syntax of Adjuncts. Cambridge University Press, Cambridge (2001)
23. Geurts, B., van der Slik, F.: Monotonicity and processing load. J. East Asian Linguist. **22**(1), 97–117 (2005)

24. Giannakidou, A., Cheng, L.: (In) definiteness, polarity, and the role of wh-morphology in free choice. J. East Asian Linguist. **23**(2), 135–183 (2006)
25. Ginzburg, J., Sag, I.: Interrogative Investigations. CSLI Publications, Stanford (2000)
26. Grewendorf, G., Sabel, J.: Scrambling in German and Japanese: adjunction versus multiple specifiers. J. East Asian Linguist. **17**(1), 1–65 (1999)
27. Hole, D.: Focus and Background Marking in Mandarin Chinese. Routledge, London (2004)
28. Horn, L.: The border wars: a Neo-Gricean perspective. In: von Heusinger, K., Turner, K. (eds.) Where Semantics Meets Pragmatics, pp. 21–48. Elsevier, Amsterdam (2006)
29. Huang, J.: Move WH in a language without WH movement. J. East Asian Linguist. **1**(4), 369–416 (1982)
30. Jackendoff, R.: Semantic Interpretation in Generative Grammar. MIT Press, Cambridge (1972)
31. Jacobs, J.: Funktionale satzperspecktive und illokutionssemantik. Linguistische Berichte **91**, 25–58 (1984)
32. Kitagawa, Y.: Anti-scrambling. Manuscript, University of Rochester (1990)
33. Ko, H.: Syntax of why-in-situ: merge into. Nat. Lang. Linguist. Theor. **23**(4), 867–916 (2005)
34. Kratzer, A.: Scope or pseudoscope? Are there wide-scope indefinites? In: Rothstein, S. (ed.) Events and Grammar, pp. 163–196. Springer, Dordrecht (1998)
35. Kratzer, A.: A note on choice functions in context. Manuscript, UMass Amherst (2003)
36. Krifka, M.: Quantifying into question acts. Nat. Lang. Semant. **9**(1), 1–40 (2001)
37. Lahiri, U.: Questions and Answers in Embedded Contexts. Oxford University Press, Oxford (2002)
38. Lang, E., Steinitz, R.: Können Satzadverbiale performativ gebraucht werden. In: Motsch, W. (ed.) Kontexte der Grammatiktheorie, pp. 51–80. Akademie Verlag, Berlin (1978)
39. Law, P.: Adverbs in A-not-A questions in Mandarin Chinese. J. East Asian Linguist. **15**(2), 97–136 (2006)
40. Lawler, J.: Any question, pp. 163–173 (1971)
41. Oshima, D.Y.: On factive islands: pragmatic anomaly vs. pragmatic infelicity. In: Washio, T., Satoh, K., Takeda, H., Inokuchi, A. (eds.) JSAI 2006. LNCS (LNAI), vol. 4384, pp. 147–161. Springer, Heidelberg (2007). doi:10.1007/978-3-540-69902-6_14
42. Paris, M.-C.: Nominalization in Mandarin Chinese: the morphene "de" and the "shi" ... "de" constructions. Ph.D. thesis, Université de Paris 7 (1979)
43. Pesetsky, D.: Wh-in-situ: movement and unselective binding. In: Reuland, E., ter Meulen, A. (eds.) The Representation of (In) Definiteness, pp. 98–129. MIT Press, Cambridge (1987)
44. Pesetsky, D.: Phrasal Movement and Its Kin. MIT Press, Cambridge (2000)
45. Reinhart, T.: Pragmatics and linguistics: an analysis of sentence topics in pragmatics and philosophy. J. East Asian Linguist. **27**(1), 53–94 (1981)
46. Reinhart, T.: Quantifier scope: how labor is divided between QR and choice functions. Linguist. Philos. **20**(4), 335–397 (1997)
47. Rizzi, L.: Relativized Minimality. MIT Press, Cambridge (1990)
48. Rizzi, L.: On the position Int(errogative) in the left periphery of the clause. Linguist. Philos. **14**, 267–296 (2001)

49. Rizzi, L.: Locality and left periphery. In: Belletti, A. (ed.) Structures and Beyond: The Cartography of Syntactic Structures, vol. 3, pp. 223–251. Oxford University Press, Oxford-New York (2004)
50. Rooth, M.: Topic accents on quantifiers. In: Carlson, G., Pelletier, F.J. (eds.) Reference and Quantification: The Partee Effect, pp. 1–23. CSLI Publications, Stanford (2005)
51. Saito, M.: Long distance scrambling in Japanese. J. East Asian Linguist. **1**(1), 69–118 (1992)
52. Shyu, S.-I.: The syntax of focus and topic in Chinese. Ph.D. thesis, University of Southern California (1995)
53. Soh, H.L.: Object scrambling in Chinese. Ph.D. thesis, Massachusetts Institute of Technology (1998)
54. Soh, H.L.: Wh-in-situ in Mandarin Chinese. J. East Asian Linguist. **36**(1), 143–155 (2005)
55. Starke, M.: Move dissolves into merge: a theory of locality. Ph.D. thesis, University of Geneva (2001)
56. Stepanov, A., Tsai, W.-T.D.: Cartography and licensing of wh-adjuncts: a cross-linguistic perspective. Nat. Lang. Linguist. Theor. **26**(3), 589–638 (2008)
57. Szabolcsi, A.: Quantification. Cambridge University Press, Cambridge (2010)
58. Szabolcsi, A., Zwarts, F.: Weak islands and an algebraic semantics for scope taking. J. East Asian Linguist. **1**(3), 235–284 (1993)
59. Tomioka, S.: Pragmatics of LF intervention effects. J. Pragmat. **39**(9), 1570–1590 (2007)
60. Wachowicz, K.: Q-morpheme hypothesis, performative analysis and an alternative. Questions, pp. 151–163. Springer, Dordrecht (1978)
61. Winter, Y.: Choice functions and the scopal semantics of indefinites. Linguist. Philos. **20**(4), 399–467 (1997)
62. Xu, L.: Topicalization in Asian languages. In: van Riemsdijk, H., Everaert, M. (eds.) The Blackwell companion to syntax, pp. 137–174. Wiley Blackwell, London (2006)
63. Yang, B.: Intervention effects and the covert component of grammar. Ph.D. thesis, National Tsinghua University (2009)
64. Yang, B.: Intervention effects and wh-construals. J. East Asian Linguist. **21**(1), 43–87 (2011)

Syntax Annotation of the Georgian Literary Corpus

Liana Lortkipanidze(✉), Nino Amirezashvili, Ana Chutkerashvili, Nino Javashvili, and Liana Samsonadze

Archil Eliashvili Institute of Control Systems of the Georgian Technical University, Tbilisi, Georgia
l_lordkipanidze@yahoo.com, ninomaskh@yahoo.com, annachutkerashvili@yahoo.com, ninojavashvili@yahoo.com, liasams@yahoo.com

Abstract. It is very important to draw out deeply annotated text corpora in order to solve theoretical and applied tasks of the Georgian language. While syntactically annotated corpora are now available for English, Czech, Russian and the other languages, for Georgian they are rare. The environment, developed by our research group, offers several NLP applications, including a module of the morphologic, syntactic and semantic level, a Universal Networking Language interface and a natural language interface to access SQL type databases.

The paper gives the description of the automatic syntactic analyzer of the Georgian Language. It includes syntactic and morphologic levels of the Georgian language model. The basis of the linguistic model of the Georgian text syntax annotation is the dependency grammar.

Keywords: Morphological analysis · Parsing · Syntactic relation · Dependent term · Dominant term

1 Introduction

At present, over 120 corpora for the European and the other languages are available; the largest of them containing hundreds of millions of words [1–3]. As for Georgian, annotated corpora did not exist until 2013 when the first version of the Georgian Dialect Corpus was compiled [4]. Since then, the Georgian corpus linguistics has been developed rapidly and several groups of researchers have been working on it.

In order to solve theoretical and applied tasks of the Georgian language it is very important to draw out deeply annotated textual corpus. During the last few years, morphologic, syntactic and semantic marking of the Georgian literary texts has been worked up in the department of Language and Speech Systems at Archil Eliashvili Institute of Control Systems of The Georgian Technical University. In the pilot version of the corpus all the novels of Otar Tchiladze, the famous writer of the XX century have been analyzed.

The following three levels have been envisaged – Morphologic, Syntactic and Semantic. The corpus annotation is different on each level:

Morphologic annotation: for every word, along with the normal form (lemma) and the part of speech, a complete set of morphological attributes is specified;

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 89–97, 2017.
DOI: 10.1007/978-3-662-54332-0_6

Syntactically tagged texts: every word is ascribed a syntactic relations marker and a syntactic link marker.

Semantically tagged texts: every word is ascribed a semantic marker.

Semantic markers of the word contain information about the field of the usage of the word. Marker points at the words that can be combined with the main word according to their meaning and shows what kind of semantic relation is between them. Lexical-semantic information that is ascribed to a word in the texts is categorized into the following order: 1. The main characteristics of a word – part of speech; 2. Lexical-semantic information; 3. Word building (derivation) features [5].

The Georgian texts are annotated with dependency structures – formalism is considered more suitable for the Georgian language with its relatively free order of words than constituent structures. The structure not only contains information to which words of the sentence are syntactically linked, but also relegates each link to one of the several dozen syntactic types. This is an important feature, since the majority of syntactically annotated corpora (both already available and under construction) represent the syntactic structure by means of constituents.

The closest analogues to our work are Dependency Treebank for Russian linguistic corpora [3] and Prague Dependency Treebank (PDT) – an annotated corpus of Czech collected at Charles University in Prague [6].

Below markup format (Sect. 2), annotation tools and procedures (Sect. 3), and types of linguistic data included in the markup (Sect. 4) are described.

2 Markup Overview

We have tried to design syntactic analysis of a sentence the way that clearly introduces distinguished syntactic pairs, or relations between the members of a sentence; at the same time it enables to restore its structural tree.

The most natural solution to meet these representations is an XML-based markup language. We have tried to make our format compatible with TEI [7], introducing new elements or attributes only in situations where TEI markup does not provide adequate means to describe the text structure in the dependency grammar framework.

The types of information about text structure that must be encoded in the markup and the respective tags/attributes used for carrying this information are listed below.

A special container element <S> is used to delimit sentence boundaries. The element has an attribute ID that supplies a unique identifier for the sentence within the text; the identifier may be used for storing information about extra-sentential relations in the text. There are two other attributes attached to the sentence: STATUS is used to visualize the respective sentence as one needing further (or no) editing and WNUM is used to show the word number in the sentences.

It also has an attribute – COMMENT, used by linguists for storing notes and observations on a particular syntactic phenomena encountered in the sentence.

The words are demarcated by a container element <W>. Like sentences, words may have a unique ID attribute that is used for referring the position of word within the sentence.

Morphological information is ascribed to the word by means of two attributes attached to the <W> tag: LEMMA – normalized word form and MOFE – list of morphological features.

To annotate the information about the syntactic dependencies, we use two other attributes attached to the <W> element: DOM is the ID of the dominant word, W depends on it and SYFU is the syntactic relation label.

3 Annotation Tools and Procedures

In the input of the system, there is a corpus. As an output, a linguist gets the text divided into the sentences, where the title form, morphologic characteristics and syntactic characteristics has been added to each word-form. The relations that provide the connection of a word-form to the other members of a sentence condition the syntactic characteristics.

The procedure of the corpus data acquisition is semi-automatic. A computer using a morphological analyzer of general purpose and syntax parser engine generates an initial version of markup; after that, the results of the automatic processing are submitted to human post-editing.

This has become necessary because the automatically annotated text contains all possible versions of morphologic and syntactic parsing and ambiguity must be removed manually. The system marked 60% of the words with homonymous morphological markers, 28% – with no homonymous markers in the texts we have processed. Among the 12% of the remaining words, some were the author's occasional forms and the others were misprints.

To support the creation of annotated data, a variety of tools have been designed and implemented. All tools are Windows applications written in C++.

The sentence border marking program converts texts recorded in Word – RTF format into an SQL database table format. All sentences are given a unique identifier by referencing the name of the texts (see Fig. 1). Selection of a sentence in the text editor for processing according to the texts recorded in the base of the corpus is possible. In the lower window of Fig. 1 information according to the sentences of the texts in the form of the table, as well as the selected sentence in the middle of the window, is displayed. "S_ID" indicates the identifier of a sentence in the text.

"St" – status of the sentence processing: "Y" shows that everything is all right and "?" shows that something is wrong. A sentence is written in the field of the "Sentences". The editor can move to the grammatical annotation and can correct annotation results by double-clicking this field; the number of the words in the sentence is written in the field of "WNum". The editor can write any comments in the field of "Comments".

The amount of manual labor required for building annotations depends on the complexity of the input data. Most sentences can be reliably processed without any human intervention; in this case, a linguist should only look through the result of the processing and endorse it. If the Sentence Properties contain errors, a linguist can edit it using a user-friendly interface (see screenshot below). This mode involves manual correction of the morphologic or the syntactic markers to every word in a sentence.

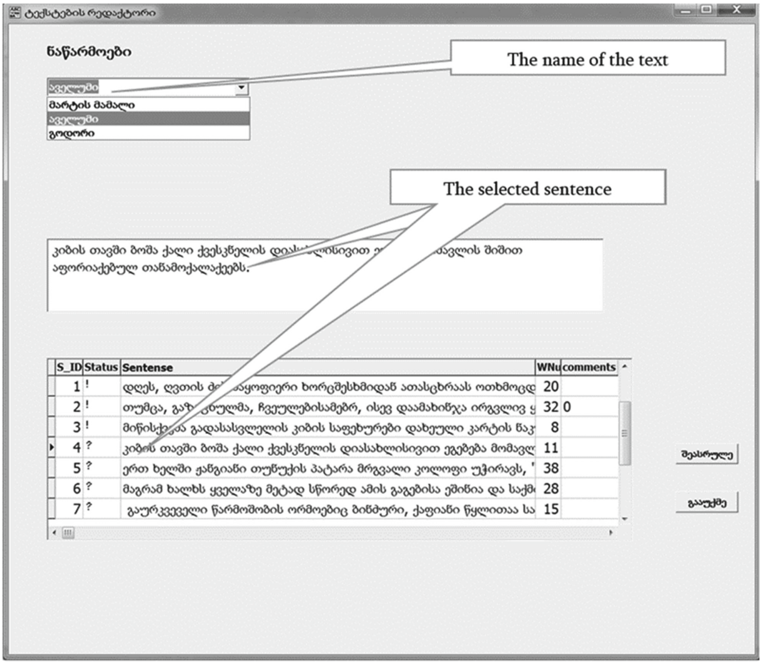


Fig. 1. The window of the text-processing program

If a linguist is uncertain whether the morphologic/syntactic properties of the word are adequate, he/she may mark it as "doubtful". For that he/she has to write the sign "?" in the field of "St" (status).

Figure 2 presents the main dialog window for editing sentence properties. An operator can edit the markups directly in any sentence of the text. The selected sentence for analysis is written in the top line of the edit window:

k'ibis tavši boša kali kvesk'nelis diasaxlisivit egebeba momavlis šišit aporiakebul tana-mokalakeebs – 'At the top of the stairs a gypsy woman welcomes her citizens, who are anxious with fear about the future, like an underworld hostess'.

If a linguist is uncertain whether the morphologic/syntactic properties of the word are adequate, he/she may mark it as "doubtful". For that he/she has to write the sign "?" in the field of "St" (status).

Figure 2 presents the main dialog window for editing sentence properties. An operator can edit the markups directly in any sentence of the text. The selected sentence for analysis is written in the top line of the edit window:

k'ibis tavši boša kali kvesk'nelis diasaxlisivit egebeba momavlis šišit aporiakebul tana-mokalakeebs – 'At the top of the stairs a gypsy woman welcomes her citizens, who are anxious with fear about the future, like an underworld hostess'.

How we know, an ambiguous form is one that can have two or more meanings. You can see more than one dictionary entry with the same lexeme in the program tool. Each entry represents one meaning. When an ambiguous form occurs in the text, program asks you to choose between the multiple features.

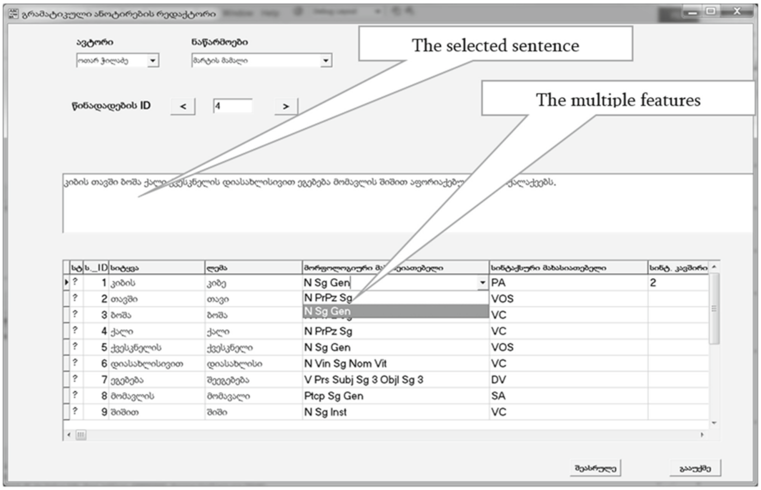


Fig. 2. The main dialog window for editing of the sentence properties

The information about particular words is written in a list: for this example, the first word *k'ibis* 'of the stairs' has an identifier ID = 1; the lemmatized form is *k'ibe* 'stair'; its morphological feature list – [N Sg Gen] consists of: part-of-speech – N (noun), number.

– Sg (singular) and case – Gen (genitive case). The word depends on a word with ID = 2 with the adverbial syntactic relation (link type is Att(N) + Adv(N)).

The editor interface is simple. All fields can be edited except the first and the second (the words of the sentence and their place number in the sentence). With double-clicking the field of characteristic the word properties list will pop up. The editor can see all morphologic/syntactic properties of the single words (see Fig. 2).

4 Types of Linguistic Data Included in the Markup

The morphologic analysis we use is based on the computational dictionary. Lists of lemmas and affixes structure the dictionary. They connect each other due to the corresponding identifier of the patterns of morphotactics representations. The morphologic dictionary that is incorporated in the program application of the parser includes 100 000 lexemes of the contemporary Georgian language. The functioning of the program application relies on morphologic generator, which, in its turn, depends on the morphologic data of each unit of the computational dictionary. Morphologic generator is used to generate a separate lexeme from lemma and to analyze a word-form. The morphology analyzer gives one or more probable derivation patterns. [4, 8, 9].

The morphological analyzer assigns features for every word. The feature set for Georgian includes part of speech, number, case, auxiliary form (nominal part of compound predicate), tense, person of subject and object.

The syntactic analyzer divides a sentence into the syntactic pairs and describes them. (Two members of a sentence that are connected in a certain way make a syntactic unit. This type of unit is called Syntactic Pair or Syntagma. A syntactic pair is the smallest syntactic unit of a sentence).

A parse tree is represented by binary mutual oriented connections between the words of a sentence. There is a dominant term and a dependent term in each connection.

In order to get a syntactic tree the following rules must be fulfilled:

1. Tree wholeness must be maintained. It is forbidden to divide a sentence into more than two trees.
2. Every word of the sentence must be involved into a tree structure and each of them must have a dominant term.

Syntactic Connection is considered as formal means to express syntactic relation between the members of the sentence. There are three types of the syntactic relation in Georgian: Agreement, Government and Adjoin [10].

Agreement is a syntactic connection when the dependent term gets the shape of the dominant term (did-i burt-i – ‘a big ball’ (nom. case), did-ma burt-ma – (erg. case)).

There are four cases of agreement connection in Georgian:

1. A noun agrees with the noun, adjective, numeral, pronoun and participle in case:
2. *maγali mta* – ‘a high mountain’ (Nom. Case), *maγalma mtam* – (Erg. Case);
3. A Noun agrees with the other Noun, Adjective, Numeral, Pronoun and Participle in Number as well as in Case. *mta-ni maγal-ni* – ‘high mountain’, *mta-ta maγal-ta* – ‘high mountains’;
4. A Noun agrees with the Verb in Person: *me c’avedi* – ‘I went’, *šen c’axvedi* – ‘you went’, *is c’avida* – ‘he/she went’;
5. A Noun agrees with the Verb in Number (it depends is it animated or not): xeebi dgas – ‘the trees is standing’ (instead of ‘trees are standing’).

Government is a syntactic connection when one term requires a certain form from the other, which it does not have.

There are the following cases of government connections:

1. Noun governs the other Noun in Case
2. Verb governs a Noun in Case
3. A Formless Word (Preposition or Adverb used as a Preposition) governs a Noun in Case
4. Verb governs a Noun with a Preposition.

Adjoin is a syntactic connection when the dominant term connects the dependent term by content without any grammatical expression. The dependent term is a word with unchangeable form (gvian dabrunda – ‘(he/she) returned late’).

Six basic types of relation are defined according to which part of speech the dominant term of the syntactic pair belongs to.

There are two kinds of Nominal relations: Substantive (of Noun) and Adjective (Adjective, Numeral and Pronoun). In substantive relation the dominant word is a noun and the dependent word may be a noun, an adjective, a numeral, a pronoun and a verbal noun, rarely – an adverb.

In order to write syntactic relations easily the following conventional signs are used: SC _ Syntactic Connection. On the right of the equation a dominant term and on the left a dependent term are placed. For example, SC = N/N means that a syntactic pair consists of two nouns, the first one is the dominant and the other is the dependent one.

The types of Nominal connections:

1. SC = N/N – *botli c'qali* – 'a bottle of water';
2. SC = N/Adj – *didi saxli* – 'a big house';
3. SC = N/Num – *mesame c'igni* – 'the third book';
4. SC = N/Pron – *ramdenime st'udent'i* – 'a few students';
5. SC = N/Adv – *gvian γamit* – 'late at night'.

In the types of verbal connections, the dominant term is a verb connecting a noun or an adverb. The first one expresses an objective connection and the second – Adverbial modifier. We consider the connection of Verbal-Noun words with the Verbal connection.

The types of Verbal connections:

1. SC = V/N – *daxat'a qvavili* – '(he/she) painted a flower';
2. SC = V/Adv – *daxat'a kargad* – '(he/she) painted well';
3. SC = V/Num – *xutni movidnen* – 'Five (of them) came'.
4. SC = V/VN (Verbal Noun) – *gak'eteba minda* – '(I) want to do'.

An Adjective appears as a dominant very rarely. Its dependent term may be a Noun or an Adverb.

1. SC = Adj/N – *lomivit ʒlieri* – 'strong as a lion';
2. SC = Adj/Adv – *ʒalian k'argi* – 'very good'.

A Numeral does not appear as a dominant term in a syntactic pairs very often either.

1. SC = Num/Num – *p'irveli xuti* – 'the first five';
2. SC = Num/Pron – *qoveli asi* – 'every hundred';
3. SC = Num/N – *siit meate* – 'the tenth on the list';
4. SC = Num/Adv – *bolodan meore* – 'the second from the end'.

Pronoun word connections:

1. SC = Pron/Pron – *qvelaferi es* – 'all these';
2. SC = Pron/Adj – *k'argi vinme* – 'someone good';
3. SC = Pron/Num – *erti ram* – 'one thing';
4. SC = Pron/Ger – *qoveli morbenali* – 'every runner'.

Adverbial word connections:

1. SC = Adv/N – *c'qlis dasalevad* – 'to drink water';
2. SC = Adv/Adv – *gvian γamit* – 'late at night'.

As a result, there are 23 syntactic pairs realized in Georgian, which differ only by part of speech variation. Let us consider syntactic pairs by their syntactic function in the sentence.

According to the theory of L. Tesnièr the hierarchy and the direction between the word connections is determined by the well-ordered actant structure of a Predicate [11]. This point of view is proved in the Georgian language. The Verb-Predicate has the ability to connect with the other members of the sentence. That is because of the originality of the Georgian verbs [12]. It contains grammatical information necessary for building a syntactic structure of the sentence. It is a sentence on its own. In the Georgian language, a predicate may be simple or compound. Compound predicate is considered as a syntactic pair. The following pairs differ according to variety of the nominal parts of the compound Predicate. According to what are the nominal parts of the compound predicate expressed with the following types are defined:

1. Nominal part of a compound predicate is a Noun (CP = V+N) – *bavšvebi arian* – 'They are children';
2. Nominal part of a compound predicate is an Adjective (CP = V+Adj) – *guli k'etili hkonda* (He/she had a kind heart);
3. Nominal part of a compound predicate is a Numeral (CP = V+Num) – *ormoci unda iqos* (there must be forty);
4. Nominal part of a compound predicate is a Pronoun (CP = V+Pron) – *šen xar* – 'you are';
5. Nominal part of a compound predicate is a Participle (CP = V+Prtcp) – *dač'rili unda qopiliqo* – 'He/she/it must have been wounded'.

The Predicate forms the largest number of the syntactic pairs with the other members of the sentence and the Subject forms the lesser number. The other members of the sentence form syntactic pairs as well.

Thus, 53 types of syntactic pairs have been defined according to these four criteria:

1. The dominant and the dependent term in the syntactic pair;
2. The syntactic relation type of the syntactic pair;
3. The part of speech of the dominant in the syntactic pair;
4. The syntactic role of the dominant in the sentence.

Let us consider an example (A dominant term is marked with bold font. The information about pairs (which parts of speech they are expressed with) is given in brackets and the syntactic function that has the given member in the sentence is ascribed on the left of the brackets):

k'ibis tavši boša kali kvesk'nelis diasaxlisivit egebeba momavlis šišit apor-iakebul tana-mokalakeebs – 'At the top of the stairs a gypsy woman welcomes her citizens, who are anxious with fear about the future, like an underworld hostess'.

1. *kali egebeba* – 'woman welcomes' – S(N) + P(V), Government in Case, Agreement in Person and Number;
2. *boša kali* – 'gypsy woman' – Att(N) + S(N), Agreement in Case;
3. *egebeba tanamokalakeebs* – 'welcomes citizens' – P(V) + IndObj (N), Government in Case;
4. *aporiakebul tanamokalakeebs* – 'anxious citizens' – Att(Participle) + IndObj (N), Agreement in Case;
5. *šišit aporiakebul* – 'anxious with fear' – IndObj (N) + Att(Prtcp), Government in Case;
6. *momavlis šišit* – 'fear about the future' – Att(N) + IndObj (N), Government in Case;
7. *diasaxlisivit egebeba* – 'welcomes like a hostess' – Adv(N) + P(V), Government in Case;

8. *kvesk'nelis diasaxlisivit* – 'underworld hostess' – Att(N) + Adv(N), Government in Case;
9. *tavši egebeba* – 'welcomes at the top' – Adv(N) + P(V), Government in Case;
10. *k'ibis tavši* – 'At the top of the stairs' – Att(N) + Adv(N), Government in Case.

5 Conclusion

At present, the full syntactic markup has been generated for 25310 sentences (95240 words). The given approach permits to include all information expressed by morphologic and syntactic means in contemporary Georgian. We expect that the new corpus will stimulate a broad range of further research and development projects.

The corpus will be publicly available on the page http://geocorpora.gtu.ge/#/texts.

The Rustaveli Foundation of Fundamental Research in Georgia has supported this work with a grant No. 31/65. The project title is "The Full (Morphological, Syntactic, and Semantic) Annotation System of the Georgian Language Corpora".

References

1. Lu, X.: Computational Methods for Corpus Annotation and Analysis. Springer Science + Business Media Dordrecht (2014)
2. Niladri Sekhar Dash: Language corpora: past, present and future. Mittal Publications, 2009 – Computational linguistics
3. Russian linguistic corpora: http://www.ruscorpora.ru
4. Beridze, M., Lortkipanidze, L., Nadaraia, D.: Dialect dictionaries in the Georgian dialect corpus. In: Aher, M., Hole, D., Jeřábek, E., Kupke, C. (eds.) TbiLLC 2013. LNCS, vol. 8984, pp. 82–96. Springer, Heidelberg (2015). doi:10.1007/978-3-662-46906-4_6
5. Amirezashvili, N., Chikoidze, G., Chutkerashvili, A., Lortkipanidze, L., Samsonadze, L., Javashvili, N.: The contribution of lexical functions in an explanatory-combinatorial dictionary. In: Proceedings of the Archil Eliashvili Institute of Control Systems of the Georgian Technical University 19, 2015
6. Hajicova, E., Panevova, J., Sgall, P.: Language resources need annotations to make them really reusable: the prague dependency treebank. In: Proceedings of the First International Conference on Language Resources & Evaluation, pp. 713–718 (1998)
7. Text Encoding for Interchange, TEI Guidelines: http://www.tei-c.org/Guidelines/
8. Lortkipanidze L.: Record and reproduction of morphological functions. In: Proceedings of the 5-th Tbilisi Symposium on Language, Logic and Computation. ILLC, University of Amsterdam CLLS, Tbilisi State University, pp. 105–111 (2003)
9. Lortkipanidze, L.: Interactive system for compilation of multilingual concordancers. In: Conference abstracts of the 6-th International Contrastive Linguistics Conference (ICLC6), Berlin (2010)
10. Kvatchadze, L.: Syntax of modern georgian language (in Georgian). Rubikoni Tbilisi (1996)
11. Gamkrelidze, T., Kiknadze, Z., Shaduri, I., Shengelaia, N.: A course in theoretical linguistics (in Georgian). Tbilisi University Press, Tbilisi (2003)
12. Chikoidze G.: Net representation of morphologic processors. archil eliashvili institute of control systems (in Russian). Tbilisi, Monograph (2004)

Frame Theory with First-Order Comparators: Modeling the Lexical Meaning of Punctual Verbs of Change with Frames

Sebastian Löbner(✉)

General Linguistics, Heinrich Heine Universität Düsseldorf,
Düsseldorf, Germany
loebner@phil.hhu.de

Abstract. The first part of the paper proposes a formal foundation of a theory of frames. Frames are embedded into a general ontology. On this background, we introduce a formal model-theoretic semantics for frames and thereby an interface to formal semantics. The model-theoretic semantics allows us to define central notions of frame theory such as the "satisfaction type" for a node in a given frame, and a semantic definition of subsumption. The second part presents a case study of decomposition: frames for subtypes of intransitive punctual verbs of change, such as punctual 'grow' and 'go from A to B'. We introduce times, events, and time-dependent attributes in the ontology. A crucial element of the analysis is "comparators", a novel type of attribute in frame theory. Comparators are partial two-place attributes that compare two individuals of the same sort and return comparison values such as '=' vs. '≠' or '<' vs. '=' vs. '>'. Comparators allow us to model within frames and AVMs conditions in terms of basic abstract relations. The approach proposed offers simplifications of alternative proposals for the frame-theoretical decomposition of these types of verb: (i) standard PL1 is used as a frame description language; (ii) with comparators, the use of non-functional relations as additional components in frames can be avoided; (iii) meanings of punctual verbs of change can be represented within one frame.

Keywords: Frames · Ontology · Comparator · Mereotopology · Time · Time-dependence · Degree achievement · Verb of change · Verb of locomotion · Decomposition

1 Background

The work presented here has emanated from joint projects involved with a novel theory of frames (*see* Acknowledgments). The research initiative aims at a formally precise and cognitively plausible theory of frames as a general format of conceptual representation, in particular, but by no means exclusively, as a general format of semantic representation, including lexical decomposition and compositional meaning. The enterprise aims to test the 'Frame Hypothesis' that goes back to Barsalou's work ([2, 3, 4]): that frames constitute the general format of representation in the human cognitive system.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 98–117, 2017.
DOI: 10.1007/978-3-662-54332-0_7

Löbner [9] discusses the implication of the Frame Hypothesis for cognition and linguistic theory in general, and for syntax and semantics in particular.

This paper adds various innovations, modifications, and extensions of earlier approaches such as Petersen [16] and Kallmeyer and Osswald [8]. These include the following major points:

- relation to a general frame ontology as a global model for admissible frames;
- inclusion of time and time-dependent attributes in the frame ontology, along with explicit time elements in frames;
- introduction of novel two-place "comparator" attributes that capture basic binary relations between entities of the same sort, e.g., equality or order relations.

Inclusion of time into frame representations addresses a basic challenge to frame theory: the representation of dynamic concepts such as the meanings of verbs of change (see Naumann [15] for introductory discussion). The apparatus developed will be applied in a decompositional representation of punctual change of state verbs.

2 Frames Related to a Global Frame Ontology

2.1 A Simple Frame Example

Frames in the sense of the frame hypothesis are recursive attribute value structures with exclusively functional relations; they will be given a formal definition in the next subsection. A simple example is the frame for a 'male person with blue eyes' given in Fig. 1. It is a frame for an entity typed as a person that is assigned values for two attributes: EYES and GENDER. The GENDER attribute is specified as 'male', which may be taken as standing for a type of gender, rather than just the individual gender ♂. The value of the attribute EYES is the eyes of the person (as one complex entity); these are specified with the attribute COLOR as blue. The value specification 'blue' is not a single discrete color value but stands for a range of color hues.[1] Figure 1 displays two common formats of representing frames: the left part of the figure shows a frame diagram with labeled nodes and arrows; the right figure is an equivalent attribute-value matrix (AVM).

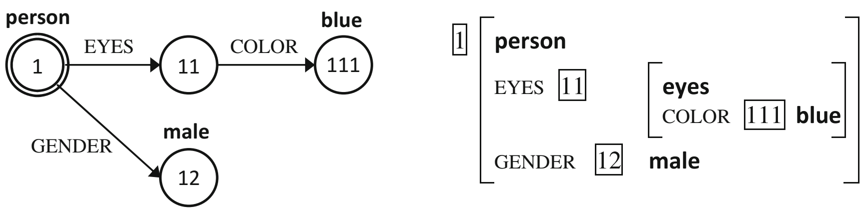


Fig. 1. Frame example in diagram and AVM representation

[1] For the sake of simplicity, the fact is ignored that it is not the whole eyes that are specified for color, but rather the iris. The complex aspects of color predication are not at issue here.

The frame diagram displays four circles, or nodes; these contain numbers as labels; the number label is essentially a variable for the individual represented by the node. The node labeled[2] with [1] represents the entity which the frame is taken to describe; it is called the "central node" and is marked by a double line border. The arrows represent attributes as indicated by their labels; the arrows end in the nodes that represent the values of the attributes. Three nodes carry type information: **person**, **blue**, and **male** indicating that [1] is a person, [111] is a blue color, [12] is a male gender.

The attribute value matrix in the right part of Fig. 1 contains the same information as the diagram. It represents the element [1] as the entity described by the whole structure; [1] is typed as a person. That person has two attributes, EYES and GENDER. The attribute EYES takes the value [11], further typed by the embedded matrix: it is **eyes** (a tautological categorization deriving from the fact that [11] is the value of the EYES attribute), and it has the attribute COLOR with value [111], categorized as **blue**. The GENDER attribute of [1] has the value [12], of type **male**.

It should be pointed out that in natural language, terms for attributes are systematically also used for their values. We talk of 'the color [attribute term] of the cocktail' but also of 'the colors [value term] red, blue, and pink'. This circumstance has caused some confusion. In this article we will distinguish the two senses of attribute terms typographically, using SMALL CAPS for their use as attribute terms, and **special bold type** in their use as value, i.e. type, terms.[3]

Frames are essentially graphs in the mathematical sense, i.e. sets of pairs of things, called "nodes". The pairing is the result of relations obtaining between the paired elements. In the case of frames, the pairs are ordered, rendering a directed graph ("digraph" in mathematical jargon), and the relations are attributes. The first member of a pair is the bearer (argument) of the attribute, the second member of the pair is the value. In the use of graphs for frame representation, different attributes may connect nodes. Crucially, the attributes are functions: one node cannot be connected by the same attribute to two nodes representing different values.

The term "graph" strongly appeals to network diagram representations like the left structure in Fig. 1; the connections between nodes are called "edges" or "arcs". Frame graphs have topological properties that also appeal to the network diagrams: for example, we will assume that frame graphs in general are connected. Still, in spite of their name, graphs are not drawings, but abstract mathematical structures. The diagrams are a means of *visualization*. The nodes carry identifying labels, or indices. The edges, too, carry labels that are read to define the attribute which an edge stands for. The nodes may carry predicative type information in addition to being indexed.

For later application, the general definition of a frame structure needs to admit n-place attributes ($n \geq 1$). N-place attributes result in a functional connection between an n-tuple $\vec{x}$ of argument nodes and an additional node for the value. We will represent such constellations as can be seen in Fig. 2. Edges that connect tuples of nodes are technically called "hyperedges" in graph theory.

[2] The frame around the number is omitted when the label is written into a frame diagram node.

[3] For discussion of the relation and the difference between attribute concepts and concepts for their values see Petersen [16]: 162ff and Löbner [10]: 30–34.

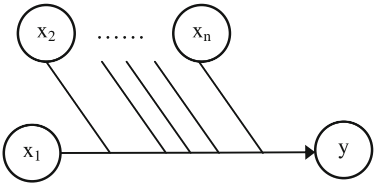


Fig. 2. Frame diagram for an n-place attribute assignment

Definition 1. Frame structure
A frame structure is a sextuple ⟨V, r, A, att, T, typ⟩, such that

a. V is a finite set of nodes.
b. r ∊ V is a distinguished node, the "central node".
c. A is a finite set of edge labels.
d. att is a function that maps pairs of an n-tuple of nodes ($n \geq 1$) and an edge label on another node.
e. T is a set of type labels.
f. typ is a partial function that assigns type labels to nodes.
g. With $E =_{df} \{\langle \vec{x}, y\rangle \mid \exists a \in A\ \mathsf{att}(\vec{x}, a) = y\}$, ⟨V, E⟩ is a connected digraph with n-to-1 hyperedges ($n \geq 1$).

A digraph is connected if there is a connection along one or more edges, in either direction, between any two nodes of the graph. The frame represented by the diagram in Fig. 1 is connected. The central node is understood to represent the type of object, or the individual object, that the frame describes. In the example, the central node is the node [1]; its status is indicated by the double-line border.

The above definition of frame structure is essentially equivalent to the definitions given in Petersen [16], except for the fact that the definition here is more general while Petersen distinguishes different, more specific types of frame, such as frames for sortal, relational, and functional nouns (see Löbner [10]: 41–47). The frame structures used below for verb meanings/event concepts are relational concepts: they have a referential central node representing the event described and argument nodes for the verb's role arguments. The definition here is also compatible with the definition of frames in Kallmeyer and Osswald [8]; their definition, however, allows arbitrary relations between the nodes of a frame structure, in addition to the attributes.

2.2 Global Frame Ontologies

The two structures in Fig. 1 are both essentially two-dimensional *expressions*. Usually, these types of structure are used with a presupposed interpretation of the attribute and type labels. When applied in semantics, these labels are in need of a precise formal interpretation. We will define an underlying ontology for frame interpretation and relate all frames to it. In an additional step we will translate frame structures into first-order predicate logic in order to provide an interface to truth-conditional semantics.

Any framework of representation employing frames depends on ontological assumptions as to which attributes can plausibly figure in a frame and which types are available for the values they can take. The frame representations used here are primarily representations in terms of attributes of the entities represented and, secondarily, in terms of the types of attribute values. Therefore the frame ontology is based on attributes. This distinguishes the approach from frameworks like Carpenter's [5] theory of typed feature structures: these are based on a given system of types with a semi-lattice structure defined by type subsumption. The "types" in the theory proposed here correspond to Carpenter style "types" in that they subsume cases of the same description (like, for example, **blue-eyed man**). However, the types as introduced here are essentially derivative. In general, attributes are partial functions; they come with a domain of application, i.e. a type of things that are eligible for carrying this attribute, and they come with a codomain, the type of possible values. Being partial functions, they need not return a value for everything in their domain. The attribute COLOR, for example, is defined for physical bodies and it takes colors as values (including the value 'colorless').

The attributes of the human frame ontology form an infinite space. Attributes can build chains of arbitrary length by applying an attribute to the value of the first attribute, another one to the value of the second, and so on. Also, arbitrary attributes can be defined ad hoc, drawing on existing concepts. Nevertheless, frame theory needs a consistent and plausible framework of available, well-defined attributes that constitute legitimate elements of frame representations.

In a really cognitive approach, the attributes and types would have to be attribute *concepts* and type *concepts*. This kind of approach is not spelt out yet. The frame ontology to be introduced here will be an "ontological" ontology, of functions and entities in the world as we refer to when using language. This decision enables a straightforward comparison with analyses in the truth-conditional semantic paradigm, which is based on an "ontological" ontology, too.

A frame ontology has a non-empty universe of individuals, a set U. U is partitioned into sorts: any individual is of exactly one sort; there is no overlap of sorts. In a hierarchy of types, the sorts are maximal types. To give a few examples: one will assume the sorts of persons, of physical objects, of numbers, of temperatures, of weights, of colors, of truth-values, and so on.

It is assumed that attributes are always restricted to one sort, but their domain need not exhaust the sort it is a subset of. Also, attributes return values of only one sort. This appears necessary for the ontological distinction of different attributes denoted by a polysemous attribute term. For example, the term *weight* can be used as denoting the weight attribute of physical objects having mass, but it can also be used for a particular aspect, roughly importance, of things like arguments or decisions. A frame ontology properly defined will provide different attributes for the 'weight' of physical objects, of arguments, and of decisions, respectively, because these three kinds of entity are of different ontological sorts as are the values returned by the 'weight' attribute.

Attributes may have more than one argument; for example, DISTANCE and RELATIONSHIP would be two-place attributes.

The attributes form a space which will be postulated to be closed under functional composition. For example, if HAIR is assumed to be an attribute of persons, and the attribute COLOR is applicable to human hair, functional composition yields the attribute HAIR COLOR for persons with hair. We will further postulate that for injective one-place attributes, the inverse is also in the ontology. Injective attributes are 1-to-1 mappings. For example, every person has a body, whence there will be the attribute BODY available for persons; conversely, every body belongs to exactly one person, and there will be the inverse attribute BODY^{-1} in the ontology; it returns the body-owner for every body.

Every attribute A is associated with the types that constitute its domain of application and the range of values it can take. We denote the domain as dom(A) and the codomain as cod(A). If A is n-place (n > 1), its domain is the Cartesian product of n types $\text{dom}_i(\text{A})$, i = 1,…, n. We will assume that the set of types is closed under intersection and that for each individual in U there is a corresponding atomic type, an atom. If an attribute is applicable to all members of a type **t**, then the image of **t** is also a type in the ontology; conversely, if **t** is a subtype of the codomain of an attribute, then its preimage is a type in the ontology. For example, one will assume that there is a type **hair** included in the domain of the attribute COLOR; thus the ontology will contain the type **hair color** as the image of **hair** under the COLOR mapping. Conversely, there is the type **red** of red colors within the codomain of the COLOR attribute; its preimage is the type of red objects; it intersects with the type **hair** to form the type **red hair**. Due to the conditions on the ontology, the system of types is closed under attribute-related operations. It does not, however, exhaust the powerset of U – not if there is more than one sort in the ontology, which we will certainly want to assume.

According to the definition to follow, the sorts are an a priori part of the ontology, while the types (except of the sorts) should be considered derivative of the system of sorts and attributes. Equally a priori is the universe as such, a non-empty set whose elements are the individuals of the system and correspond to the atomic types. The types form hierarchical systems, but the hierarchies are each restricted to one sort.

Definition 2. Sorted frame ontology
A sorted frame ontology $\mathfrak{O}$ is a quadruple $\langle \text{U}, \mathcal{S}, \mathcal{A}, \mathcal{T} \rangle$ such that

a. U, the universe, is a non-empty set of individuals.
b. $\mathcal{S}$, the system of sorts, is a partition of U: every individual in U belongs to exactly one sort in $\mathcal{S}$.
c. $\mathcal{A}$, the set of attributes, is a set of non-empty partial functions A: $\text{U}^n \longrightarrow \text{U}$. The attributes are restricted to sorts: for every A: $\text{U}^n \longrightarrow \text{U}$, there are sorts $\mathbf{s}_1,\ldots, \mathbf{s}_n$, $s \in \mathcal{S}$ such that $\text{dom}_i(\text{A}) \subseteq \mathbf{s}_i$ for i = 1,…, n and $\text{cod}(\text{A}) \subseteq \mathbf{s}$.
d. $\mathcal{T}$, the set of types, is a proper subset of $\wp(\text{U})$: every type **t** is a subset of U. For every $t \in \mathcal{T}$, there is an $s \in \mathcal{S}$ with $\mathbf{t} \subseteq \mathbf{s}$: types contain individuals of only one sort.

Closure conditions on the set $\mathcal{A}$ of attributes

e. $\mathcal{A}$ is closed under functional composition.
f. If a one-place attribute $\text{A} \in \mathcal{A}$ is injective, there is a partial function $\text{A}^{-1} \in \mathcal{A}$, such that for every $x, y \in \text{U}, \text{A}^{-1}(y) = x$ iff $\text{A}(x) = y$.

Closure conditions on the set $\mathcal{T}$ of types

g. Every sort is a type: $\mathcal{S} \subseteq \mathcal{T}$.
h. For every $x \in \text{U}$, $\{x\} \in \mathcal{T}$. $\{\{x\}: x \in \text{U}\}$ is the set of atomic types in $\mathcal{T}$.

i. If $\text{A} \in \mathcal{A}$, $\mathbf{t} \in \mathfrak{T}$, $\mathbf{t} \subseteq \text{dom(A)}$, then the image of $\mathbf{t}$ under A, $\text{A}[\mathbf{t}]$, is in $\mathfrak{T}$;
if $\text{A} \in \mathcal{A}$, $\mathbf{t} \in \mathfrak{T}$, $\mathbf{t} \subseteq \text{cod(A)}$, then the preimage of $\mathbf{t}$ under A, $\text{A}^{-1}[\mathbf{t}]$ is in $\mathfrak{T}$.
j. For every $\mathbf{t}, \mathbf{t}' \in \mathfrak{T}$, $\mathbf{t} \cap \mathbf{t}' \in \mathfrak{T}$.
k. $\mathfrak{T}$ contains no other types than those defined by (h)–(k).

While this definition is completely abstract, we want to invest it with the informal constraint that the attributes to be postulated in any frame ontology used for the analysis of natural language are to be cognitively plausible, i.e. plausible candidates for attributes of which humans can be expected to be able to have cognitive representations in their minds.[4] We now introduce the notion of a frame structure related to a given ontology.

Definition 3. Frame structure related to an ontology
For a frame structure $\langle$V, r, A, att, T, typ$\rangle$ related to the ontology $\mathfrak{O} = \langle \text{U}, \mathcal{S}, \mathcal{A}, \mathfrak{T} \rangle$
a. the elements of V are variables for individuals in U;
b. the elements of A are labels for attributes in $\mathcal{A}$;
c. the elements of T are labels for types in $\mathfrak{T}$.

2.3 A Formal Semantics for Frames

A frame structure related to an ontology receives a straightforward semantics by using identical constants as labels in the frame structure and as expressions in the metalanguage of the ontology. A frame structure describes a structure in the ontology with as many elements as there are nodes in the frame structure indexed with a variable. For example, the frame structure in Fig. 1 describes triples $\langle$x, y, z$\rangle$ such that x is a person of male gender, with eyes y of color z of a blue color. Hence it describes the type of blue-eyed male persons in the ontology. As a descriptor of a type, the frame *constitutes a concept*. Note that, unlike a predicate expression in a logic language under truth-conditional interpretation, a frame is not merely associated with a type to fit its truth conditions; rather it describes the type it represents by use of particular criteria (i.e. attributes and value assignments) which are irreplaceable components of the concept.

In order to provide an interface to truth-conditional semantics, we translate frame structures into an appropriate first-order predicate logic (PL1) language. The type of PL1 language needed for frame representation is considerably restricted: it has neither negation, nor disjunction, nor universal quantification.

Definition 4. PL1 frame language associated with an ontology
For a given frame ontology $\mathfrak{O} = \langle \text{U}, \mathcal{S}, \mathcal{A}, \mathfrak{T} \rangle$, the associated language PL-$\mathfrak{O}$ is a first-order predicate logic language with the following elements:

[4] We will make use of an implicitly presupposed ontology when we discuss examples below. To come up with a concrete definition of a frame ontology for semantic analysis is a task for semantic and ontological theory for decades of research.

a. individual terms, including individual variables and individual constants for individuals in U,
b. type constants: terms for types in $\mathfrak{T}$; type constants include expressions of the form '{i}' (i an arbitrary individual constant) for the atomic types in $\mathfrak{T}$,
c. n-place function constants: terms for the attributes in $\mathcal{A}$,
d. $\in$ for statements of the form $\ulcorner i \in \mathbf{t} \urcorner$, with individual term i and type term **t**,
e. = for statements of the form $\ulcorner i_1 = i_2 \urcorner$ with individual terms i_1, i_2,
f. $\wedge$ propositional conjunction,
g. $\exists$ existential quantifier.

The model-theoretic interpretation of PL-$\mathfrak{O}$ is obvious: all constants are interpreted as denoting the individuals, attributes, and types they denote in the metalanguage of the ontology. Complex expressions are interpreted as usual. In the PL-$\mathfrak{O}$ applied here, we use framed symbols for natural numbers as individual variables. There is a straightforward way of translating a frame structure into a PL-$\mathfrak{O}$ representation, rendering a canonical satisfaction formula (SatFor) that is unique except for the order of conjuncts.

Definition 5. Canonical satisfaction formula
If $\mathfrak{f} = \langle V, r, A, T, \mathsf{att}, \mathsf{typ} \rangle$ is a frame based on the ontology $\mathfrak{O} = \langle U, \mathcal{S}, \mathcal{A}, \mathfrak{T} \rangle$, the canonical satisfaction formula for $\mathfrak{f}$ – SatFor($\mathfrak{f}$) – is the conjunction of the following PL-$\mathfrak{O}$ statements:

a. for all $i_1, \ldots, i_n, j \in V$, all $\textsc{a} \in A$,	$\ulcorner \textsc{a}(i_1, \ldots, i_n) = j \urcorner$	if	$\mathsf{att}(i_1, \ldots, i_n, \textsc{a}) = j$,
b. for all $i \in V$, $\mathbf{t} \in T$	$\ulcorner i \in \mathbf{t} \urcorner$	if	$\mathsf{typ}(i) = \mathbf{t}$
c. for all $i \in V$, $u \in U$	$\ulcorner i = u \urcorner$	if	$\mathsf{typ}(i) = \{u\}$

d. If att and typ are empty and $V = \{r\}$, then SatFor($\mathfrak{f}$) is $\ulcorner r = r \urcorner$.

Example. The frame structure in Fig. 1 yields the SatFor in (1), which can be simplified by variable elimination to the equivalent in (2):

(1) $\boxed{1} \in \textbf{person} \wedge \textsc{eyes}(\boxed{1}) = \boxed{11} \wedge \textsc{color}(\boxed{11}) = \boxed{111} \wedge \boxed{111} \in \textbf{blue} \wedge$
$\textsc{gender}(\boxed{1}) = \boxed{12} \wedge \boxed{12} \in \textbf{male}$

(2) $\boxed{1} \in \textbf{person} \wedge \textsc{color}(\textsc{eyes}(\boxed{1})) \in \textbf{blue} \wedge$
$\textsc{gender}(\boxed{1}) \in \textbf{male}$

Related to an appropriate ontology, a frame structure represents a type in the ontology for every variable/node it contains. The type is defined by existential closure applied to the remaining variables. For the three variables of the example frames, the types are:

(3) a. $\{x : \exists y \exists z (x \in \textbf{person} \wedge \textsc{eyes}(x) = y \wedge \textsc{color}(y) = z \wedge z \in \textbf{blue} \wedge$
$\textsc{gender}(x) \in \textbf{male})\}$
i.e. the type of male persons with blue eyes

b. $\{y : \exists x \exists z (x \in \textbf{person} \wedge \textsc{eyes}(x) = y \wedge \textsc{color}(y) = z \wedge z \in \textbf{blue} \wedge$
$\textsc{gender}(x) \in \textbf{male})\}$
i.e. the type of blue eyes of male persons

c. $\{z : \exists x \exists y (x \in \textbf{person} \wedge \textsc{eyes}(x) = y \wedge \textsc{color}(y) = z \wedge z \in \textbf{blue} \wedge$
$\textsc{gender}(x) \in \textbf{male})\}$
i.e. the type of blue colors of the eyes of male persons.

The resulting type may be empty, depending on facts given in the ontology. The "satisfaction type" of a frame is the type it describes with respect to the central node; in general, a satisfaction type can be defined for every variable/node in the frame:

Definition 6. Satisfaction type
Let $\mathfrak{f} = \langle \mathrm{V}, \mathrm{r}, \mathrm{A}, \mathrm{T}, \mathsf{att}, \mathsf{typ} \rangle$ be a frame based on the ontology $\mathfrak{O} = \langle \mathrm{U}, \mathcal{S}, \mathcal{A}, \mathcal{T} \rangle$.
The satisfaction type of $\mathfrak{f}$ is the type
$\mathrm{SatType}(\mathfrak{f}) =_{\mathrm{df}} \{\mathrm{r} : \exists_{\neq \mathrm{r}}\ \mathrm{SatFor}(\mathfrak{f})\}$
where $\exists_{\neq \mathrm{r}}\ \mathrm{SatFor}(\mathfrak{f})$ is the existential closure of $\mathrm{SatFor}(\mathfrak{f})$ for all variables in V except r.
More generally, if $\mathrm{i} \in \mathrm{V}$ is a node in $\mathfrak{f}$, the satisfaction type of $\mathfrak{f}$ for i is:
$\mathrm{SatType}(\mathfrak{f}, \mathrm{i}) =_{\mathrm{df}} \{i : \exists_{\neq \mathrm{i}}\ \mathrm{SatFor}(\mathfrak{f})\}$

The notions defined so far allow for a semantic definition of frame subsumption.

Definition 7. Subsumption
Let $\mathfrak{f}_1$ and $\mathfrak{f}_2$ be frames related to the same ontology, let x and y be nodes in $\mathfrak{f}_1$ and $\mathfrak{f}_2$, respectively. $\mathfrak{f}_1$ for node x subsumes $\mathfrak{f}_2$ for node y if the satisfaction type for the first includes the satisfaction type for the second:
$\mathfrak{f}_1(\mathrm{x}) \sqsubseteq \mathfrak{f}_2(\mathrm{y})\quad \mathrm{iff}_{\mathrm{df}}\quad \mathrm{SatTyp}(\mathfrak{f}_1, \mathrm{x}) \supseteq \mathrm{SatTyp}(\mathfrak{f}_2, \mathrm{y}).$
In particular, $\mathfrak{f}_1$ subsumes $\mathfrak{f}_2$:
$\mathfrak{f}_1 \sqsubseteq \mathfrak{f}_2\quad \mathrm{iff}_{\mathrm{df}}\quad \mathrm{SatTyp}(\mathfrak{f}_1) \supseteq \mathrm{SatTyp}(\mathfrak{f}_2).$

The usual definition of subsumption for frames and AVMs relates to the structure of a frame: a frame subsumes another frame if there is a node in the second frame for every node in the first frame with compatible respective characteristics (cf., e.g., the definition in Kallmeyer and Osswald [8]: 280 ff.). The formal semantics defined for frames here allows a semantic definition of subsumption that is independent of the form in which the information in a frame is arranged. For example, we might replace the frame structure in Fig. 1 by the ones in Fig. 3. The three frame structures mutually subsume each other. The left frame contracts the attribute chain COLOR(EYES(…)) to the functional composition EYE COLOR; the respective parts of the SatFor's are logically equivalent. In the right frame, the attribute EYES is replaced by its inverse EYES-OWNER. Both replacements are legitimate due to the closure conditions on $\mathcal{A}$ in the general definition of ontologies. Thus defined, subsumption is basically logical entailment for frame structures or AVMs taken as logical expressions.

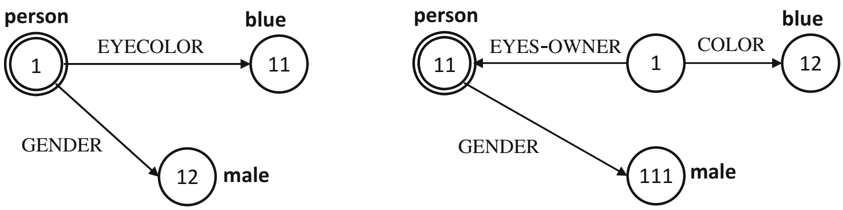


Fig. 3. Alternative frame structures for 'blue-eyed male person'

3 Comparators

One point of debate in frame theory is the question of how to model relations in a framework that is restricted to graphs with exclusively functional connections. This paper offers a proposal for modeling certain basic intrasortal relations. Kallmeyer and Osswald [8] proposed to model such relations by admitting non-functional relations as an additional type of component in a frame. This step is in conflict with the Frame Hypothesis mentioned in the beginning since Barsalou frames are supposed to employ exclusively functional attributes as relations in frames. We will introduce special two-place attributes to capture this type of intrasortal relation; we call them "comparators".

Comparators are binary attributes that apply to arguments of the same sort. Their values are the outcome of comparison with respect to certain basic criteria. For example, a comparator might compare two real numbers and return one of the comparison values '=', '>' and '<'. The three values are mutually exclusive alternatives, whence the comparator is a functional relation. (Note that there is no comparator function which could in addition also return '$\leq$' and '$\geq$'; there are however comparators that return either '>' or '$\leq$', or either '<' or '$\geq$'.) The comparison values are individuals returned as values by the comparison function; they are not to be confused with the general relations usually denoted with the same symbols. Comparator values are ontologically of their own sort, depending on the domain of comparison.

Comparators are cognitively highly plausible conceptual operations. Even the most primitive organisms are capable of comparisons. For humans, comparators are involved in recognition and categorization to name only a few cognitive functions.[5]

Depending on the sort applied to, there may be more than one comparator definable. The basic equality comparator is defined for every sort. If there are partial or linear orders defined within a sort, there are corresponding comparators; partial orderings include part-whole relations.

Definition 8. First-order comparators[6]
First-order comparators in an ontology $\mathfrak{O} = \langle \mathrm{U}, \mathcal{S}, \mathcal{A}, \mathcal{T} \rangle$ are two-place attributes with both arguments of the same sort; they return comparison values.
For every sort $\mathsf{s} \in \mathcal{S}$, the **standard comparator** $©_\mathsf{s}$ is defined as
$©_\mathsf{s}(\mathrm{x}, \mathrm{y}) =_{\mathrm{df}}$ '$\neq$' / '=' iff x$\neq$y / x=y.

When we use comparators, we will express the comparisons in the satisfaction formulae in the conventional way; we will write 'x = y' instead of '$©_\mathsf{s}$(x, y) = =', and so on.

[5] Barsalou ([4]: 601) uses comparators (although not called so) in his model of truth and falsity for propositions. Comparators check whether an internal simulation [roughly, a frame] can be "map[ped] successfully into a perceived scene". Truth and falsity is envisaged in our approach as a further application of comparators, probably of second order and hence outside the scope of frame theory sketched here.

[6] These comparators are first-order in that they compare individuals.

4 Time and Tensed Ontologies

We will now integrate time into the ontology, a step that allows the modeling with frames of time dependence by using explicit time representations. In their frame analysis of temporal situation structure, in Kallmeyer and Osswald [8], Naumann [15], and Gamerschlag, Geuder, and Petersen [7], the authors relate to time implicitly by employing attributes such as RESULT; their frames do not contain times explicitly. The approach taken here follows a strategy of maximizing explicitness in frame representation; in particular, it tries to represent all arguments of attributes in the frame structure. In addition to this aspect of expressivity, the incorporation of explicit time parameters in verb meanings is also probably necessary for modeling tense and aspect in a frame theory of composition. I take it, along with many theories of tense, that tense is a predication about times, more specifically about the time occupied by an event e, if e is the event referred to, and aspect is perfective.[7]

Independently, the question arises if the assumption that times rather than just events figure in human cognitive representations is psychologically realistic. According to experimental evidence (Roberts, Coughlin, and Roberts [17]) there is positive evidence of time representation in cognition even for pigeons. This justifies the assumption of the sort **time** in our ontology.

We extend the definition of the ontology by introducing the sort **time**, where times are understood as being intervals on the time axis. For the comparison of times we apply the system of temporal relations introduced in Allen [1] which are mutually exclusive; we will use only three of them, the relation 'm' (meet) and its inverse 'mi', and the "earlier" relation '<'. Two times x and y "meet" iff x is earlier than y and immediately adjacent to y; mi obtains between x and y iff y m x. Rather than the traditional $\mathbb{R}^1$ topology assumed e.g. by Dowty [6] for the time axis, it appears psychologically more adequate to apply a topology where two time intervals can be connected without overlapping or with one being open and the other one closed. In such a mereotopology, it would be possible to model two consequent days as two closed time intervals that meet without overlap, the first one ending with 12:00 p.m. and the next one beginning with 0:00 a.m.; 0:00 a.m. would be the point in time immediately following 12.00 p.m. Other units of time like years, months, weeks, hours, and so on would be modeled analogously. This appears to me more in accordance with intuition than assuming that either 0:00 a.m. or 12:00 p.m. does not belong to a full day. A topology that allows for connection without overlap is even more plausible for the cognition of physical space where we are obviously willing to assume that it is possible that, say, two boxes with plane surfaces can be piled upon each other with nothing in between and no parts shared. Mereotopologies that implement the notion of connection without overlap are introduced in Varzi [18] or Muller [14] for systems of spatio-temporal reasoning. Allen's temporal relations can be defined for such a mereotopology if we take 'adjacent' as 'connected without overlap'.

Introducing time into the ontology involves two major changes and additions: (i) certain attributes will be defined as being time-dependent, i.e. two-place attributes

[7] See Löbner [11], Chap. 6, for a basic outline of the theory of tense and aspect supposed here.

with an additional time argument; (ii) events will be related to times by means of certain event attributes. As to the first point, it is to be observed that according to common ontological understanding the values of some attributes are time-dependent, while the values of other attributes are not. In general, attributes relating things to their origin are not time-dependent; most property attributes, however, are, since properties can change during the lifespan of things.

Definition 9. Tensed ontology
A tensed ontology $\mathfrak{O}^t = \langle \mathrm{U}, \mathcal{S}, \mathcal{A}, \mathcal{T} \rangle$ fulfills the following conditions in addition to conditions a.–k. in Definition 2:

l. There is a sort **time** in $\mathcal{S}$, of non-empty time intervals (including points in time) with the properties and relations as to be defined in an appropriate mereotopology of **time**.
m. There is a comparator $©_{\mathbf{time}}$ in $\mathcal{A}$, defined for the sort **time** that assigns values for the Allen relations to pairs of times.
n. There is a sort **event** in $\mathcal{S}$, of events.
o. Three attributes map events on time:
TE(e) = the time occupied by the event e,
TB(e) = the time before the event e,
TA(e) = the time after e;
for every event e: TB(e) < TE(e) < TA(e); the three times need not be adjacent.
p. Time-dependent attributes: There are two-place attributes A: $\mathbf{t}_1 \times \mathbf{time} \longrightarrow \mathbf{t}_2$, for some types $\mathbf{t}_1$, $\mathbf{t}_2$.
q. Homogeneity condition for attributes with a time argument:
If an attribute A assigns the value v to a time t and possibly further arguments, then A returns the same value v for all non-empty subintervals of t.

The homogeneity condition is a novel kind of constraint. It is also necessary for other attributes, e.g. COLOR: COLOR does not yield a unique value for a would-be argument in case it is of more than one color like, for example, most flags. In general, an attribute A underlies a homogeneity condition with respect to an argument x iff a value assignment to x by A is true iff the same assignment holds for all relevant proper parts of x. This is tantamount to the condition that the predication λx A(x) = y (for any y $\in$ cod(A)) is summative in the sense of Löbner ([12]: 237).

5 Lexical Frames for Intransitive Punctual Verbs of Change

We will now present proposals for representing the lexical meaning of certain subtypes of punctual verbs of change. Verbs of change have been a topic in decompositional analysis since Dowty [6]. Modeling change is a particular challenge to frame theory since the original notion of frame is static. The proposal differs from existent alternatives in various ways.

- Unlike the proposals in Kallmeyer and Osswald [8], Gamerschlag et al. [7], and Naumann [15], it employs explicit time reference.
- Unlike in Gamerschlag et al. [7], and Naumann [15], the meanings of verbs of change are represented in a single frame.
- Unlike in Kallmeyer and Osswald [8], dynamic verb meanings are modeled with comparators rather than by adding non-functional relations to the frames used; also the logical language used for frame description here is more conventional.

The frame model to be proposed represents punctual verbs of change: the event consists of a change in time relating to the time before and the time after the event e. There is a condition that holds at the time before e and does not hold at the time after, or vice versa. The frame imposes no conditions on the time TE(e) that the event itself occupies. The transition may be continuous or an instantaneous change; it may be temporally extended or not. It is in this sense that these verbs are punctual: the time TE(e), for all the verb concept tells us, might as well be just a point in time. We need to assume that the criterial times TB(e) and TA(e) are immediately adjacent to TE(e). Otherwise, there might be more than one change of the kind between TB(e) and TA(e). We also need to assume that the times TB(e) and TA(e) are homogeneous with respect to the criterial condition in order to prevent there being further changes *within* TB(e) or TA(e); this constraint will be captured by the general homogeneity condition q in Definition 9 as the criterial condition will be in terms of the values of a time-dependent attribute. Note that TB(e) and TA(e) can be just points in time; the homogeneity condition does not require that there be extended time intervals of no change before or after the event. The question whether or not TE(e) itself should be imposed a condition to the extent that e must not host further changes back and forth will be left open here. I take it that we ought to allow for this; for example, the sentence *the light went on* may be about a neon light that goes on and off and on again several times until it is permanently on; to give a different example, we may say *the price of the share rose today* after a day of the price constantly changing up and down.

5.1 A General Frame for Punctual Change

These conditions still leave the proper determination of TB(e) and TA(e) to the individual context of interpretation. Even so, I assume that 'the time before the event' and 'the time after the event' are pragmatically admissible determinate notions, i.e. legitimate time-dependent attributes in the underlying ontology. TB(e) and TA(e) are determined by the conditions that they "meet" TE(e) and are homogeneous with respect to the criterial condition. If, and how far, TB(e) extends into the past and TA(e) into the future does not make any difference.

The constellation of the event and its attributes TB, TE, and TA is depicted in Fig. 4. The pale arrow labeled 'time' is not part of the frame; it is only added for illustrating the temporal relationships. Comparators define the Allen relation from TA(e) to TB(e) as 'm' and the Allen relation from TB(e) to TE(e) as 'mi'.

From now on, we will write the atomic values of comparator attributes right into the nodes of the frame diagrams, as is done in Fig. 4 with the value nodes for the two

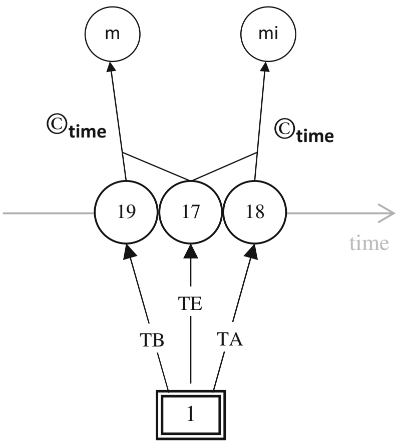


Fig. 4. Frame for an event and the related times

temporal comparison attributes. Thus, the node inscription ‘cv’ replaces the type annotation ‘{cv}’ and overwrites the variable labeling the node. We will render the respective conjunct in the SatFor as ‘©..(i, j) = cv’ rather than as ‘©..(i, j) = k ∧ k ∈ {cv}’. We will simplify ‘©..(i, j) = cv’ further to ‘i cv j’, using the symbols for the comparator values as relation symbols between individual terms in the associated PL1 language.

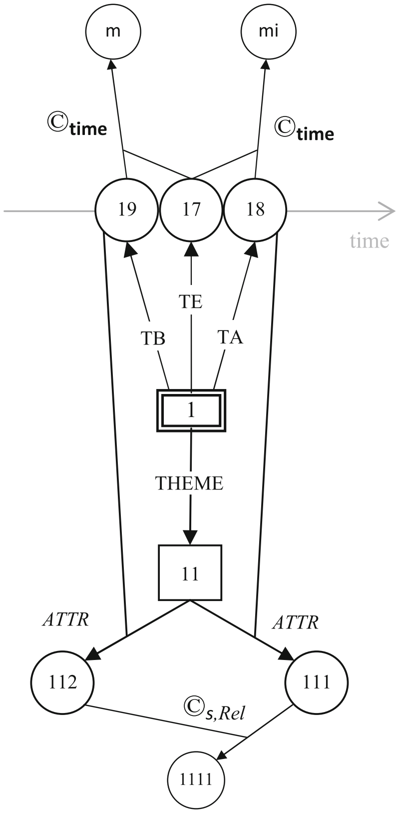
For the subtypes of verbs of change considered here, the criterial condition of change is in terms of a time-dependent attribute of the theme argument that takes on different values for TB and TA. We will represent the theme argument by a rectangular node in the frame diagram, thereby indicating that it represents an open argument. Argument nodes can be considered providing an interface to syntax, but they do not receive a different interpretation than the other nodes in the frame.

Figure 5 (next page) displays the general schema of a frame for a verb that denotes a punctual change of its theme argument in terms of the attribute *ATTR*;[8] the difference between the state before and after the change is captured by the value of some comparator $©_{s,Rel}$ that is defined on the sort **s** of entities which the attribute *ATTR* takes as values. The value of the comparator has to be such that it entails inequality, as does e.g. ‘<’. The change is punctual in the sense that it does not impose any conditions on the time TE itself.

Discussion. The frame in Fig. 5 models a variant of Dowty’s BECOME operator in its second version related to time intervals (cf. the discussion in [6], pp. 139ff.). BECOME operates on a proposition φ; [BECOME φ] is true at an interval I if there are intervals J and K such that ([6], p. 141, (11’)):

---

[8] Note that $\mathfrak{F}_{\mathbf{PunCh}}$ is not a frame structure in accordance with Definition 1 as it contains the attribute *variables* ‘ATTR’ and ‘$©_{s,Rel}$’; $\mathfrak{F}_{\mathbf{PunCh}}$ is a frame schema.



$$
\begin{aligned}
& \mathrm{TE}(\boxed{1}) = \boxed{17} \\
\wedge\ & \mathrm{TB}(\boxed{1}) = \boxed{19} \\
\wedge\ & \boxed{19}\ \mathsf{m}\ \boxed{17} \\
\wedge\ & \mathrm{TA}(\boxed{1}) = \boxed{18} \\
\wedge\ & \boxed{18}\ \mathsf{mi}\ \boxed{17} \\
\wedge\ & \mathrm{THEME}(\boxed{1}) = \boxed{11} \\
\wedge\ & \mathit{ATTR}(\boxed{11}, \boxed{18}) = \boxed{111} \\
\wedge\ & \mathit{ATTR}(\boxed{11}, \boxed{19}) = \boxed{112} \\
\wedge\ & ©_{s,Rel}(\boxed{111}, \boxed{112}) = \boxed{1111}
\end{aligned}
$$

Fig. 5. Frame schema $\mathfrak{F}_{\mathbf{PunCh}}$ for a punctual change of state of the theme in the attribute *ATTR*, canonical satisfaction formula

(i) $\neg\varphi$ is true in J and J contains the initial bound of I,
(ii) φ is true in K and K contains the final bound of I,
(iii) I is the smallest time interval that fulfills the conditions (i) and (ii).

The frame schema $\mathfrak{F}_{\mathbf{PunCh}}$ represents a BECOME event where φ is defined in terms of the value of the attribute *ATTR* of the theme changing into what it becomes. The three times $\boxed{19}$ (time before), $\boxed{17}$ (time the event occupies), and $\boxed{18}$ (time after), correspond to Dowty's time intervals J, I, and K, respectively. As is common practice in many variants of Montague Grammar, Dowty uses a logical language without expressions that refer to times; time-dependence is spelt out in the model-theoretic interpretation. It is therefore not possible to express in the formal frame language used here what would be Dowty's φ for the frame in Fig. 5. If we omitted the time argument of the attribute of the theme, we would get '*ATTR*($\boxed{11}$) = $\boxed{111}$' as the operand φ of Dowty's BECOME.

There are differences, though, between the analysis proposed here and Dowty's. One concerns the minimality condition (iii). Dowty's definition of BECOME aims at exactly delimiting the interval in which the change takes place. In view of examples like those mentioned above (the neon light and the share price examples), I consider this condition possibly too strong. The second difference concerns the comparison between the two states; Dowty's BECOME operator models a change from $\neg\varphi$ to φ, while the comparator model allows for a wider range of relationships between the state before and the state after. The third difference is the respective topology of time assumed.

5.2 Punctual Degree Achievements

The general frame $\mathfrak{F}_{\mathbf{PunCh}}$ can be spelt out for different types of intransitive punctual verbs of change. One such type is punctual degree achievements[9] with a lexically specified scale of change. Examples are cases like punctual intransitive 'grow':

(4) *the number of participants in my seminar grew by 2*

The lexical frame for 'grow' has SIZE as the attribute *ATTR* of change. The respective comparator is $©_{\mathbf{size},<}$ where **size** is the sort of sizes, i.e. the sort of things that can be values of the attribute SIZE; this sort carries a linear ordering $<$; in the case of *grow*, the comparator returns '>'.

Punctual verbs cannot be used in the progressive[10] as they do not denote an event that can be parted into subevents of equal kind. In addition to the punctual use of *grow* illustrated in (4), there are senses of English *grow* where the verb denotes a continuous change on the size scale. These are not captured with the frame in Fig. 5. Modeling continuous change would require TE(e) to be temporally extended and would call for imposing a monotonicity condition on TE(e) to the extent that the value of SIZE is monotonically increasing during TE(e). Other punctual verbs denoting degree achievements on specific scales would be represented with a different instantiation of *ATTR*: *rise* with HEIGHT, *widen* with WIDTH, and so on.

In Japanese, there are degree achievement verbs which in general defy progressive use, i.e. a progressive reading with the continuative-*te iru* form.[11] These include for example *hutoru* 'get fat(ter)' and *yaseru* 'get thin(ner)'.

In general, punctual degree achievements are characterized by relating to a theme attribute *ATTR* that takes values of a sort that is ordered and therefore has a comparator that returns corresponding values such as '>' or '<'.

5.3 Punctual Verbs of Change into a Specific State

Consider *go on* (of lamps etc.) in the sense of changing into the change of being [switched] on. The criterial attribute (to be named properly) can take either of two values, 'on' and 'off'. The value of that attribute at TA(e), i.e. $\boxed{18}$, is specified as 'on', the comparator is the standard © for this sort and returns '$\neq$'. Japanese punctual verbs

9 See Dowty [6]: 88ff for the notion of degree achievement verb.

10 The notion 'progressive' is to be taken in the functional, semantic sense, as relating to a certain variant of imperfective aspect, not in the morphological sense.

11 See Martin [13], p. 518 for the punctuality of the Japanese verbs mentioned here and below.

of change into a specific state include *aku* ‘come open’, *kowareru* ‘break’ (intransitive), or *sinu* ‘die’. In general, this group of verb can be modeled with the frame in Fig. 5 with a specification of ATTR and its value $\boxed{111}$, and the comparator value being ‘$\neq$’.

5.4 Punctual Verbs of Locomotion

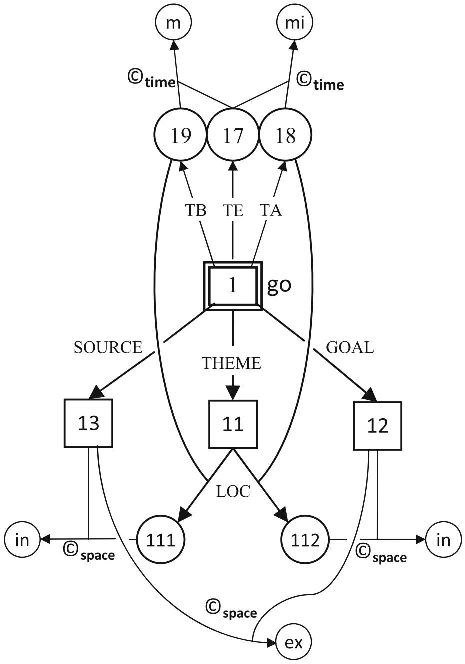
Verbs of locomotion from one location to the other can be represented with a frame very similar to $\mathfrak{F}_{\mathbf{PunCh}}$ – provided the verbs are conceived as punctual. It is not clear, if English has such verbs, but Japanese does, e.g. *iku* ‘go’, *kuru* ‘come’, *kaeru* ‘return’, *otiru* ‘fall’, *deru* ‘emerge from’, *hairu* ‘enter’, and others. The attribute of change is the location of the subject referent. For the construction ‘x *ga* [NOM] B *ni* [LOC] *ik-*’[12], we assume that both x, the theme, and B, the goal, are arguments of the verb. The goal attribute will usually be specified as a region that covers more than the space taken in by the theme when it is there. We cannot, therefore, identify the referent of the goal specification with the location of the theme that results from the motion; rather, we have to model the resulting state as the location of the theme being *within* the specified goal region. For the spatial relation, we can again use a comparator, $©_{\mathbf{space}}$, based on an appropriate mereotopology. We will use two comparator values, ‘in’ for x being within y and ‘ex’ for x being outside of y. These two are admissible values of the same comparator because the two spatial relations exclude each other.

We will model the Japanese punctual verb *iku* ‘go’ in the full construction ‘x *ga* [NOM] y *kara* [SOURCE] z *ni* [GOAL] *ik-*’, meaning ‘x go from y to z’ in a punctual sense. The verb meaning is modeled as ‘LOCATION(THEME(e)) be within SOURCE(e) at TB(e) and within GOAL(e) at TA(e), where SOURCE(e) is outside of GOAL(e)’. Thus the frame involves three applications of the spatial comparator; they encode (i) the relation between the location of the theme and the source, (ii) the relation between the location of the theme and the goal, and (iii) the relation between source and goal. The latter is necessary because this type of verb is not applicable to situations where the source is inside the goal region or overlaps with it: A statement like *John went from Tokyo to Japan* violates a presupposition of ‘go from SOURCE to GOAL’. This presupposition is modeled explicitly by the third comparator condition. Figure 6 displays the frame and the corresponding canonical satisfaction formula.

The frame can be accommodated to simpler cases such as the punctual Japanese *hair-u* ‘enter’. This verb has only two arguments, theme and goal. There is a condition on the location of the theme at TA(e): it is within the goal, and a second condition on the location of the theme at TB(e): it is outside of the goal. Switching these two conditions yields a frame for the verb *de-ru* ‘leave, emerge from’.

The canonical satisfaction formulae can be considerably simplified; if one omits the general conditions on TB, TE, and TA, and the general condition on the spatial relation between SOURCE and GOAL, the remaining conjuncts can be reduced to two, if one makes use of variable elimination. For the sake of comparability with other approaches, we

[12] *ik-* is the bare stem of the verb, not inflected for tense. The citation form of Japanese verbs carries the present tense ending *-u* or *-ru*.



$$
\begin{aligned}
& \text{TE}(\boxed{1}) = \boxed{17} \\
\wedge\ & \text{TB}(\boxed{1}) = \boxed{19} \\
\wedge\ & \boxed{19}\ \mathsf{m}\ \boxed{17} \\
\wedge\ & \text{TA}(\boxed{1}) = \boxed{18} \\
\wedge\ & \boxed{18}\ \mathsf{mi}\ \boxed{17} \\
\wedge\ & \text{THEME}(\boxed{1}) = \boxed{11} \\
\wedge\ & \text{GOAL}(\boxed{1}) = \boxed{12} \\
\wedge\ & \text{SOURCE}(\boxed{1}) = \boxed{13} \\
\wedge\ & \boxed{13}\ \mathsf{ex}\ \boxed{12} \\
\wedge\ & \text{LOC}(\boxed{11}, \boxed{19}) = \boxed{111} \\
\wedge\ & \boxed{111}\ \mathsf{in}\ \boxed{13} \\
\wedge\ & \text{LOC}(\boxed{11}, \boxed{18}) = \boxed{112} \\
\wedge\ & \boxed{112}\ \mathsf{in}\ \boxed{12}
\end{aligned}
$$

Fig. 6. Frame for punctual 'THEME go from SOURCE to GOAL', canonical satisfaction formula

replace the variable '$\boxed{1}$' by the conventional event variable 'e'. The remaining two conjuncts capture the idiosyncratic meaning components of punctual *go*/*iku*.

(5) punctual '*go from* SOURCE *to* GOAL', Japanese 'SOURCE *kara* GOAL *ni ik-'*
LOC(THEME(e), TB(e)) in SOURCE(e) ∧
LOC(THEME(e), TA(e)) in GOAL(e)

The respective meaning components of *hairu* and *deru* are given in (6) and (7):

(6) punctual '*enter* GOAL', Japanese 'GOAL *ni hair-*'
LOC(THEME(e), TB(e)) ex GOAL(e) ∧
LOC(THEME(e), TA(e)) in GOAL(e)

(7) punctual '*leave* SOURCE', Japanese 'SOURCE *o de-*'
LOC(THEME(e), TB(e)) in SOURCE(e) ∧
LOC(THEME(e), TA(e)) ex SOURCE(e)

6 To Be Continued

The proposal developed here introduces further modules of the 'Düsseldorf frame theory'. Major points are (i) the definition of a global frame ontology, (ii) the introduction of times, events, and time-dependent attributes in the ontology, a step paving the way for explicit representation of time, time-dependence, and time parameters of events in frames, and (iii) the introduction of two-place comparator attributes.

The latter enable the modeling of basic intrasortal relations within a framework with exclusively functional frame-internal relations. Comparators are waiting for further applications, e.g. to mereological relationships, or to the modeling of degrees.

The inclusion of time-dependent attributes raises the general logical and ontological question as to which attributes are time-dependent and which are not. In particular the question arises if, for principal reasons, dynamic verbs involve at least one time-dependent attribute of at least one verb argument.

Time-dependent attributes require a homogeneity constraint on the assignment of values. This raises another general question: which attributes in general underlie a homogeneity constraint?

In view of the range of possible application of these general theoretical considerations, the proposed lexical analysis of verb meanings is, of course, very selective. Further extensions would have to address non-punctual verbs of change, as well as all the other well-known aspectual classes.

Acknowledgments. The theory presented here arose from joint work at Düsseldorf University in the Research Unit RU 600 'Functional Concepts and Frames' (2005–2011), continued in the Collaborative Research Centre CRC 991 'The Structure of Representations in Language, Cognition, and Science' (2011–2019 or later), both financed by the German Science Foundation; see http://www.sfb991.uni-duesseldorf.de/en/ for the CRC. The author profited tremendously from extended discussion with Ralf Naumann, Wiebke Petersen, Henk Zeevat, and Timm Lichte. Reference to Roberts et al. [17] is owed to Tobias Kalenscher; Christopher Habel referred the author to the work on mereotopology mentioned above.

References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **26**, 832–843 (1983)
2. Barsalou, L.W.: Cognitive Psychology: An Overview for Cognitive Scientists. Lawrence Erlbaum Associate, Hillsdale (1992)
3. Barsalou, L.W.: Frames, concepts, and conceptual fields. In: Lehrer, A., Kittay, E.F. (eds.) Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization, pp. 21–74. Lawrence Erlbaum Associate, Hillsdale (1992)
4. Barsalou, L.W.: Perceptual symbol systems. Behav. Brain Sci. **22**, 577–660 (1999)
5. Carpenter, B.: The Logic of Typed Feature Structures. Cambridge University Press, Cambridge (1992)
6. Dowty, D.R.: Word Meaning and Montague Grammar. Reidel, Dordrecht (1979)
7. Gamerschlag, T., Geuder, W., Petersen, W.: Glück auf, der Steiger kommt: a frame account of extensional and intensional *steigen*. In: Gerland, D., Horn, C., Latrouite, A., Ortmann, A. (eds.) Meaning and Grammar of Nouns and Verbs, Studies in Language and Cognition, vol. 1, pp. 115–144. Düsseldorf University Press, Düsseldorf (2014). http://dup.oa.hhu.de/517/
8. Kallmeyer, L., Osswald, R.: Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. J. Lang. Model. **1**, 267–330 (2013)

9. Löbner, S.: Evidence for frames from human language. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Frames and Concept Types, Studies in Linguistics and Philosophy, vol. 94, pp. 23–67. Springer, Heidelberg (2014)
10. Löbner, S.: Functional concepts and frames. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Meaning, Frames, and Conceptual Representation. Studies in Language and Cognition, vol. 2, pp. 13–42. Düsseldorf University Press, Düsseldorf (2015). http://dup.oa.hhu.de/1/
11. Löbner, S.: Understanding Semantics, 2nd edn. Routledge, London and New York (2013)
12. Löbner, S.: Polarity in natural language: predication, quantification, and negation in particular and characterizing sentences. Linguist. Philos. **23**, 213–308 (2000)
13. Martin, S.E.: A Reference Grammar of Japanese. Yale University Press, New Haven and London (1975)
14. Muller, P.: Topological spatio-temporal reasoning and representation. Comput. Intell. **18**, 420–450 (2002)
15. Naumann, R.: An outline of a dynamic theory of frames. In: Bezhanishvili, G., Löbner, S., Marra, V., Richter, F. (eds.) TbiLLC 2011. LNCS, vol. 7758, pp. 115–137. Springer, Heidelberg (2013). doi:10.1007/978-3-642-36976-6_9
16. Petersen, W.: Representation of concepts as frames. In: Skilters, J. et al. (eds.) Complex Cognition and Qualitative Science. The Baltic International Yearbook of Cognition, Logic and Communication, vol. 2, pp. 151–170. University of Latvia, Riga (2007)
17. Roberts, W.A., Coughlin, R., Roberts, S.: Pigeons flexibly time or count on cue. Psychol. Sci. **11**, 218–222 (2000)
18. Varzi, A.C.: Parts, wholes, and part-whole relations: the prospects of mereotopology. Data Knowl. Eng. **20**, 259–286 (1996)

Semantic Predictions in Natural Language Processing, Default Reasoning and Belief Revision

Ralf Naumann and Wiebke Petersen(✉)

Institut für Sprache und Information, Universität Düsseldorf, Düsseldorf, Germany
petersen@phil.hhu.de

Abstract. Formal semantic theories are designed to explain how it is possible to produce and understand an infinite number of sentences on the basis of a finite lexicon and a finite number of composition rules. According to this architecture, language comprehension completely proceeds in a bottom-up fashion only driven by linear linguistic input thereby leaving no room for a predictive component which allows to make expectations about upcoming words. This is in stark contrast to neurophysiological research in the past decades on online semantic processing which has provided ample evidence that prediction plays indeed an indispensable role in language comprehension (the brain as a *prediction machine*, [Ber10]). In this article, we present an extension of formal semantic theory that allows to make predictions of upcoming words. The basic intuition is: predictions are based on incomplete information. Drawing (defeasible) conclusions based on such information can be modeled by default reasoning. Since predictions can go wrong, a second strategy for retracting wrong guesses is needed in order to integrate (unexpected) words into the prior context. This is modeled by belief revision. We model both processing stages, making predictions about upcoming words and integrating them into the prior context, and relate the models to the empirical findings in neurophysiological research.

Keywords: Default logic · Modal logic · Cognitive semantics · System Z · N400 · Late positivity

1 The Brain as a Prediction Machine

In formal semantic theories meaning is taken to be a relation between language and the external world (or reality). This relation is defined inside a logical theory, e.g. some form of type logic, using notions like 'reference', 'satisfaction' and 'truth'. On this view the main goal of natural language semantics is a definition

We thank our anonymous reviewers who helped to improve this paper by providing an elaborated and constructive feedback to former versions. The research was supported by the German Science Foundation (DFG) funding the Collaborative Research Center 991.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 118–145, 2017.
DOI: 10.1007/978-3-662-54332-0_8

of the truth for sentences in a natural language. This goal is achieved by giving a recursive and compositional analysis of the well-formed expressions of a language. Based on a finite lexicon and a finite set of composition rules, it then becomes possible to both produce and parse an infinite set of sentences none of which needs to be stored in the brain. This characterization is still valid for dynamic approaches like DRT or DPL in which the notion of truth is replaced by that of a relation between (information) states.

From a psycholinguistic or neurophysiological point of view the concept of meaning endorsed in formal semantics is quite unsatisfactory since it completely leaves out the question of how language is processed in the brain. A prime example that has emerged during the last three decades both in behavioral and electro-physical research are predictions or expectations of upcoming words in a given context.[1] Consider the example in (1) taken from [FK99, 469].

(1) Getting himself and his car to work on the neighboring island was time consuming. Every morning he drove for a few minutes and then boarded the ….

When asked, most people end the second sentence with the word ‘ferry’. This behavior is remarkably robust across individuals and it is empirically defined in terms of a word’s cloze probability[2] in a given (sentential) context. For example, in (1), ‘ferry’ has highest cloze probability (CP) and is therefore the *best completion* (BestComp). Since none of the individual words in (1) is strongly semantically related to ‘ferry’, it seems most likely that the context preceding ‘. . . ’ together with world knowledge is used during language processing to pre-activate semantic properties which best apply to (the concept expressed by) ‘ferry’ but not to the same degree to other vehicles like gondolas or airplanes. On this interpretation, both world knowledge and context play a crucial role in setting up semantic properties on the basis of which an expectation (or prediction) for an upcoming word is formed.

According to Baggio and Hagoort, examples like (1) show that formal semantics ‘is by design insensitive to differences between words of the same syntactic category denoting objects of the same type’, [BH11, 1343]. Their own example is (2).

(2) Last Friday the cruiser Arberia entered the *port/hippodrome* of Trieste.

They argue that the difference between the two continuations after ‘entered the’ must be semantic in nature because pragmatic deviance like the violation of a

[1] ‘Prediction’ must not be understood as a conscious or strategic process. Rather, prediction is understood as the unconscious activation of semantic properties of upcoming words prior to their occurrence, [FK99, 487].

[2] Cloze probability: participants in an offline norming task are presented sentence frames like that in (1) and are asked to fill in the dots with the first word that comes to their mind. The proportion, ranging from 0 to 1, of respondents supplying a particular word is defined as the cloze probability of this word in that context.

Gricean conversational maxim does not occur (if one assumes that both sentences are false at speech time). In addition, the difference has nothing to do with the way the world looks like.

However, note that Baggio and Hagoort's argument is based on the implicit assumption that the problem arises only at the level of integration/composition. After 'port' or 'hippodrome' have been semantically recognized, they have to be integrated or combined with (the semantic representation of) the previous context. For 'port', being a best completion, this should pose no problems whereas for 'hippodrome' this integration should be much more difficult, if not impossible, given that this word is not only semantically unrelated but almost semantic anomalous to the semantic properties of the context. Though integration and prediction are closely related, the problem of how predictions and/or expectations can be represented in formal semantic theories cannot be reduced to simply incorporating it into the integration/composition mechanism.

If prediction (or expectation) is understood in the sense that it is based on the pre-activation of semantic features of words which are not yet presented to the comprehension system, the problem of combining or of integrating that word with the current semantic representation does arise only at a second stage. In a first stage, the semantic features of the expected upcoming, not yet presented, word are activated simultaneously (or in parallel) with the semantic features of words that have already been recognized and combined with the prior context.[3] Thus, there must be a separate mechanism which makes it possible to deduce semantic features (σ) from information that is already part of the semantic representation (τ) of the prior context and world knowledge (τ') stored in Long Term Memory (LTM). Then, using τ and τ', σ is deduced. Prediction is closely related to integration. Since predictions are risky – they can go wrong – there needs to be an additional (or subsequent) mechanism that deals with wrong guesses by explaining how they can be retracted. Exactly at this point prediction becomes related to integration/composition. Predicted semantic features are used to build up a semantic representation of the upcoming word, which eventually is integrated with the prior context. If a prediction turns out to be wrong because a non-expected word is encountered, integration is successful only if the wrong guesses are first retracted because otherwise combining the predicted with the actual encountered features results in an unsatisfiable semantic representation. Since predicted and actual features are combined, semantic anomalies like 'hippodrome' in (2) is a limiting case of the prediction-integration mechanism.

The above considerations lead to the following questions. At the empirical level one gets: (i) what neurophysiological evidence is there to support a distinction between prediction and integration/composition? (ii) given that there is a distinction between prediction and integration/composition, what type of

[3] Note that the pre-activated features used to predict upcoming words cannot simply be part of information about arguments, say, of verbs or common nouns. For example, 'board' in isolation does not prime (semantic features of) 'ferry' as opposed to (semantic features of) other semantically possible arguments like 'gondola' or 'airplane'.

information is predicted (atomic vs. decompositional in terms of semantic features)?, and (iii) predictions can be wrong; is there any empirical evidence for a stage in online semantic processing at which wrong predictions are retracted? If yes, how is this stage related to integration? These questions will be the topic of the next section where we will review electrophysiological experiments involving event-related brain potentials, in particular the N400 and two kinds of late positivity.

When implementing a predictive mechanisms in a formal semantic theory, the two principle questions are (i) in what exactly does this mechanism consist?, and (ii) where in the overall architecture of such a theory is it to be located? These questions will be the topic of the second part of this paper (see Sect. 3). In the last part of the paper (Sect. 4), we introduce wide-spread alternative approaches to interpret results on the N400 and P600 components (N400 as an index of semantic integration and P600 as an index of syntactic processing) and briefly discuss their shortcomings and possible implications for our theory.

2 Semantic Processing Online: Evidence from ERPs

For semantic processing, an important event-related potential (ERP) component[4] is the N400. It is a broad, negative-going deflection that starts around 200–300 ms after a word has been presented, either auditory or visually, and peaks around 400 ms after stimulus onset. In neuroscience there is an ongoing dispute of whether the N400 reflects semantic prediction and lexical retrieval or semantic integration operations [BFH12]. In the following, we focus on the former view; Sect. 4 critically discusses the latter approach. Thus, our approach builds on the hypothesis that the N400 is an index that allows one to examine the impact and the extent long-term memory (LTM) have on on-line semantic sentence processing. Its amplitude for a word in a given context is modulated (though not monotonic to) the word's off-line cloze probability. It was first observed in case of semantic anomalies like 'I like my coffee with cream and socks'. However, each word in a sentence elicits an N400. Furthermore, it does not even require a sentential context as shown by semantic priming tasks which involve the presentation of a semantically related or unrelated word before a target word: coffee – tea vs. chair – tea. Here 'tea' yields a larger N400 when followed after 'chair'. Note that, the N400 is *not* sensitive to negation. E.g., both 'A carrot is a fruit' and 'A carrot is not a fruit' generate more N400 activity than 'A carrot is a vegetable'.

[4] An event-related potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event. An ERP component is a portion of an ERP waveform that has a characteristic shape, timing and amplitude distribution across the scalp and a well-characterized pattern of sensitivity to experimental manipulations or neural source, [KF11,LPP08]. It is important to note that the common statement that a word does not elicit an ERP component (which will be used in this paper as well) is a simplification. It is meant that it does not trigger a brain response that significantly differs from the baseline response triggered by some control word.

2.1 Fine-Grained Expectations: Semantic Features are Pre-activated

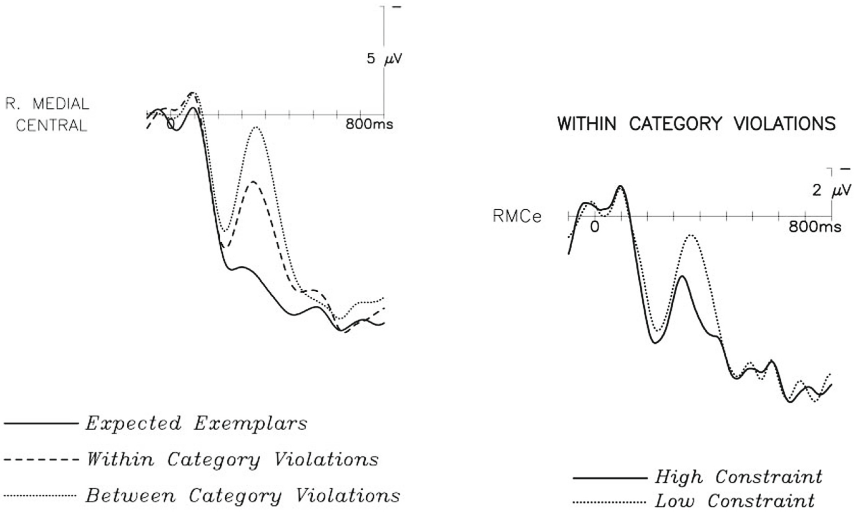
In their seminal paper [FK99], the authors investigated the following three questions w.r.t. predictions using N400 effects: (i) what type of information is predicted in a given context?, (ii) what influence do different kinds of constraining contexts have on those predictions?, and (iii) what influence do semantic relations between different target words have on the predictions? The experimental design consisted of pairs of sentences which were read by participants for comprehension. The first sentence established an expectation for a particular exemplar of a semantic category, syntactically realized by a common noun, while the second ended either (a) with this best exemplar, (b) an unexpected exemplar from the same (expected) category or (c) an unexpected exemplar from another category. An example is given in (3).

(3) They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of *palms/pines/tulips*.

Two of the three words belonged to the same taxonomic category. For example, both 'palm' and 'pine' are subtypes of the category 'tree'. The third member, 'tulip' in (3), did not belong to that category but, importantly, there was a (common) category to which all three words (or the concepts expressed by them) belong: plant. Unexpected exemplars from the same category are *within-category-violations* (WCV) whereas unexpected exemplars from another category are *between-category-violations* (BCV). Completions were ranked according to their offline cloze probability (CP, cf. footnote 3). Best completions ('palm') have highest CP. Both WCVs and BCVs had the same low CP in a given context. Additionally, sentential contexts were divided into two groups: strongly constraining and weakly constraining contexts. This distinction was defined by a median split on the CP of the best completions. For strongly constraining contexts best completions had an average value of 0.896 and in weakly constraining contexts of 0.588. WCVs and BCVs always had a CP < 0.05 across both sentential constraints. (3) is an example of a strongly constraining whereas (4) is a weakly constraining context.

(4) The gardener really impressed his wife on Valentine's day. To surprise her, he had secretly grown some *roses/tulips/palms*.

The following results were found. Overall, the N400 amplitude was significantly larger (i) for BCVs than for WCVs and (ii) for WCVs compared to best completions, i.e. one got BestComp < WCV < BCV (see Fig. 1 for details). Strongly constraining contexts are associated with overall slightly higher, i.e. more positive, amplitudes than weakly constraining contexts, [FK99, 481]. However, there was a difference w.r.t. the factor 'constraint' for WCVs. Such violations elicited a less enhanced N400 amplitude in strongly constraining compared to weakly constraining contexts (cf. Fig. 1). For both BestComp and BCVs, by contrast, there were no significant differences between the two kinds of contexts.



(a) taken from [FK99, p. 481] (b) taken from [FK99, p. 483]

Fig. 1. (a) Comparison of N400s for BestComp (expected exemplar), WCV (within category violation), and BCV (between category violation). (b) Comparison of N400s for WCVs in strongly constraining (high constraint) and weakly constraining contexts (low constraint).

The consequences which these results have for an account of online semantic processing are the following (for details, see [FK99]). First, the information provided by the context must be rather specific. This follows from the difference in N400 amplitude between BestComp and WCVs. If only general taxonomic, say category level, information were available, members of the same category, say 'palm' and 'pine', should elicit similar brain responses. Second, the N400 is sensitive to category violations. Words that are unexpected but belong to the same category as the best completion are processed differently from unexpected ones belonging to a different category, though both words have the same (low) CP. Second, predictions/expectations come in degree and depend on the strength of the context.

According to [FK99, 489], these results constitute evidence for the view that what gets pre-activated and what is stored in LTM are semantic features of concepts expressed by words and not (discrete) atoms like 'ferry' or 'palm'. The features that get activated are those associated with the best completion(s), i.e. those words having the highest CP in the given context, plus possibly features that can be inferred using world knowledge. For example, in (3) the context together with world knowledge pre-activates such features as 'tropical', 'resort', 'adornment', 'tree', and 'evergreen' since 'palm' is the best completion having the highest CP. Since three of those features equally apply to 'pine', its N400 amplitude though larger than that for 'palm' is smaller than that for 'tulip', for which only one feature applies. For a strongly constraining context, the number of pre-activated features is greater and therefore more constraining than the number of such features in a corresponding weakly constraining context. The more features of an upcoming word get pre-activated, the higher the probability is that even for unexpected but semantically related words (that belong to the same category) there is sufficient overlap with those features so that lexical access is facilitated. Hence, since in a strongly constraining context the number of pre-activated semantic features is greater than in a corresponding weakly constraining context, WCV should elicit a lower N400 amplitude in strongly constraining than in weakly constraining contexts, as borne out by the empirical data. Furthermore, since the overlap between pre-activated and actual semantic features is equally low for BCVs, the amplitude of the N400 should be the same for strongly constraining and weakly constraining contexts, again in line with the empirical data. Consequently, predictions/expectations should be graded and these degrees should be reflected in the corresponding amplitudes of the N400. But this is exactly what happens: BestComp $<$ WCV $<$ BCV across sentence constraint. In sum, if in a particular context a part of the semantic features representing a word A in the brain, say 'palm', is (pre-)activated, the comprehension system is better prepared to access and semantically process another word B, say 'pine', whose set of semantic features has a greater overlap to that of A than a word C, say 'tulip', for which this overlap is smaller.

2.2 The Risk of Pre-activation: Wrong Guesses

Predictions are risky because they can turn out to be wrong. E.g., if in the context of (3) 'palm' is predicted but 'pine' is eventually found, some expectations are wrong and must be deleted or retracted. Thus, there should be a stage in online semantic processing during which wrong guesses are undone. One candidate for such an operation is semantic integration. There are at least two kinds of evidence for drawing a distinction between a prediction stage in which possibly wrong features of the upcoming word are predicted and an integration stage of the semantic of the actual encountered word in which wrong features are deleted and new ones are added. First, if the N400 would be related not only to the prediction stage, but to the stage of semantic integration as well, words with the same meaning should elicit identical or very similar N400 effects, However, this is not the case as shown by the following empirical result. [DUK05] used sentence pairs like those in (5) where the sentence frame ended either with 'a' or 'an'. Since these two articles have exactly the same meaning, they should elicit the same N400 effects.

(5) The day was breezy so the boy went out to fly a/an ...

[DUK05] found a larger N400 amplitude for 'an' compared to 'a'. Since both articles have the same meaning, there should be no difference in brain response when it comes to integrating it with the semantic representation of the previous context because the semantic relation to this context must be exactly the same for both words. By contrast, if one assumes that the context preceding the article establishes a particular prediction for the most expected word 'kite', this difference can easily be explained. Since 'kite' begins with a consonant, 'a' is expected and not 'an'.

Second, there is post-N400 brain activity which is related to the semantic distinctions on which the N400 is based: late positivities. [FWODK07] considered pairs of sentences like those in (6).

(6) a. The children went outside to *play/look*. (strongly constraining context)
b. Joy was too frightened to *move/look*. (weakly constraining context)

In both kinds of context the unexpected ending, 'look' for example, had the same (low) CP.[5] In addition, the unexpected ending was not semantically related to the best completion and were considered plausible in an off-line norming task. Thus, any difference w.r.t. N400 effects could be attributed to the constraint of the sentence context. The results of the experiment showed that the N400 effects were graded by CP. The N400 amplitude was smallest for the best completion in the strongly constraining context; it was intermediate for the best completion in the weakly constraining context and highest for the unexpected completion for both kinds of constraint. However, the unexpected ending differed w.r.t. another ERP-component: a late frontal positivity between 500 and 900 ms over frontal electrode sites emerged for unexpected words in strongly constraining but not in weakly constraining contexts. The authors comment [FWODK07]: 'This processing stage thus seems to be sensitive to the greater degree of mismatch between the rich information provided by a strongly constraining sentence and an unrelated (though plausible) unexpected word, leading to the possibility of surprise and/or increased resource demands entailed by the need to override or suppress a strong prediction for a different word or concept.' This result was reproduced by [DQK14] using sentences like that in (7).

(7) For the snowman's eyes the kids used two pieces of coal. For his nose they used a *carrot/banana/groan* from the fridge.

According to [DQK14], the contexts in (7) were strongly constraining since the mean CP of the best completion, 'carrot', was 73.9%. Besides a best completion there was a semantically related and plausible continuation, 'banana', and a semantically unrelated and implausible (or impossible) continuation, 'groan' in (7). The CP for both kinds of continuation was equally low: $<0.01\%$. In addition to the late frontal positivity the authors found an increased parietal post-N400 positivity (PNP) for unexpected and semantically implausible words. Importantly, this positivity was not exhibited by unexpected but plausible words like 'banana' in (7). Similarly, the late frontal positivity was only found for plausible but not for implausible (impossible) continuations.

Since both kinds of late positivity are not graded (in contrast to N400 effects) and apply only to one particular type of unexpected continuations, they can neither be taken to simply reflect some process of plausibility evaluation nor be interpreted as a 'mismatch' detector.

[5] Cloze probabilities: 'play': 91%; 'move': 31% and 'look': 3% in both contexts.

When taken together, the results in this section provide evidence for the following picture of online semantic processing.[6] Semantic processing in the brain unfolds over several stages, [FK99,DQK14,BFH12]. The first stage is indexed by the N400 and has to do with lexical access. Semantic features of an upcoming word are activated in parallel with features of words that are currently being processed in order to access that word in LTM. The more features are already activated, the easier it is to retrieve that item from LTM. At the neural level this correlation is reflected by the amplitude of the N400: the greater the overlap with pre-activated features, and, therefore, the less features have to be additionally activated, the lower the amplitude. At this stage the item is not (yet) integrated with the semantic representation of the context. When it comes to integration, indexed by the two late positivities, what is at stake are no longer those features that are common to both the pre-activated and the actually encountered set but those feature which do not apply to the semantic representation of the word encountered. Two principle cases have to be distinguished: the target word is either of a type to which the best completion belongs or not. In the first case those features that have been pre-activated but which do not apply to the semantic representation of the word encountered have to be retracted. By contrast, in the second case, e.g. 'groan' in (7), a different strategy must be chosen because the semantic representation of the target word is incompatible with the semantic constraints imposed by the context.

Thus, we have arrived at the following three constraints on a formal semantic theory: (a) there must be a mechanism which combines semantic information already present in the context and world knowledge to deduce information about upcoming, but not yet presented words; (ii) the combinatory process must be sensitive to a semantic decomposition in terms of semantic features in order to account for the graded character of expectations; and (iii) there must be a separate mechanism for retracting wrong guesses made on the basis of incomplete information.

3 The Formal Theory: Defaults and Belief Revision

The description at the end of the previous section suggests that online semantic processing involves some kind of nonmonotonicity. Reconsider example (3); after semantically processing the context prior to the target word at the end of the second sentence, all that is known about the concept expressed by that word is (i) the resort is supposed to look tropical and that (therefore) (ii) something is planted along the driveway. From this information conclusions about semantic features of the theme argument are drawn. Likely candidates are (a) type=plant, (b) category=tree, and (c) habitat=tropics. However, these conclusions are *defeasible*. If the upcoming word is eventually semantically recognized, the predictions made on the basis of the prior context can turn out to be false. This always happens for within-context-violations and between-context-violations. E.g., if in (3) 'pine' is the theme argument, 'habitat=tropics' turns

[6] Section 4 discusses alternative interpretations of the results..

out to be false though the other predictions turn out to be true. As an effect, 'habitat=tropics' has to be withdrawn because it is not part of the semantic representation of 'pine'. By contrast, for the best completion 'palm', all information predicted before the word is encountered applies.

Nonmonotonicity will be modelled by default rules. Such rules describe the expectations of the comprehension system. Schematically, such expectations have the form $A \Rightarrow B$, with A being some piece of (factual) information provided by the context through bottom-up processing and B being the conclusions which normally follow from A. Here, 'normally' refers to the fact that A is all that is known about an object. The conclusions B are defeasible. For example, if in addition to $A \Rightarrow B$ one has $C \Rightarrow \neg B$ and $C \Rightarrow A$ then C is an exceptional A w.r.t. the property expressed by B. Thus, if in addition to A C is also known about the object (so that A is not only known), $\neg B$ should (normally) be true of the object. Applied to our running example of the resort which should look tropical, one has $A \doteq (i) \wedge (ii)$; $B \doteq (a) \wedge (b) \wedge (c)$ and $C = (\text{sort=pine} \vee \text{sort=tulip})$. Thus, additional factual information can invalidate a prior inference based on less specific information. One therefore has: if both $A \Rightarrow B$ and $A \wedge C \Rightarrow \neg B$, $A \wedge C \Rightarrow \neg B$ should be used to draw the (default) conclusion $\neg B$ since one has $A \wedge C \supset A$, i.e. the antecedent of the second default rule $A \wedge C \Rightarrow \neg B$ implies that of the first one $A \Rightarrow B$. This reflects the fact that during online semantic processing conclusions drawn by more specific (less incomplete) information always overwrite conclusions drawn on the basis of less specific (more incomplete) information. What is required, therefore is an ordering on default rules which reflects this strategy. Since default conclusions can turn out to be wrong, there must be an additional mechanism of how to retract such wrong guesses. On the account just sketched, semantic processing therefore not only comprises decompositional semantic representations of items in the lexicon together with a set of recursive composition rules but, in addition, the following two components: (i) a set of default rules, which are used to draw defeasible conclusions (B) from factual information (A), and (ii) a mechanism for retracting conclusions got from applying rules in (i).

The relation to the ERP components, the N400 and the two kinds of late positivity, is the following. Default rules are correlated to the N400 and therefore to the first stage of online semantic processing. The relevant parameter is the difference between those semantic features derived after semantic recognition of the target word and those features derived prior to that recognition. This difference reflects the additional features that have to be activated. The two late positivities are correlated with those semantic features that were predicted prior to the semantic recognition of the target word but which turn out to be false and which therefore have to be retracted.

We will develop the formal theory in two steps. Building on [Bou94], we begin by defining default rules as a conditional $\Rightarrow$ in a modal logic with a Kripke-style semantics based on a normality ordering which reflects the expectations a comprehender has for a particular constituent of a sentence in a given context. Such models are the appropriate level to reason about the whole set of defaults represented by that model. Which default conclusions can be drawn depends on the available factual information. Such reasoning is best modeled in a particular model based on a (priority) ordering on defaults. This leads to system Z, [GP92], which will be introduced in the second step.

3.1 Formal Theory I: ⇒ and *CO*-models

The conditional logic chosen is that of Boutilier, [Bou94]. In this theory, the conditional connective ⇒ is not a primitive but is defined inside a modal logic using modal operators. One reason for choosing this framework is its generality. Besides default reasoning, it also allows to model belief revision. In addition, Boutilier's logic incorporates other approaches, in particular that of Pearl, [Pea90], in the sense that those logics are equivalent to fragments of Pearl's logic. This makes it possible to use either of these formalisms, depending on the context.

The basic idea underlying [Bou94] is to order situations (modeled as possible worlds in terms of valuations in a Kripke model) according to some measure of normality. This measure is represented by an accessibility relation $\geq_N$ on worlds. One has $w \geq_N v$ iff v is at least as normal as w. $w >_N v$ holds if v is strictly more normal than w, that is if $w \geq_N v$ and not $v \geq_N w$. The relation $\geq_N$ is required to be (i) transitive and (ii) totally connected from which together reflexivity follows: (i) $\forall uvw : u \geq_N w \wedge w \geq_N v \supset u \geq_N v$, and (ii) $\forall wv : w \geq_N v \vee v \geq_N w$. Models in which (i) and (ii) hold consist of totally ordered clusters of worlds, where a cluster is any maximal set of worlds s.t. $w \geq_N v$ for each w, v in this set, i.e. the elements of a cluster are all equally normal and the cluster is maximal w.r.t. this condition. If the set of worlds is finite, this chain of clusters has both a minimal and a maximal element. Furthermore, this ordering determines a normality ranking for each cluster and, therefore, for each world in W.[7]

Next, the language L_{Frame} is defined. As was shown in the first section, the information predicted is rather specific. We will therefore use a frame-based approach [Pet07]. Frames are recursive rooted attribute-value structures.[8] A modal language for talking about such structures is given by a set $\{P_\sigma\}_{\sigma\in\Sigma}$ of sort symbols ($\Sigma = \{tree, palm, \ldots\}$) and a set $\{Attr_{at}\}_{at\in ATTR}$ of attribute symbols ($ATTR = \{habitat, look, \ldots\}$). Elements of $\{P_\sigma\}_{\sigma\in\Sigma}$ are interpreted as unary relations and elements of $\{Attr_{at}\}_{at\in ATTR}$ as binary relations on a set of nodes.

[7] The ordering $\geq_N$ depends both on the kind of context and the comprehender. The dependency on the context corresponds to the distinction between strongly constraining and weakly constraining contexts. In a strongly constraining context there are more expectations than in a weakly constraining context. The dependency on a comprehender is illustrated by the following example concerning the moral value system of a comprehender. [BHN+09] presented examples like 'I think euthanasia is an acceptable course of action' to members of a relatively strict Dutch Christian party and to non-Christian respondents with sufficiently contrasting moral value systems. The result was that for both groups there was an enhanced N400 though it was larger for members of the strict Dutch Christian party.

[8] Note that [Pet07] allows unrooted frames as well, but such frames are of no interest for our purpose.

Formulae are of the form $at_1 : at_2 : \ldots at_n = \sigma$, expressing that the value at the end of the sequence of attributes $at_1 : at_2 \ldots at_n$ is of sort σ. They therefore express properties of nodes, as can be seen by looking at the standard translation of such a formula in first-order logic: $\lambda x \exists y_1 \ldots \exists y_n . at_1^*(x, y_1) \wedge \ldots \wedge at_n^*(y_{n-1}, y_n) \wedge \sigma^*(y_n)$ (see [PO14] for details). By interpreting such formulae at the root of a frame, a frame can be described by what is true at its root. On this perspective, frames can be taken as points (possible worlds) in a model. Formulae of the form $at_1 : at_2 \ldots at_n = \sigma$ are then atomic propositions in the language L_{Frame}. In addition, L_{Frame} has three modal operators $\Box$, $\overleftarrow{\Box}$ and $\Box_>$. While $\Box A$ refers to all accessible (i.e. equally or more normal) worlds in the ordering $\geq_N$, $\overleftarrow{\Box} A$ means that A is true at all inaccessible worlds, i.e. at all worlds which are strictly less normal than the world at which $\overleftarrow{\Box}$ is evaluated. $\Box_>$ is the strict variant of $\Box$. Models for L_{Frame} are defined below.

Definition 1 (A CO-model; [Bou94, 101]). *A CO-model is a triple $\langle W, \geq_N, V\rangle$ s.t. (i) W is a non-empty, finite set of worlds, (ii) $\geq_N$ is a binary relation on W that is transitive and totally connected and (iii) V is a valuation function for the atomic formulas in L_{Frame}.*

Truth of a formula is defined as follows.

Definition 2. *Let $M = \langle W, \geq_N, V\rangle$ be a CO-model with $w \in W$. The truth of a formula A at w in M is defined inductively by*

(i) $M \models_w A$ *iff* $w \in V(A)$ *for atomic sentence* A.
(ii) $M \models_w A \supset B$ *iff* $M \models_w B$ *or not* $M \models_w A$.
(iii) $M \models_w \neg A$ *iff not* $M \models_w A$.
(iv) $M \models_w \Box A$ *iff for each* v *s.t.* $w \geq_N v : M \models_v A$.
(v) $M \models_w \overleftarrow{\Box} A$ *iff for each* v *s.t.* $w \not\geq_N v : M \models_v A$.
(vi) $M \models_w \Box_> A$ *iff for each* v *s.t.* $w >_N v : M \models_v A$.

In terms of $\Box$ and $\overleftarrow{\Box}$ the following modal operators are defined.

Definition 3 (Defined modal operators).

1. $\Diamond A \equiv_{df} \neg\Box\neg A$.
2. $\overleftarrow{\Diamond} A \equiv_{df} \neg \overleftarrow{\Box} \neg A$.
3. $\overleftrightarrow{\Box} A \equiv_{df} \Box A \wedge \overleftarrow{\Box} A$.
4. $\overleftrightarrow{\Diamond} A \equiv_{df} \Diamond A \wedge \overleftarrow{\Diamond} A$.

One has: $\Diamond A$ is true at $w \in W$ iff A is true at some equally or more normal world v; similarly, $\overleftarrow{\Diamond} A$ holds at w just in case A holds at some strictly less normal world v; $\overleftrightarrow{\Box} A$ holds at a world w iff A is true at each world $w \in W$; $\overleftrightarrow{\Diamond} A$ is true at w iff A is true somewhere in the model, i.e. if there is a world $v \in W$ at which A is true. The conditional $\Rightarrow$ is defined in Definition 4.

Definition 4 ([Bou94, 104]). $A \Rightarrow B \equiv_{df} \overleftrightarrow{\Box} \neg A \vee \overleftrightarrow{\Diamond} (A \wedge \Box(A \supset B))$.

According to Definition 4, $A \Rightarrow B$ is true at a world w just in case either A is false at every world in the chain of worlds, i.e. the conditional is satisfied vacuously, or at the most normal A-worlds ($A \supset B$) holds. The truth of $A \Rightarrow B$ is independent of a particular possible world. If $A \Rightarrow B$ holds at some w, then it holds at all $v \in W$. This follows from the fact that the disjuncts in the definition of $\Rightarrow$ are modally decorated by $\overleftrightarrow{\Box}$ and $\overleftrightarrow{\Diamond}$, respectively. As a consequence, the truth of $A \Rightarrow B$ only depends on the complete ordering of worlds.

A CO-model represents the set of default rules Δ_D of a comprehender w.r.t. an argument (or a constituent) of a sentence in a given context. Together with factual information A got from bottom-up processing of the prior context (and, possibly, world knowledge), default rules $A \Rightarrow B$ are used to (defeasibly) infer B. More generally, one has: the local epistemic state of a comprehender w.r.t. an upcoming word is a quadruple $ES = \langle \Gamma, \Gamma^*, \Delta_D, \Delta_E \rangle$. Δ_D is a set of defaults of the form $A \Rightarrow B$ and Δ_E is the set of expectation rules given by the corresponding material conditionals $A \supset B$.[9] Γ is the set of factual information about the word. Before the word is semantically recognized it contains information got from the context. Upon recognition of the word, sortal information, e.g. $sort = palm$ is added. Γ^* is a set of default conclusions pertaining to the target word. They are inferred using Γ and Δ_E.

The reason for distinguishing Γ and Γ^* is directly related to the way semantic information is used in default rules $A \Rightarrow B$. The antecedent contains factual information from bottom-up semantic processing. This information is stored in Γ. By contrast, the information B in the consequent of a default rule is used to build up a partial semantic representation of an upcoming word. Since this information is in general defeasible (the problem of 'wrong guesses'), it is not directly integrated with the factual information stored in Γ but stored separately in Γ^*. This reflects the distinction between *lexical access* (first stage of semantic processing) and *integration* (second stage of semantic processing). During semantic processing, Γ and Γ^* are constantly updated whereas both Δ_D and Δ_E remain fixed.

3.2 Defaults and Online Semantic Processing

Next we will apply CO-models to online semantic processing. As our running example we will take (3), repeated below for convenience.

(8) They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of *palms/pines/tulips*.

After processing (8) up to the final world, the comprehender has got the following factual information which is relevant for drawing default conclusions about the object planted.

[9] The reason for distinguishing Δ_D and Δ_E will become clear if a ranking on the set Δ_D of default rules using System Z is defined. See below for details.

(9) a. *resort:look=tropical.*
b. *resort:driveway:adornment*$=\top$.

Let this information be A_0. This information is related to the following default rules.

(10) a. $A_0 \Rightarrow$ *resort:driveway:adornment:type=plant.*
b. $A_0 \Rightarrow$ *resort:driveway:adornment:category=tree.*
c. $A_0 \Rightarrow$ *resort:driveway:adornment:sort:habitat=tropics.*

When taken together, one gets default rule r_0 in (11).

(11) $r_0 : A_0 \Rightarrow$
resort:driveway:adornment:type=plant $\wedge$
resort:driveway:adornment:category=tree $\wedge$
resort:driveway:adornment:sort:habitat=tropics.

The material conditional r_0^* corresponding to r_0 is (12).

(12) $r_0^* : A_0 \supset$
resort:driveway:adornment:type=plant $\wedge$
resort:driveway:adornment:category=tree $\wedge$
resort:driveway:adornment:sort:habitat=tropics.

What happens if the upcoming word is eventually encountered and semantically recognized? In our frame theory, the information provided by a common noun like ‘palm’ is taken as sortal information. In (8), this is the value of the *sort*-attribute. Thus, if ‘palm’ is semantically recognized

resort:driveway:adornment:sort=palm

is added to Γ. The default rule corresponding to this information is r_1.

(13) r_1: $A_0 \wedge$ *resort:driveway:adornment:sort=palm* $\Rightarrow$
resort:driveway:adornment:type=plant $\wedge$
resort:driveway:adornment:category=tree $\wedge$
resort:driveway:adornment:sort:habitat=tropics.

Rule r_1 differs from r_0 in one respect. Its antecedent is more specific than that of r_0 ($A_1 \supset A_0$). This reflects the fact that r_0 is used in a situation of incomplete information, i.e. the upcoming word has not yet been semantically recognized whereas r_1 is used after that recognition has taken place. The consequents are the same because ‘palm’ is the best completion and therefore all predicted properties apply to the word encountered. The general pattern between these two default rules is given in (14).

(14) $r_0 : A \Rightarrow B.$
$r_1 : A \wedge C \Rightarrow B.$

This pattern can be taken as showing that encountering the best completion amounts to a confirmation of the expectations drawn when this word is not yet encountered.[10] The situation is different if instead of the best completion a within-context-violation like ‘pine’ is found. Similar to the case of ‘palm’, new sortal information is added to the factual information,

Γ: *resort:driveway:adornment:sort=pine.*

One also has that the antecedent of the corresponding default rule is more specific than that of r_0. But in this case the two consequents are logically incompatible because B_0 contains *resort:driveway:adornment:sort:habitat=tropics* whereas B_2 contains *resort:driveway:adornment:sort:habitat=moderate.*

(15) r_2: $A_0 \wedge$ *resort:driveway:adornment:sort=pine* $\Rightarrow$
resort:driveway:adornment:type=plant $\wedge$
resort:driveway:adornment:category=tree $\wedge$
resort:driveway:adornment:sort:habitat=moderate.

The general relation between the two default rules is given in (16).

(16) $r_0 : A \Rightarrow B.$
$r_2 : A \wedge C \Rightarrow \neg B.$

The case for ‘tulip’ should by now pose no problems. The default rule is r_3.

(17) r_3: $A_0 \wedge$ *resort:driveway:adornment:sort=tulip* $\Rightarrow$
resort:driveway:adornment:type=plant $\wedge$
resort:driveway:adornment:category=flower $\wedge$
resort:driveway:adornment:sort:habitat=moderate.

[10] According to rule r_0, an expectation w.r.t. to the theme argument of ‘plant’ does not include sortal information. Thus, there is no bias towards any tropical tree in the context of A_0. For example, both ‘palm’ and ‘eucalyptus’ are equally expected. However, if ‘palm’ is the best completion one may argue that this information is already activated prior to the encounter of the argument. Thus, rule r_0 seems to apply to weakly constraining and not to strongly constraining contexts. However, if sortal information is part of the consequent of the default rule, alternatives (‘eucalyptus’) to the best completion (‘palm’) are excluded. E.g., rule r_0 becomes r_{00}.

(i) r_{00} $A_0 \Rightarrow B_0 \wedge$ *resort:driveway:adornment:sort=palm.*

Using r_{00}, r_1 becomes redundant because upon encountering ‘palm’ no new information needs to be added. Rule r_1 is replaced by the following rule for the sort ‘eucalyptus’.

(ii) $r_1 : A_0 \wedge resort{:}driveway{:}adornment{:}sort{=}eucalyptus \Rightarrow$
$B_0 \wedge resort{:}driveway{:}adornment{:}sort{=}eucalyptus.$

An open empirical question is the relation between N400 effects both in strongly constraining and weakly constraining contexts for ‘palm’ and ‘eucalyptus’, i.e. two concepts that are of the same type, here ‘plant’, but also of the same category. here ‘tree’, and that both fulfill the conditions specified in the consequent of rule r_0.

Similar to the case of 'pine', the consequent is logically incompatible with that of r_0 (and also with that of r_2). In contrast to 'pine', there are two conjuncts which are logically incompatible. Besides the one specifying the value of the HABITAT-attribute, this also holds for the value of the SORT-attribute.[11]

A drawback of the rules r_1–r_3 is that they contain redundant information. This is the case whenever they contain information that is also specified in the rule r_0. This information will not be retracted even when a non-best completion is encountered. An alternative is to only specify that information which is incompatible with information given by r_0. Applied to the processing level, this means that once a feature is activated it need not be activated a second time. At the formal level, one uses the following property of formulae.

Definition 5 (Downward closed property). *A formula A is downward closed iff* $\overset{\leftrightarrow}{\Box}(A \supset \Box_{>} A)$.

According to this definition, a formula is downward closed if its truth at a world w implies that it holds at all strictly more normal worlds. The revised rules $r'_1 - r'_3$ are given in (18).

(18) r'_1: $A_0 \wedge$ *resort:driveway:adornment:sort=palm* $\Rightarrow$ *true.*
r'_2: $A_0 \wedge$ *resort:driveway:adornment:sort=pine* $\Rightarrow$
resort:driveway:adornment:sort:habitat=moderate.
r'_3: $A_0 \wedge$ *resort:driveway:adornment:sort=tulip* $\Rightarrow$
resort:driveway:adornment:category=flower $\wedge$
resort:driveway:adornment:sort:habitat=moderate.

A possible model for the default rules is given in Fig. 2. This model is based on a knowledge base corresponding to our running example: the objects planted are either palms, pines or trees and there are no 'abnormal' instances of those sorts.[12]

In L_{Frame}, the four clusters can be formally characterized as follows. To begin, note that the formula $\Box A \wedge \overset{\leftarrow}{\Box} \neg A$ holds at a world w_0 if A is true at all equally or more normal worlds w_1 whereas at all worlds w_2 which are strictly less normal

[11] One may argue that rules $r_1 - r_3$ are strict and not defeasible. For example, a palm is a tree and not a flower. However, in the present context we are interested in the way a comprehender uses information, both top-down and bottom-up, to build a semantic representation of a constituent. What matters, therefore, is the relation between the various rules he uses (the priority ordering) and not the status of an individual rule as defeasible or strict. For example, rule r_2 has a higher priority than rules r_0 and r_1 because it describes a situation which is assumed to be less normal. In addition, not all conjuncts in the consequent of a rule are non-defeasible, given the antecedent. For example, the tropics are only normally the habitat of palms, but they grow in moderate habitats as well (e.g., in botanical gardens in Europe).

[12] These restrictions are due to the fact that we do not have any information about the way, say, orchids (tropical flowers) or palms whose habitat are not the tropics are semantically processed online. Additional experimental data is needed to tackle this question.

Cluster	**0** (palm)	**1** (pine)	**2** (tulip)
tropics	true	false	false
tree	true	true	false
flower	false	false	true
plant	true	true	true

Fig. 2. Possible model for the running example

Cluster	$\Box A \wedge \overleftarrow{\Box} \neg A$
0	$A \doteq habitat{=}tropcis$
1	$A \doteq category{=}tree$
2	$A \doteq type{=}plant$

Fig. 3. Relation between clusters and properties of objects adorning the driveway

A is false. This formula can therefore be seen as expressing a kind of 'frontier'. All worlds above the frontier satisfy A whereas all worlds below it fail to satisfy it. The relation between this formula, properties of the objects adorning the driveway and clusters are shown in Fig. 3. Thus, cluster 0 is a frontier for the property *habitat=tropcis* (for ease of readability, only the last attribute of a chain of attributes is displayed) whereas clusters 1 and 2 are frontiers for the properties *category=tree* and *type=plant*, respectively. This correlation between clusters and properties shows that of the three properties assumed in a most normal situation, *habitat=tropics* is the least entrenched one or the first to be given up. Similarly, *type=plant* is the most entrenched one whereas *category=tree* has a position intermediate between those two properties. Intuitively, one can say that 'tropics'-worlds only see 'tropics'-world and similarly for 'tree'- and 'plant'-worlds. The difference shows up if one looks backwards. 'tree'-worlds are either seen by 'non-tropics'-worlds or if in the same non-minimal cluster, 'tree'-worlds are always 'non-tropics'-worlds: $\overleftrightarrow{\Box}(tree \supset (tree \wedge \neg tropics) \vee \overleftarrow{\Box} \neg tropics)$. Thus, 'tropics'-worlds are more normal than 'tree'-worlds. Furthermore, one has $\overleftrightarrow{\Box}(flower \supset \Diamond\Box tree)$: 'flower'-worlds are no more normal than 'tree'-worlds. Finally, one has $\overleftrightarrow{\Box}(plant \supset (tree \vee flower))$. The above properties are global in the sense that their truth is independent of a particular world.

General CO-models are appropriate for specifying global properties of the local epistemic state of a comprehender w.r.t. an upcoming word. If a comprehender uses a CO-model to draw conclusions, it is more convenient to use a particular CO-model which is based on a priority ordering on default rules.

3.3 Formal Theory II: Defining an Ordering on Defaults

An ordering on default rules can be defined using procedure Z, [GP92]. Defeasible rules can be verified, falsified or satisfied at a world w.

Definition 6 (Verifying, falsifying and satisfying a default rule). *A possible world w in a model M* verifies *a conditional $A \Rightarrow B$ iff $M \models_w A \wedge B$; it* falsifies *$A \Rightarrow B$ iff $M \models_w A \wedge \neg B$, and it satisfies $A \Rightarrow B$ iff $M \models_w A \supset B$.*

The derivation of a Z-ordering of default rules is based on the notion of *toleration*, Definition 7.

Definition 7 (Toleration). *Δ_D is said to* tolerate *a default $A \Rightarrow B$ iff there is a world w that verifies $A \Rightarrow B$ and falsifies no rule in Δ_D, i.e.*

(19) $A \wedge B \wedge \bigwedge_{r_j \in \Delta_D} A_j \supset B_j$.

Toleration is used to define a natural ordering on a set of defaults by partitioning this set. The procedure for finding this partition works as follows. Let Δ be the set of defaults. In a first step all rules in Δ which are tolerated by all other rules are in Δ_0. Next, the set $\Delta' = \Delta - \Delta_0$ is considered. All rules in Δ' which are tolerated by all other rules in Δ' are in Δ_1. Next, the set $\Delta'' = \Delta' - \Delta_1$ is considered. Continuing in this way, yields a partition $\Delta_0, \Delta_1, \ldots, \Delta_n$ of Δ (provided Δ is consistent). This procedure is defined inductively in Definition 8 where $\Gamma(\Delta)$ is the set of defaults in Δ which are tolerated by Δ.

Definition 8 (Partition of a set of defaults). $\Delta_0 = \Gamma(\Delta)$ *and* $\Delta_{\tau+1} = \Gamma(\Delta - (\bigcup_{\sigma \leq \tau} \Delta_\sigma))$

Given this partition of Δ, the rank of a default $A \Rightarrow B \in \Delta$ is defined by $Z(A \Rightarrow B) = \tau$ iff $A \Rightarrow B \in \Delta_\tau$. The intuition is that lower ranked defaults are more general and have a lower priority.

Next, the ranking of a world w is defined. The rank of a world w is the smallest integer τ s.t. all defaults having a rank higher or equal to τ are not falsified by w. This condition is expressed by: w satisfies $\bigcup_{\sigma \geq \tau} \Delta_\sigma$ or, equivalently by $Z(w) = min\{\tau : M \models_w A \supset B$ for all $r \in \Delta$ and $Z(r) \geq \tau\}$. The intuition is that lower ranked worlds are more normal. Thus, the Z-ranking on worlds determines a unique preferred structure Z_T.

The rank of a (non default) formula A is defined as follows.

(20) $\kappa^z(A) = min\{i \,|\, A \wedge \bigwedge_{j:Z(r_j) \geq i} A_j \supset B_j \, is \, satisfiable\}$.

Using this ranking on formulae, a formula B is said to be Z-entailed by a formula A iff the worlds in which A and B hold are strictly lower ranked than the worlds in which A and $\neg B$ hold, that is if the rank of $A \wedge B$ is strictly lower than the rank of $A \wedge \neg B$.

Definition 9 (Z-entailment). *A formula B is Z-entailed by a formula A w.r.t. Δ, written $A \vdash_Z B$, iff $\kappa^z(A \wedge B) < \kappa^z(A \wedge \neg B)$.*

(19) and (20) can be used to construct a theory *Th(A)* which characterizes precisely the set of conclusions B that defeasibly follow from factual information A, given a set Δ_D of default rules: $A \vdash_Z B$ iff *Th(A)* $\supset B$.

(21) $Th(A) = A \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A)} A_i \supset B_i$.

In our application, A is always factual information about an upcoming word (or an argument) got by bottom-up processing and stored in Γ. Δ_D (or Δ_E) is a set of default rules (expectations) which pertain to this argument. In our running example, this is the theme argument of the verb 'plant' in a given context. The $A_i \supset B_i$ are elements of Δ_E, i.e. material counterparts of default rules in Δ_D. The elements of Γ^* are those B_i which follow from *Th(A)*, i.e. from A and the $A_i \supset B_i$ with $A \supset A_i$.

3.4 Drawing Default Conclusions from Factual Information

Let us next apply system Z to our running example. We first construct a Z-ranking on $\Delta_D = \{r_0, r_1, r_2, r_3\}$. Rules r_1 and r_0 are tolerated by all the other rules. The following valuation verifies both rules:

resort:look=tropical $\wedge$ *resort:driveway:adornment:*$\top$
$\wedge$ *resort:driveway:adornment:type=plant* $\wedge$
resort:driveway:adornment:category=tree $\wedge$
resort:driveway:adornment:sort:habitat=tropics $\wedge$
resort:driveway:adornment:sort=palm.

Furthermore, one sets $\neg$*resort:driveway:adornment:sort=X* for $X \in$ {*pine*, *tulip*}. Since the antecedents of the rules r_2 and r_3 are pairwise logically incompatible, each rule tolerates the others. For example, verifying r_2 requires

resort:driveway:adornment:sort=pine.

Setting $\neg$*resort:driveway:adornment:sort=tulip* satisfies r_3. Here it is assumed that one has e.g. $tree \supset \neg flower$. Therefore, for $j \neq k$ with $j, k \in \{2, 3\}$ we get that if a world verifies $A_j \Rightarrow B_j$, it satisfies $A_k \Rightarrow B_k$ because A_k is false at that world. The Z-ranking on rules is $\Delta_0 = \{r_0, r_1\}$ and $\Delta_1 = \{r_2, r_3\}$.

As long as no factual information about the theme is given, one has $A = true$. No conclusions using the set of expectations Δ_E can be drawn. Furthermore, $\Gamma = \{true\}$, $\Delta_D = \{r_0, r_1, r_2, r_3\}$, $\Delta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$ and $\Gamma_0^* = \emptyset$. After processing the prior context, one has $A = A_0$ and $\Gamma = \{A_0\}$. Since $\kappa^Z(A_0) = 0$, one gets $Th\,(A_0) = A_0 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_0)=0} A_i \supset B_i$. Thus, $\Delta_D = \{r_0, r_1, r_2, r_3\}$ and $\Delta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$. The set of defeasible consequences Γ_0^* is deduced from $A = A_0$ and $A_0 \supset B_0$ yielding $\Gamma_0^* = \{B_0\}$. If 'palm' is encountered, the sortal information *sort=palm* is added to Γ so that $A = A_1$. Since $\kappa^Z(A_1) = 0$, one has $\Delta_D = \{r_0, r_1, r_2, r_3\}$ and $\Delta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$. The set of defeasible consequences is got from A_0, A_1, and $A_0 \supset B_0$ and $A_1 \supset B_1$, which yields $\Gamma_1^* = \{B_1\}$ since $B_1 \supset B_0$. If instead of r_1, r_1' is used no new (defeasible) information is added to Γ^*.

If a within-context-violation or a between-context-violation is encountered, the new sortal information is *sort=pine* or *sort=tulip* in our running example. It is added to Γ, yielding $A = A_2$ ('pine') or $A = A_3$ ('tulip'). In contrast to A_0 or

A_1, one has $\kappa^Z(A_2) = \kappa^Z(A_3) = 1$ so that $Th\,(A_2) = A_2 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_2)=1} A_i \supset B_i$ and $Th\,(A_3) = A_3 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_3)=1} A_i \supset B_i$. This means that the situation is not described as most normal. As a result, r_0 and r_1 can no longer be used. One rather gets $\Delta_D = \{r_2, r_3\}$ and $\Delta_E = \{r_2^*, r_3^*\}$.

For A_2 ($\doteq$ *sort = pine*), the conclusions one gets are given by A_2, $A_2 \supset B_2$, yielding B_2. Using r_2' instead of r_2^*, one has $B_2 = \{habitat{=}moderate\}$, i.e. $\Gamma_1^* = \{habitat{=}moderate\}$. For the BCV 'tulip', the situation is similar. Conclusions are got from A_3 and $A_3 \supset B_3$, yielding B_3. Using r_3' instead of r_3^*, the new derived information is *habitat=moderate* and *category=flower*, i.e. $\Gamma_1^* = \{habitat{=}moderate, category{=}flower\}$. Both for A_2 and A_3, it is not possible to directly add B_2 or B_3 to Γ^*, i.e. to use $\Gamma_0^* \cup \Gamma_1^*$. This would result in an unsatisfiable set because one would have both *habitat=tropics* (from the previous application of rule r_0 prior to the semantic recognition of the theme) and *habitat=moderate* from applying r_2^* or r_3^*. In addition r_3^* yields *category=flower* which conflicts with *category=tree*, again got from applying r_0 prior to encountering the theme argument. Despite the fact that Γ_0^* (got from applying r_0^*) and Γ_1^* (the information got from applying r_2^* or r_3^*) are logically incompatible, their union contains all semantic features necessary for building up a semantic representation of the theme argument.

Let us take stock and compare a best completion, a within-context-violation and a between-context-violation. One has: (a) in each case sortal information is added to the default conclusions got prior to encountering the argument, (b) they differ w.r.t. the set $\Gamma_1^* - \Gamma_0^*$, and (c) they differ w.r.t. the set $\Gamma_0^* - \Gamma_1^*$. The set $\Gamma_1^* - \Gamma_0^*$ is the set of semantic features that have to be activated in addition to those that were activated prior to the semantic recognition of the target word. By contrast, the set $\Gamma_0^* - \Gamma_1^*$ (using the rules r_i and not the rules r_i') is the set of semantic features that have to be retracted because they are 'wrong guesses'. Now consider the two hypotheses in (22).

(22) (i) The set $\Gamma_1^* - \Gamma_0^*$, i.e. the set of additional features to be activated, is related to the N400 effect, i.e. it is related to the first stage of online semantic processing: semantic access.

(ii) Predicting semantic features of an upcoming word can lead to wrong guesses. Those wrong guesses must be eliminated before the semantic representation of the target word can be added to the representation of the prior context. The set $\Gamma_0^* - \Gamma_1^*$, containing those wrong guesses, is related to the two late positivities and therefore to the second stage of online semantic processing.

In the introduction it was argued that online semantic processing can be split in (at least) two separate stages: lexical semantic access, indexed by the N400, and semantic integration, indexed by two late positivities. The former, lexical access, is based on predictions which are made prior to encountering the target word, and, therefore, on the basis of incomplete information. Such predictions are risky because they can turn out to be wrong. On the account presented in this paper, wrong guesses are directly related to the two stages of online semantic processing.

A wrong guess activates less semantic features of the actual target word; thus, lexical access is aggravated. Accessing additional semantic features comes with a computational cost because information needs to be retrieved from LTM. This cost is reflected in an enhanced amplitude of the N400. This is only one side of the coin. The other is, of course, that a wrong guess has to be retracted. This follows from the fact that predictions, be they based on incomplete information or on information after the word is encountered, are related to accessing the associated features in LTM. Thus, once a semantic feature has been activated using rule r_0, it has to be retracted if it turns out to be wrong after the target has been semantically recognized and before the target is integrated into the prior context. This operation is related to the second stage, the integration stage. As a result, integration becomes a two stage process: first retracting wrong guesses and only then integrating the semantic representation of the target with the representation of the prior context. The above correlations will be explained in more detail in the following sections.

3.5 The N400 and Default Reasoning

We hypothesize the following correlation between the N400 effect and default reasoning.

(23) *Correlation N400 – default reasoning:*
The N400 effect is monotonic to the difference between semantic features got after semantic recognition of the target word and prior to its semantic recognition.

According to (23), the N400 effect is correlated to the difference between the pre-activated features $\Gamma^*_{r_0}$ if only rule r_0 is used, representing the most normal expectations, and those features contained in the consequent of the rule used after the upcoming word is eventually being semantically recognized. One calculates the cardinality of $\Gamma^*_{r_i} - \Gamma^*_{r_0}$. The greater this cardinality, the greater the N400 effect. Thus, the N400 is a measure of the cost of activating additional semantic features after recognition of the target word. For our running example, this correlation is shown in Fig. 4.

word encountered	$\lvert\Gamma^*_{r_i} - \Gamma^*_{r_0}\rvert$	N400 amplitude
palm (best completion)	0	base line
pine (within-category violation)	1	a > base line
tulip (between-category violation)	2	b > a

Fig. 4. Default rules and N400 effects

If 'palm' is encountered, rules r_0 and r_1 apply. As shown above, no new features need to be activated so that all features already pre-activated become part of the frame-representation of the concept expressed by this word. If 'pine' is encountered, neither rule r_0 nor rule r_1 apply. Instead rule r_2 is used. In this case only one feature does not apply: *habitat=tropics* so that one new feature *habitat:moderate* of the concept expressed by 'pine' must be activated. For 'tulip', the situation is similar. The difference is that even fewer pre-activated features apply: *type=plant*. Therefore, more additional features have to be activated: *category=flower* and *habitat=moderate*.

The above criterion for the amplitude of the N400 can be refined in the following way. Instead of just counting the number of attributes, one considers in addition the degree of entrenchment of an attribute. For example, the attribute 'category', with its value 'plant', is more entrenched than the attribute 'type', with values 'tree' or 'flower'. Formally, such distinctions can be made in an extension of system Z, system Z^*, [GP91]. In Z^*, a default rule is of the form $A \overset{\delta}{\Rightarrow} B$. Intuitively, δ is a measure of strength or the degree of surprise of finding the default rule violated. Applied to the above example, one gets: the value of δ for *type=plant* is greater than that for *category=tree*.

3.6 Late Positivities and Belief Revision

Frontal Late Positivity. One difference between a best completion on the one hand and a within-context-violation and a between-context-violation on the other is the fact that for the latter but not for the former there are wrong guesses. At the formal level, this corresponds to the distinction between $\Gamma_0^* - \Gamma_1^*$ being empty or not. If this set is empty, the default conclusions drawn before the target word is encountered are completely confirmed. Formally, this process is an addition. First, A, the sortal information, is added to Γ^* and next $\Gamma^* \cup \{A\}$ is added to Γ.

(24) a. $\Gamma_{i+1}^* = \Gamma_i^* \cup \{A\}$.
b. $\Gamma_{i+1} = \Gamma_i \cup \Gamma_{i+1}^*$.

If a non-best completion is encountered, processing is more involved. This is a simple consequence of the fact that the comprehender knows that the situation described is not most normal and that therefore at least some of his expectations are not satisfied. First, default rule r_0 can no longer be used because the theory w.r.t. the target word has changed. Instead of $Th(A_0)$, $Th(A_i)$ with $2 \leq i \leq 3$ has to be used. Second, $Th(A_0)$ and $Th(A_i)$ are incompatible. Using (21), this is the case for the information B_i contained in the consequent of rule r_i. For example, if 'pine' is encountered, one gets *habitat=moderate*, which is incompatible with *habitat=tropics* got from applying r_0 during the first stage. Let this information, i.e. *habitat=moderate*, be A. One then has $\Gamma^* \models \neg A$. Therefore, it is not possible to simply add A to Γ^* because this would result in a set which is not satisfiable. Rather, one first has to retract $\neg A$ from Γ^*. Only after this has been done, the addition operation given in (24) can be applied. Formally, this amounts to a revision operation in terms of the Levi-identity.

(25) Levi-identity: $KB \overset{*}{-} A = (KB \overset{\cdot}{-} \neg A) + A$.

Revising a knowledge base KB with A amounts to first making KB consistent with A by removing $\neg A$ from KB and then adding A to the resulting KB from which $\neg A$ has been retracted. For a best completion, the retraction step does not apply because there is no new default information which is incompatible with information got during the first stage. As an effect, revising reduces to a simple addition. In the present context, KB is always Γ^*, i.e. the set of default conclusions got by applying r_0. A is the conjunction of the literals in the consequent of rule r_i, i.e. the conjunction of those literals which differ in the value assigned to an attribute from those in the consequent of r_0. Thus, the retraction operation is always applied to Γ^* and therefore to defeasible conclusions. This reflects the fact that defeasible information, i.e. information got from top-down processing using default rules, is always less entrenched than information got by bottom-up processing.

We hypothesize the following relation between the frontal late positivity and the formal process described above.

(26) *Correlation frontal late positivity – belief revision:*
A frontal late positivity is triggered whenever $\Gamma^* \overset{*}{-} A$ is a *proper* revision, i.e. if there is a non-empty retraction operation. In this case default conclusions drawn before the target word is encountered have to be retracted.

Parietal Late Positivity. As was shown in the previous section, the revision of Γ_i^* by the new information got from processing the target word is successful, not only for the best completion but also for a within-context-violation or a between-context-violation. This follows from the fact that both kinds of violation satisfy the minimal appropriateness condition imposed on the theme argument of the verb 'plant', namely the type of the object must be 'plant'. At the level of CO-models, this is expressed by the integrity constraint, $\overset{\leftrightarrow}{\Box}$ (*type=plant*). The effect of this constraint is that any attempt to update with information which does not satisfy this constraint, say *sort=groan* $\wedge$ *type=sound*, will fail because it leads to an inconsistent knowledge base. One has both *type=plant* and *type=sound*. The only way of blocking this result is to retract *type=plant* from the knowledge base. However, this is not admissible because it violates the integrity constraint (or, from the point of view of an attribute-value structure, the appropriateness condition). We hypothesize the following relation between a parietal late positivity effect and our model of semantic processing.

(27) *Correlation parietal late positivity – belief revision:*
A parietal late positivity is triggered whenever an integrity constraint is violated s.t. a 'normal' revision operation is not applicable.

It seems that a sentence like 'For the snowman's eyes the kids used two pieces of coal. For his nose they used a groan from the fridge' is interpretable only if at least part of the sentence is not taken in its literal sense. For example, the word 'groan' could be used in such a way that it refers to a nose-like object which emits a groan-like noise when squeezed. The general interpretative strategy behind this example can be described as follows. The information provided by the head noun is *not* taken as specifying the sort of the frame (e.g., it is a groan) but rather as giving a particular property of the object denoted by that frame, e.g. the value of a sound quality. The task of making sense of such sentences, then, consists in finding a frame s.t. (a) it satisfies the appropriateness condition, and (b) it has an attribute whose value can be of the sort denoted by the head noun.

Parietal late positivity also shows that during online semantic processing conclusions derived from more specific information do not always win out. Though it is true that semantic features got after semantic recognition overwrite contrary semantic features got prior to semantic recognition, features imposed on the argument can never be overwritten. Formally, this is expressed by having such a constraint be true at all worlds in a CO-model.

On the account developed above, integration/composition is always done w.r.t. a consistent set of features that are either imposed, predicted prior to recognition or got after semantic recognition. As a consequence, integration/composition are sensitive to differences between words of the same syntactic category denoting objects of the same type.

Summarizing, we have arrived at the following correlations. Late positivity effects are triggered whenever a prediction must be given up. Frontal late positivities are correlated with wrong guesses which do not violate a sortal type restriction on an argument of the verb. This is the case for within-context-violations and for between-context-violations. In this case integration, defined by the revision operation *, is possible. By contrast, for parietal late positivity effects, a violation of such a sortal type restriction occurs. In such a case normal integration is not possible.

4 Comparison to Other Approaches

In this section we will compare our model with other approaches and discuss some possible objections. First, we summarize the main theses underlying our model.

1. The N400 effect is related to lexical retrieval of items in LTM. In particular, it is directly related to the number of additional features (attribute-value pairs) that must be activated compared to the features which have already been activated during prediction. Hence, the N400 is not related to integration and/or composition.
2. The two late positivities are related to integration/composition: in order to arrive at a consistent new semantic representation (or knowledge base), predictions that are incompatible with the information got by bottom-up processing have to be given up. Formally, this amounts to revising the predictions with the (true) bottom-up information.
3. When taken together, (1) and (2) yield the following model of semantic processing in the brain: The first stage, indexed by the N400, is related to semantic retrieval whereas the second stage is related to integration/composition,

which consists in a revision component made up by a retraction followed by an addition.

Thesis (1) is incompatible with a widely held view of N400 effects: the *integration view*. We will therefore begin by providing a critical assessment of this view in Sect. 4.1. According to thesis (2), the P600 is a semantic effect, However, this is at odds with the widely held view according to which it reflects syntactic violations and syntactic repairing. Evidence for our interpretation will be discussed in Sect. 4.2.

4.1 Integration View of the N400

On the *integration view*, the amplitude of the N400 is related to the effort of integrating a word in the current context, i.e. in the semantic representation built up so far. On this interpretation, N400 effects are (i) post lexical, i.e. they occur after a word has (semantically) been recognized and (ii) result from a combinatorial process. After a word has been accessed in LTM, the task consists in combining the semantic representation of the prior context with the semantic representation of that word. Thus, at any moment during semantic processing the set of semantic features is solely built up by words that have already been processed (or recognized) and not by features of (expected) upcoming words farther down the sentence.

As noted by [Pyl12], a general problem of this account of the N400 is that the notion of 'semantic integration' is usually not sharply defined. As was already shown in the introduction, according to (most) formal semantic theories, composing a word after accessing it with the previous context does not depend on the way it is semantically related to this context in detail but merely on its general syntactic and semantic type. Furthermore, it is usually not explained why semantic expectedness and/or relatedness should affect semantic integration. Besides these theoretical weaknesses, this account also faces a number of severe empirical problems. First, the mismatch between the set of semantic features pre-activated by a prior context and a within-context-violation is greater in a strongly constraining than in a weakly constraining context so that it should be more difficult to integrate a WCV, like 'pine', in a strongly constraining context compared to a weakly constraining context. As an effect, the N400 amplitude in a strongly constraining context should exceed that in a weakly constraining context, contrary to the empirical findings. Second, the *integration view* predicts that for two words which are synonymous the difficulty in integrating them should be the same since they are necessarily semantically related to the prior context in exactly the same way. This prediction is falsified by example (5) in Sect. 2.2. Since 'a' and 'an' have exactly the same meaning, they should elicit the same N400 effects. However, the amplitude of the N400 was larger for 'an' compared to 'a'.

4.2 The P600 as a Measure of Multimodal Updating Processes

According to our model, the late positivity (P600) is related to semantic integration.[13] Some predictions that have been made have to be given up because they are incompatible with the (semantic) information provided by the target word. This integration view of the P600 seems to be at odds with the popular *syntactic view* of this ERP-component. On this view, the P600 is interpreted as indexing the difficulty of revising or repairing a syntactic analysis when the target word makes the sentence based on this analysis ungrammatical (see [BFH12, 135] and [GPKP10] for an overview).

The syntactic view of the P600 has been challenged by a number of empirical results. First, [KHGH00] (see also [BFH12]) compared sentences with long distance wh-dependencies.

(28) a. Emily wonders who the performers in the concert imitate ...
b. Emily wonders whether the performers in the concert imitate ...

Only for (28a), but not for (28b), a P600 was found. Since (28a) is neither syntactically ill-formed nor does it contain a garden-path, this effect has to be explained in a different way. [KHGH00] suggest that in this case this effect reflects a process of syntactic integration: the verb 'imitate' has to be linked to the wh-pronoun 'who' whereas no such additional operation is needed in the case of (28b). Thus, the P600 is related to integration and not only to repairing. In addition, the linking that is required can be interpreted as being semantic in nature. A further example of a semantic P600 effect is given by so-called bridging phenomena, [Bur06,Sch13].

(29) Yesterday, a PhD. student was *shot/killed/found dead* downtown. The press reported that the pistol was probably from army stocks.

Both for 'killed' and 'found dead', [Bur06] found a P600 effect but not for 'shot'. According to [Bur06], this can be explained by assuming that in the former two cases establishing a coherent discourse relation (say, elaboration) requires more inferential work. Again, this is a purely semantic (or pragmatic) task which is related to integrating new information into the semantic representation of the previous context. In their discussion of the findings by [RGF11,Bur06], [BFH12, 136] conclude that 'what their materials have in common is that they require additional processing (as compared to the control condition) in order to arrive at a coherent mental representation of what the speaker or writer meant to communicate'. The authors hypothesize that all P600 effects can be described in terms of the construction, revision, or updating of a mental representation of what is being communicated, [BFH12, 137]. They argue that on this account of the P600 the observed effect reflects the efforts in reworking an initial mental representation and not the revision of a syntactic analysis. Thus for them, the P600 reflects integration difficulty. This difficulty is determined by how much the

[13] This section owes much to the review article [BFH12].

current mental representation needs to be adapted to incorporate the current input. They summarize their view of the P600 as follows [BFH12, 138]: 'The P600 component is the brain's natural electrophysiological reflection of updating a mental representation with new information.' Hence syntactic complexities and violations elicit a P600 effect because they reflect difficulties in building up a coherent mental representation at the syntactic level. Even more important than [BFH12]'s account of the P600 effect is the way they relate the N400 to this ERP-component. According to them, the N400 reflects the retrieval of the meaning of a word from LTM, [BFH12, 128].

5 Summary

In this paper we showed how a semantic theory can be extended to incorporate a 'predictive' and a 'revision' component in order to account for neurophysiological data on online semantic processing. The general idea is to use a decompositional frame semantics based on typed attribute-value structures together with a set of default rules. Such rules are in part pragmatic because their use is context dependent. Yet, the information inferred is always part of the semantic representation of a concept in LTM since the context only determines which part of the frame representing the concept is activated.

References

[Ber10] Van Berkum, J.J.A.: The brain is a prediction machine that cares about good and bad - any implications for neuropragmatics? Ital. J. Linguist. **22**, 181–208 (2010)

[BFH12] Brouwer, H., Fitz, H., Hoeks, J.: Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. Brain Res. **1446**, 127–143 (2012)

[BH11] Baggio, G., Hagoort, P.: The balance between memory and unification in semantics: a dynamic account of the N400. Lang. Cogn. Process. **26**(9), 1338–1367 (2011)

[BHN+09] Van Berkum, J.J.A., Holleman, B., Nieuwland, M., Otten, M., Murre, J.: Right or wrong? the brain's fast response to morally objectionable statements. Psychol. Sci. **20**, 1092–1099 (2009)

[Bou94] Boutilier, C.: Conditional logics of normality: a modal approach. Artif. Intell. **68**(1), 87–154 (1994)

[Bur06] Burkhardt, P.: Inferential bridging relations reveal distinct neural mechanisms: evidence from event-related brain potentials. Brain Lang. **98**(2), 159–168 (2006)

[DQK14] DeLong, K.A., Quante, L., Kutas, M.: Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia **61**, 150–162 (2014)

[DUK05] DeLong, K.A., Urbach, T., Kutas, M.: Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat. Neurosci. **8**, 1117–1121 (2005)

[FK99] Federmeier, K.D., Kutas, M.: A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. **41**, 469 (1999)

[FWODK07] Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M.: Multiple effects of sentential constraint on word processing. Brain Res. **1146**, 75–84 (2007)

[GP91] Goldszmidt, M., Pearl, J.: System-Z+: a formalism for reasoning with variable-strength defaults. In: Dean, T.L., McKeown, K. (eds.) Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, 14–19 July, Vol. 1, pp. 399–404. AAAI Press/The MIT Press (1991)

[GP92] Goldszmidt, M., Pearl, J.: Rank-based systems: a simple approach to belief revision, belief update, and reasoning about evidence and actions. In: Nebel, B., Rich, C., Swartout, W.R. (eds.) Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR 1992), pp. 661–672. Morgan Kaufmann (1992)

[GPKP10] Gouvea, A.C., Phillips, C., Kazanina, N., Poeppel, D.: The linguistic processes underlying the P600. Lang. Cogn. Process. **25**(2), 149–188 (2010)

[KF11] Kutas, M., Federmeier, K.: Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. **62**(1), 621–647 (2011)

[KHGH00] Kaan, E., Harris, A., Gibson, E., Holcomb, P.: The P600 as an index of syntactic integration difficulty. Lang. Cogn. Process. **15**(2), 159–201 (2000)

[LPP08] Lau, E.F., Phillips, C., Poeppel, D.: A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. **9**, 920–933 (2008)

[Pea90] Pearl, J., System, Z.: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: Parikh, R. (ed.) Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, March 1990, pp. 121–135. Morgan Kaufmann (1990)

[Pet07] Petersen, W.: Representation of concepts as frames. In: Skilters, J., Toccafondi, F., Stemberger, G. (eds.) Complex Cognition and Qualitative Science, The Baltic International Yearbook of Cognition, Logic and Communication, vol. 2, pp. 151–170. University of Latvia (2007)

[PO14] Petersen, W., Osswald, T.: Concept composition in frames: focusing on genitive constructions. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Frames and Concept Types. Studies in Linguistics and Philosophy, vol. 94, pp. 243–266. Springer, Cham (2014)

[Pyl12] Pylkkänen, L.: N400. http://www.psych.nyu.edu/pylkkanen/Neural_Bases/07_slides/18_N400.pdf. Accessed 10 Sept 2015

[RGF11] Regel, S., Gunter, T.C., Friederici, A.D.: Isn't it ironic? an electrophysiological exploration of figurative language processing. J. Cogn. Neurosci. **23**(2), 277–293 (2011)

[Sch13] Schumacher, P.B.: Content and context in incremental processing: "the ham sandwich" revisited. Philos. Stud. **168**(1), 151–165 (2013)

A Probabilistic, Mereological Account of the Mass/Count Distinction

Peter R. Sutton(✉) and Hana Filip(✉)

Heinrich Heine Universität, Düsseldorf, Germany
peter.r.sutton@icloud.com, hana.filip@gmail.com
http://peter-sutton.co.uk
https://user.phil-fak.uni-duesseldorf.de/filip/

Abstract. In this paper, we attempt to answer the vexing question why it should be the case that only certain types of noun meanings exhibit a mass/count variation in the lexicalization of their semantic properties, while others do not. This question has so far remained unanswered, or been set aside. We will do so by focusing on the role of context-sensitivity (already highlighted in recent theories of the mass/count distinction), and argue that it gives rise to a conflict between two pressures that influence the encoding of noun meanings as mass or count, one stemming from learnability constraints (*reliability*) and the other from constraints on informativeness (*individuation*). This will also lead us to identifying four semantic classes of nouns, and to showing why variation in mass/count encoding is, on our account, to be expected to occur widely in just two of them. Context-sensitivity forces a choice between prioritizing individuation, which aligns with count lexicalization, and prioritizing consistency, which aligns with mass lexicalization.

Keywords: Count/Mass · Probabilistic semantics · Mereology · Vagueness · Context-sensitivity

1 Introduction

The focus of this paper is on some of the most puzzling data in the domain of the mass/count distinction, which have so far seemed intractable or have been

First, we would like to thank the participants and organizers of the *Eleventh International Tbilisi Symposium on Language, Logic and Computation*. We especially enjoyed the inspirational and welcoming environment provided by our Georgian friends and colleagues. We are very grateful to Robin Cooper for his invaluable comments and suggestions during the preparation of this manuscript. We also wish to thank the audiences at *Type Theory and Lexical Semantics* held at *ESSLLI 2015* (Barcelona, Spain), *Sinn und Bedeutung 20* (Tübingen, Germany) and *Semantics and Philosophy in Europe 8* (Cambridge UK). Finally, we thank our colleagues and collaborators in the Department of Linguistics and the affiliated *Collaborative Research Center 991* at Heinrich Heine University, and in particular the participants of the *Semantics and Pragmatics Exchange* (*SemPrE*) colloquium. This research was funded by the German Research Foundation (DFG) CRC 991, Project C09.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 146–170, 2017.
DOI: 10.1007/978-3-662-54332-0_9

set aside in current accounts. In a substantial number of cases, we observe cross- and intralinguistic variation in the lexicalization of nouns as mass or count. As is well-known, but puzzling, in languages with a fully-developed grammaticized lexical mass/count distinction, things in the world like furniture, jewelry, hair, lentils fall under a count or a mass description, but cats are uniformly describable by basic lexical count nouns, while air or mud by mass nouns. The questions to ask are: 'Why do certain types of noun meanings exhibit a mass/count variation in the lexicalization of their semantic properties, while others do not?' 'Is this variation ad hoc, arbitrary, or is it due to some general principles that underlie the form-meaning mappings in the noun domain?' We will address these questions by developing two key ideas from recent work on the mass/count distinction that both highlight the importance of context-sensitivity. First, some nouns are context-sensitive in that what counts as minimal in a number neutral predicate's denotation varies with context. This form of context sensitivity is also associated with *vagueness* (Chierchia [2]). For example, what counts, minimally, as *rice* or *mud* is context dependent. A single grain of rice or a single fleck of mud is sufficient to count as *rice* or *mud* in some contexts, but not in others. Second, what counts as 'one' unit with respect to a given noun varies with context, as well (Rothstein [20], Landman [12]). For example, in some 'counting contexts' (Rothstein [20]), a teacup and saucer will count as one item of kitchenware, in other contexts as two items, and in other cases a teacup and saucer may "simultaneously in the same context" count as both one and two items of kitchenware (Landman [12]).

We argue that a more general level of explanation underlies the impact of context on countability, namely, that context-sensitivity of either of the two varieties just mentioned can be understood as giving rise to a conflict between two pressures on how languages encode the meanings of nouns, and which lead to predictions about when exactly variation in the mass/count lexicalization patterns is to be expected. One pressure, *reliability*, is derived from learnability constraints, and has to do with consistent criteria guiding the acquisition of noun meanings and their felicitous use in a variety of contexts. The second pressure, *individuation*, is derived from constraints on informativeness, which, for nouns, as we argue, amounts to the pressure to encode what counts as 'one' entity in their denotation. In other words, and put in the simplest terms, what is understood as individuation, as a prerequisite for counting, is here recast partly in information-theoretic terms.

The two varieties of context-sensitivity (one related to 'quantity vagueness' and the other to what counts as 'one') and the conflict they generate between the pressures coming from learnability constraints and constraints on informativeness, which impact on how languages encode the meanings of nouns, leads us to identifying four semantic classes of nouns. Most importantly, we motivate why only two of these, *granulars* and *collective artifacts and homogeneous objects* are systematically subject to the striking variation in the mass and count lexicalization, while the other two, *prototypical objects* and *substances, liquids, gasses*, are not. In brief, context-sensitivity forces a choice between prioritizing individuation, which aligns with count lexicalization, and prioritizing consistency, which aligns with mass lexicalization. As far as we know, no motivation of this kind has yet been provided.

We formally represent these ideas in a probabilistic mereological theory, *probabilistic mereological Type Theory with Records* (probM-TTR) which is an enrichment of probabilistic Type Theory with Records (Cooper et al. [5]). This theory has the advantage that it provides us with rich representational means to model the key ideas and processes that, as we argue, underly the mass/count distinction: namely, vagueness, counting-context sensitivity, overlap between entities that count as one, the impact of semantic learning on meaning representations, reliability of application criteria for nouns, and why, in some cases, multiple individuation criteria are licensed.

In Sect. 2, we summarize some of the leading recent contributions to the semantics of the mass/count distinction and we highlight and connect the role of context in each of them. In Sect. 3, we argue that the two notions of context sensitivity identified in Sect. 2 can be used to demarcate four semantic classes of nouns. We then argue for the presence of two competing pressures on natural language predicates that we call *reliability* and *individuation*. In Sect. 4, we briefly introduce our formal framework, probM-TTR. In Sects. 5 and 6, we show how the two types of context sensitivity from Sect. 2 give rise to conflicts between individuation and reliability, and so also give rise to the licensing of either mass or count encoding. We summarize these findings in Sect. 7.

2 Context-Sensitivity and the Mass/Count Distinction

2.1 Vagueness as a Variation in Extensions Across Contexts: Chierchia (2010)

Chierchia's [2] main claim is that mass nouns are vague in a way that count nouns are not. While count nouns have *stable atoms* in their denotation, that is, they have entities in their denotation that are atoms in every context, mass noun denotations lack *stable atoms*. If a noun lacks stable atoms, there is no entity that is an atom in the denotation of the predicate at all contexts. In this sense then, mass nouns have only *unstable individuals* in their denotation. But counting is counting of stable atoms only. Therefore, mass nouns are uncountable.

Chierchia enriches mereological semantics with a form of supervaluationism wherein vague nouns interpreted at ground contexts have extension gaps (vagueness bands). Contexts then play the role of classical completions of a partial model in other supervaluationist formalisms such that at every (total) context, a nominal predicate is a total function on the domain.

Contexts stand in a partial order to one another such that if c' precisifies c ($c \propto c'$), then the denotation of a predicate P at c and at a world w is a (possibly not proper) subset of P at c' and w. For an interpretation function F: $F(P)(c)(w) \subseteq F(P)(c')(w)$.

On Chierchia's supervaluationist account, mass nouns such as *rice* are vague in the following way. It is not the case that across all contexts, for example, a few grains, single grains, half grains, and rice dust always count as rice. Thus these quantities of rice are in the vagueness band of *rice*. There may be some total precisifications of the ground context c, in which single grains are rice atoms. There may also be some c' such that $c \propto c'$, where half grains are rice atoms. There may also be some c'' such that $c' \propto c''$, in which rice dust particles are rice atoms. Most importantly, there is, therefore, no entity that is a rice atom at every total precisification of rice. The denotation of *rice* lacks stable atoms, but counting is counting stable atoms, and so *rice* is mass.

2.2 Disjointness in Context

Rothstein [20] argues that neither formal atomicity (defined in mereological terms with reference to a Boolean lattice structure, presupposed by Chierchia [1]), nor natural atomicity (understood in terms of a "natural unit" in the sense of Krifka [10]) are sufficient or necessary to account for the differences between mass nouns and count nouns. A major contribution of Rothstein's work is to provide a formal model of how nouns such as *fence*, *wall*, which are not naturally atomic, nonetheless exhibit the hallmark grammatical properties of count nouns.

In contrast to Chierchia's use of context, Rothstein [20] coins the term "counting context", and defines count nouns as typally distinct from mass nouns. Mass nouns are of type $\langle e, t\rangle$. Count nouns are indexed via entity–context pairs and so are of type $\langle e \times k, t\rangle$. The following is Rothstein's example. Suppose that a square field is encircled by fencing. The answer to the question *How many fences encircle the field?* is wholly dependent on context. In some contexts, it would be natural to answer four (one for each side of the field). In other contexts, it would be more natural to answer one (one fence encircling the whole field). By indexing count nouns to contexts, Rothstein is able to capture how there can be one answer to the above question in any particular context (either one or four), despite fence lacking natural atoms, atoms that are independent of counting-context.

Rothstein's and Chierchia's context differ in their formal properties. On the assumption that we restrict our discussion to what Rothstein refers to as "default contexts" (relative to which the denotations of predicates are disjoint), Rothstein's contexts are not precisifications definable as a partial order. For example, let us again consider the field surrounded by fences a, b, c, d. Then at the context, k, at which a, b, c, d each individually counts as a single fence, their sum $a \cup b \cup c \cup d$ is excluded from the denotation of *fence*, while at the context k' at which $a \cup b \cup c \cup d$ jointly count as a single fence, a, b, c, d each taken individually are excluded from the denotation of *fence*. Clearly, therefore, one context does not precisify another.

There is, however, arguably a formal connection between the use of 'context' in Rothstein [20] and Chierchia [2]. Take the following quote from Chierchia, where, for ease of comparison, we added Rothstein's fence example to his mountain(s) example.

> "We must independently require (on anyone's theory) that for a concrete sortal noun N, its atoms are chosen so as not to overlap spatiotemporally. To put it differently, a disagreement over whether what you see in (43) is one or two mountains [ONE OR FOUR FENCES, IN OUR FIELD EXAMPLE ABOVE, Sutton & Filip] is, in the first place, a disagreement on how to resolve the contextual parameters. The key difference between nouns like *heap* or *mountain* [OR FENCE, Sutton & Filip] and mass nouns like *rice* is that minimal rice amounts, once contextually set, can still be viewed as units or aggregates without re-negotiating the ground rules." Chierchia [2] p. 123.

From this point of view, we could therefore, tentatively, associate the role of ground contexts (the contexts that set the "ground rules") in Chierchia [2] with the role of counting contexts in Rothstein [20]. Ground contexts, for Chierchia, set upper bounds on precisifications. This means that they set the positive extension for predicates. Formally speaking, ground contexts set the precisification g such that for all precisifications c, if $c \propto g$, then $c = g$.[1] In this sense we have two distinct notions of context. For counting one must, as per Rothstein's account, set a schema of individuation (via a counting context). However, as per Chierchia's account, there may still be ways to resolve the extension of a predicate across contexts of use that can undermine countability by obscuring what the individuals for counting are. In Sects. 5 and 6, we make these two types of contexts explicit in our formalism and analyze how they interact.[2]

2.3 Overlap in Context

In Landman [12] the set of generators, $\mathbf{gen}(X)$, of the regular set X is the set of semantic building blocks, which are either "the things that we would want to count as one" Landman [12, p. 26], relative to a context, or are contextually determined minimal parts. If the elements in the generator set are non-overlapping, as in the case of count nouns, then counting is sanctioned: Counting is counting of generators and there is only one way to count. However, if generators overlap, as in the case of mass nouns, counting goes wrong. One of Landman's innovations is to provide a new delimitation of the two cases when this happens, and hence two subcategories of mass nouns: mess mass nouns like mud, and neat mass nouns like furniture. A mass noun is *neat* if its intension at every world specifies a regular set whose set of minimal elements is non-overlapping. A noun is a mess mass noun if its intension at every world specifies a regular set whose set of minimal elements is overlapping.

[1] Thanks to a reviewer who pointed out this formulation.

[2] Our association of ground context and counting context is only tentative, however. They may, formally, operate in a similar manner, but there are clear informal differences. For example, Rothstein's counting contexts are meant to set the ground rules in the sense of determining what count as 'one'. Chierchia's ground contexts set the ground rules more in the sense of determining the boundary between the positive extension and the vagueness band.

For Landman, counting goes wrong when the variants of the generator set have different cardinalities simultaneously, but under different "perspectives" on one and the same set of entities. Variants of a set are maximally disjoint subsets of a set. For example, for the set $X = \{a, b, c, d, a \cup b, c \cup d\}$, there are four variants of X: $v_1 = \{a, b, c, d\}$, $v_2 = \{a, b, c \cup d\}$, $v_3 = \{a \cup b, c, d\}$, $v_4 = \{a \cup b, c \cup d\}$. Clearly, therefore, the effect of deriving variants of a set can be associated with the effect of applying a default counting context (from Rothstein [20]) to a predicate: every variant marks one way that an overlapping denotation could be made disjoint.

Context, although not a prominent part of Landman's account, is mentioned in relation to neat mass nouns. His paradigm example of a neat mass noun is *kitchenware*:

> "The teapot, the cup, the saucer, and the cup and saucer all count as kitchenware and can all count as one simultaneously in the same context. ... In other words: the denotations of *neat nouns* are sets in which the distinction between *singular individuals* and *plural individuals* is not properly articulated." Landman [12] pp. 34–35.

A striking idea here is that there are contexts which allow overlap in the denotation of a noun N with respect to what counts as one N. In other words, there are contexts in which, either one simply does not apply an individuation schema, or, alternatively, that the individuation schema one applies fails to resolve overlap. The former possibility is in effect a description of Rothstein's typal distinction between mass and count nouns wherein mass nouns are not indexed to counting contexts. However, equally possible is that, for some reason, one may choose a schema of individuation that fails to remove overlap. We will motivate this latter option in Sect. 3.

3 Count/Mass Variation, Reliability, and Individuation

3.1 Four Semantic Classes of Nouns and the Variation in Mass/Count Encoding

Considering just concrete nouns, as most do, we observe a considerable amount of puzzling data with respect to the variation in mass/count encoding between and within languages. This variation is not random, however. We may distinguish five classes of nouns depending on their main lexicalization patterns. They are summarized in Table 1 where the 'Noun Class' is a cover term for the descriptive labels below it. We then argue that these may be grouped into four classes in terms of the semantic properties given in Table 2.

The first striking pattern that we observe is markedly little variation in the mass/count encoding in two groups: namely, first, there is a strong tendency for *substances, gasses, liquids* to be encoded as mass (*mud*, *blood*, *air*), and second, a strong tendency for both animate and inanimate prototypical individuals

Table 1. Classes of nouns and mass/count variation

Noun class	Examples
Proto-typical objects	*chair*$_{+\text{C}}$; *tuoli*$_{+\text{C}}$ ('chair' Finnish); *Stuhl*$_{+\text{C}}$ ('chair' German) *dog*$_{+\text{C}}$; *koira*$_{+\text{C}}$ ('dog' Finnish); *Hund*$_{+\text{C}}$ ('dog' German) *boy*$_{+\text{C}}$; *poika*$_{+\text{C}}$ ('boy' Finnish); *Junge*$_{+\text{C}}$ ('boy' German)
Collective artifacts	*furniture*$_{-\text{C}}$; *huonekalu-t*$_{+\text{C},\text{PL}}$ ('furniture' Finnish) *meubel-s*$_{+\text{C},\text{PL}}$, *meubilair*$_{-\text{C}}$ ('furniture' Dutch) *kitchenware*$_{-\text{C}}$; *Küchengerät-e*$_{+\text{C},\text{PL}}$ (German, lit. kitchen device-s) *footwear*$_{-\text{C}}$; *jalkinee-t*$_{+\text{C},\text{PL}}$ ('footwear' Finnish)
Homogeneous objects	*fence*$_{+\text{C}}$, *fencing*$_{-\text{C}}$; *hedge*$_{+\text{C}}$, *hedging*$_{-\text{C}}$ *wall*$_{+\text{C}}$, *walling*$_{-\text{C}}$; *shrub*$_{+\text{C}}$, *shrubbery*$_{-\text{C}}$
Granulars	*lentil-s*$_{+\text{C},\text{PL}}$; *linse-n*$_{+\text{C},\text{PL}}$ ('lentils' German) *lešta*$_{-\text{C}}$ ('lentils' Bulgarian); *čočka*$_{-\text{C}}$ ('lentils' Czech) *oat-s*$_{+\text{C},\text{PL}}$; *oatmeal*$_{-\text{C}}$; *kaura*$_{-\text{C}}$ ('oats' Finnish); *kaurahiutale-et*$_{+\text{C},\text{PL}}$ (Finnish, lit. oat.flake-s)
Substances, liquids, gases	*mud*$_{-\text{C}}$; *muta*$_{-\text{C}}$ ('mud' Finnish); *Schlamm*$_{-\text{C}}$ ('mud' German) *blood*$_{-\text{C}}$; *veri*$_{-\text{C}}$ ('blood' Finnish); *Blut*$_{-\text{C}}$ ('blood' German) *air*$_{-\text{C}}$; *lenta*$_{-\text{C}}$ ('air' Finnish); *Luft*$_{-\text{C}}$ ('air' German)

Table 2. Interpretation of theories of the mass/count distinction

Noun class	Variation	Chierchia [2]	Landman [12]	Rothstein [20]
Prototypical objects	Little	Not vague	Not overlapping generators	Not context sensitive
Collective artifacts & homogeneous objects	Much	Not vague	Overlapping generators	Context sensitive
Granulars	Much	Vague	Not overlapping generators	Not context sensitive
Substances, liquids, gases	Little	Vague	Overlapping generators	(Context sensitive)

(*prototypical objects*) to be encoded as count (*cat, boy, chair*). The second striking pattern is a substantial amount of variation in the encoding of *collective artifacts* as mass/count *furniture*, *footwear*, *kitchenware*; *homogeneous objects* ('homogeneous' in the sense of Rothstein [20]) like *fence*, *wall*, *hedge*; *granulars* like *rice*, *lentils*. Such observations immediately prompt the question what is the reason why the mass/count variation is rife for some nouns, but scarce for others? What semantic facts or constraints allow us to make predictions when the mass/count variation is expected?

Let us first consider in more detail the groupings which display much mass/count variation (Table 2) and their attribution of properties, which are based on Rothstein [20], Chierchia [2], and Landman [12].

Prototypical objects: These nouns are not vague in the sense of Chierchia [2]. In Landman's [12] terms, they are count and so have non-overlapping minimal generators and non-overlapping generators. In Rothstein's [20] terms, these count nouns and as such indexed to counting contexts, and they have atoms in their denotations that do not vary across counting contexts. A dog, a chair, or a boy will count as one *dog*, *chair*, or *boy* by any reasonable schema of individuation.

Collective artifacts: These nouns contain typical cases of what Chierchia [2] calls "fake mass" nouns (following a long-standing tradition), and for which Landman [12] coins the term "neat mass" nouns: e.g., *furniture*, *footwear*, *kitchenware*. Chierchia takes the mass encoding of these nouns to be independent of vagueness, because they have stable atoms. Landman takes these nouns to have overlapping generators (but their minimal generators are non-overlapping). If it is the case that, from counting context to counting context, what counts as 'one *P*' varies, then, these nouns are counting context sensitive. Most importantly, they also have count counterparts, cross- and intralinguistically. Take *footwear* versus *jalkineet*$_{+\text{C,PL}}$ ('footwear', Finnish). On Landman's account, a shoe, and a pair of shoes can count as single items of footwear simultaneously in the same context. On Rothstein's account, being indexed to a (default) counting context would prohibit this, but it is not the case that Finnish must pick a single counting context. In some contexts a pair of shoes will count as one item of footwear. In another context, a pair of shoes will count as two items of footwear.

Homogeneous objects: Following Rothstein [20], we use "homogeneous" as a description of nouns such as *fence*, *wall*, *hedge*. The homogeneity is meant to capture that, at least for relatively large samples, a single stretch of fence or wall could be viewed, in another context, as two or more stretches of fence or wall. According to Chierchia (see Sect. 2.2), nouns such as *fence* and *wall* do not denote unstable entities relative to a ground context (e.g. relative to a counting context). On Rothstein's account, these are central cases of context-indexed nouns that are counting context sensitive. Most significantly, notice that these count nouns have mass counterparts (*fencing*, *walling*, *hedging*). As mass nouns, they presumably have, on Landman's account, overlapping generators. It seems reasonable to conclude that, for example, *fencing* denotes overlapping entities that can, simultaneously and in the same context, count as single items of *fencing*. Furthermore, Landman categorizes fencing as neat mass (p.c.). This would lead one to expect a felicitous cardinality comparison with, for example, *more than* constructions. However, native speakers are divided on the felicity of this reading. If this is an accurate description, then homogeneous objects pattern along with collective artifacts as not vague, overlapping, and counting context sensitive, hence the grouping of the two in Table 2.

Granulars: The denotations of granular nouns (*rice*, *lentils*) contain small grains. On Chierchia's [2] account, these nouns are vague, since no quantities of grains or parts of grains are stable atoms (in some contexts parts of grains would suffice, in other contexts, more than one grain may be required). Notably, those mass nouns in these categories often have cross-linguistic count-counterparts. On Landman's account, these count-counterparts should have non-overlapping generators. For example, the generators of *lentil* are presumably the individual lentils (they count as one).[3] However, it is hard to see how less than or more than a single lentil could equally count as one lentil, thus these granular count nouns arguably have non-overlapping minimal generators (they are *neat*). Similarly, for nouns such as *lentil*, it is hard to see how counting context could affect this individuation criteria. If single lentils count as one on one counting context for *lentil*, then, like nouns such as *cat*, they should count as one across all counting contexts. Although nouns such as *lentil* should be indexed to counting contexts on Rothstein's account, they are not counting context sensitive. Despite the mass encoding of granular nouns such as *rice*, we take similar considerations to apply.[4] Furthermore, reasons for thinking that nouns such as *rice* are neat, not mess, are given in [19].

Substances, gasses, liquids: These nouns are also vague on Chierchia's [2] account. On Landman's [12] account, such nouns are mess mass (because they have overlapping *minimal* generators). Insofar as these nouns are rarely encoded as count, it is hard to say whether or not they are counting context sensitive. However, in Yudja (Lima [14]), at least for nouns such as *mud* which do display count noun behavior, it seems that the quantities of mud that can count as one could vary from context to context (a pile in one context, a bucketful in another). Hence, we may tentatively conclude that *mud* is counting context sensitive (hence the parenthesis in Table 2).[5]

3.2 Two Competing Pressures: Reliability and Individuation

In the formal framework we propose in Sects. 4, 5, and 6, we will investigate a hypothesis that could account for much of the cross- and intralinguistic

[3] This is a vexed issue, however. Prima facie, *rice* and *lentil-s* should be treated similarly, however the mass noun *rice* should have overlapping generators, but the count noun *lentil* should have non-overlapping generators.

[4] Actually, this issue is also somewhat vexed. Nouns such as *lentil* cause problems for Landman [12] since, if subparts of lentils are not in the generator set and constitute proper parts of elements of the generator set, then they should not be in the denotation of *lentil(s)*, but this prediction is not accurate. This problem is remedied in Landman [13], where generators are replaced by "bases".

[5] There are also nouns which denote *fibrous* entities like *hair(s)*, *string(s)* which, on the one hand seem to pattern with granulars like *rice* insofar as they denote saliently perceptually distinguishable entities and are lexicalized as mass, but on the other hand, they also pattern with context-sensitive count nouns like *fence* insofar as what counts as one is contextually determined.

mass/count data. Our hypothesis supposes that there are (at least) two competing pressures on natural languages, one derived from learnability, the other from being a tool for effective communication. We take a cue for this proposal from work on information theoretic models of communication. For an example of how this type of approach of balancing learning and communicative pressures can be used to derive a theory of vagueness in information theoretic terms, see Sutton [17]. For a comparable approach applied to ambiguity, see [15].

Generally, there is an informational trade-off between being more informative and being learnable. For example, in the extreme case, a language could have one and only one predicate to describe all entities. This would be easily learnable, but maximally underdetermined, and so be a highly inefficient means of communication. At the other extreme, one could have a lexicalized classifier for every discernible property (e.g. a different lexical items for one N, two Ns, two big Ns etc.). Each of these classifiers would be highly informative, but would make languages unstable and unlearnable, since a learner would not receive sufficient instances of all classifiers to be able to infer their denotations (this is a form of the 'bottleneck' problem as discussed in the iterative learning paradigm [9]). Typically, classifiers convey an amount of information that in some way balances these pressures.

These general pressures are instantiated in the learning of concrete nominal predicates. On the one hand, there is a pressure for nominal predicate to be informative. In these cases, the amount of information conveyed is linked to how much of the domain is excluded by a classifier. Intuitively, a predicate which allows one to *individuate*, to pick out individual entities, is more informative than one which conveys no individuation schema, hence there is a general pressure towards establishing an individuation schema if this is possible given other perceptual and/or functional properties of the entities in the denotation of the relevant predicate. Individuation can be set in information theoretic terms. If the meaning of a noun (the signal) determines a specific criteria for counting, as opposed to something more ambiguous or vague, then the message will be more informative (carry a higher informational value). For example, if N_1 specifies $\{a, b, c\}$ as countable entities with some high probability (and so excludes sums thereof), but N_2 is has a level distribution between the sets $\{a, b, c\}$ and $\{a \cup b, a \cup c, b \cup c\}$, then in information theoretic terms, N_1 carries more information than N_2.

On the other hand, there is a pressure for learnability. One's criteria for classifying should be a reliable indicator of the correct way to apply the predicate, and consistently across various contexts. Call this pressure *reliability.* In the context of countability, if the individuation criteria sometimes correctly but sometimes wrongly excludes entities from the denotation of a noun, then it is unreliable. This very simple pressure, in effect requires that the probability of correctly applying a predicate, given the individuation schema is high.

As we will discuss in Sects. 5 and 6, these pressures may sometimes push in opposing directions. However, in the case of prototypical count nouns, *reliability* pushes in the same direction as *individuation.* There is a single and specific individuation schema for cat, namely being a cat individual (a single cat).

Furthermore, being a cat individual (or a sum thereof) is a very good indicator of being in the denotation of *cat(s)*.

4 Formal Framework

4.1 Type Theory with Records (TTR)

Type Theory with Records (Cooper [6], and references therein) is a richly typed formalism with a wide number of possible applications. In the following, we discuss only its application to natural language semantics and the representation of semantic structures as a form of compositional frame semantics (for discussion see Cooper [3,6]). In its application to natural language semantics, TTR is a system that combines insights from Fillmore's frame semantics [7] and situation theory, but also from formal semantics in the Montague tradition. In this section, we briefly introduce readers to the aspects of TTR that we will use in this article. Full formal details can be found in Cooper [6].

Two formal structures that are central to TTR are *records* and *record types*. Records are approximately *situations* from situation theoretic semantics, and record types are *situation types* from the same tradition, *frames* in the sense of Fillmore, but also what act as the TTR equivalent of *propositions*, namely, intensional structures that are made true by parts of the world, i.e. records/situations.

Record Types are represented as Field-Type matrices such as the one in (1) which details a highly simplified *cat*-frame.

$$\begin{bmatrix} x & : Ind \\ s_{cat} & : \langle \lambda v_{:Ind}(\text{cat}(v)), \langle x \rangle\rangle \end{bmatrix} \tag{1}$$

The fields (to the left of the colons) contain the labels x, s_{cat} will determine what values are provided by the record (situation) to which this frame is applied. For those more familiar with frameworks such as DRT, labels can also be thought of a approximating discourse referents. To the right of the colons are types. Ind is the basic type for individuals. In the spirit of semantics in the Frege-Montague tradition, predicates are functions. $\langle \lambda v.cat(x), \langle x \rangle\rangle$ is a predicate which is a function from entities of type Ind to a type of situation. It is important to note that predicates apply to vales for labels, not labels themselves. For example, if the value for x is felix, then this will yield a type of situation, cat(felix) in which Felix is a cat.

To form *properties* (the equivalent of expression of expressions of type $\langle e, t\rangle$), frames can be abstracted over to take a record as an argument. This is shown in (2) and provides a highly simplified representation of the English *cat*. What (2) requires is an application to a situation (record) which contains an individual. Now the type is restricted to take the value for the label x in the record ($r.x$), and apply it to the type statements in the record type/cat frame.

$$\lambda r : [x : Ind]. \begin{bmatrix} s_{cat} : \langle \lambda v_{:Ind}(\text{cat}(v)), \langle r.x \rangle\rangle \end{bmatrix} \tag{2}$$

Such a record could be the one given in (3). Records are finite sets of ordered pairs of labels and values. This is shown in matrix format in (3) where the label is x and the value is felix.

$$[x = \text{felix}] \tag{3}$$

For our purposes, felix could be thought of as the actual cat, and the label is just a way of tracking and accessing this object. Applying the record in (3) to the function in (2) yields a proposition: $[s_{cat} : \text{cat(felix)}]$ which will be true just in case there is a situation in which Felix is a cat. In other words, propositions are the equivalent of $\langle s, t\rangle$ expressions, except that TTR propositions are true of situations, which are partial and more cognitively plausible as truth makers than non-partial worlds (usually understood as sets of propositions).[6]

For this very brief introduction to TTR, another important point to note is the role of agents in the formalism. Agents can make a judgement that some object or situation a is of some type T (A judges that $a : T$). In an Austinian spirit, type judgements of this kind can be true or false. In Sect. 4.2 we will expand on how the notion of an agent's judgement set is linked to a probabilistic learning model (fully detailed in Cooper et al. [4,5]).

Finally, with respect to notation, we henceforth follow the standard brevity convention in TTR by simplifying how predicates are represented. Instead of $\langle \lambda v(P(v)), \langle x\rangle\rangle$ we will use just $P(x)$. For example, the frame in (1) will, under the convention, be represented as in (4).

$$\begin{bmatrix} x & : Ind \\ s_{cat} & : \text{cat}(x) \end{bmatrix} \tag{4}$$

4.2 Probabilistic Type Theory with Records (prob-TTR)

A full outline of prob-TTR may be found in Cooper et al. [4,5], we again introduce only that which will be necessary for our purposes. The central enrichment of TTR made in prob-TTR is to replace truth/falsity conditions of judgements with probability conditions. In later work, Cooper et al. [5] say that this is the probability of a judgement being the case, however, more in the spirit of the learning centric approach detailed in prob-TTR, we find that a more informative gloss on probability value for a judgement is the probability an agent ascribes to a competent speaker making that judgement (estimated with respect to her linguistic experiences and learning data).[7] Once integrated with a Bayesian learning model, probabilistic judgements are symbolized $p_{A,\mathfrak{J}}(a : T) = k$ or the probability that agent A judges that a is of type T with respect to her judgement set

[6] Another feature of TTR is that types are inherently intensional. This is because types are themselves viewed as objects to which other objects/situations belong, not merely as sets of objects/situations. As such, two distinct types may be coextensional.

[7] This formulation is due to Shalom Lappin p.c.

$\mathfrak{J}$ is $k \in [0, 1]$. Judgement sets record type judgements made of particular situations along with a probability value. Judgement sets are updated and form the basis for novel type judgements by the agent. The value k in (5) will represent the prior probability an agent A has for some individual being a cat, given her judgement set $\mathfrak{J}$. Conditional probabilities are then computed as in (6) using a type theoretic version of Bayes' Rule where $||T||_{\mathfrak{J}}$ is the sum of all probabilities associated with T in $\mathfrak{J}$.

$$p_{A,\mathfrak{J}}(s : \begin{bmatrix} x & : Ind \\ s_{cat} & : \text{cat}(x) \end{bmatrix}) = k \tag{5}$$

$$p_{A,\mathfrak{J}}(s : T_1 | s : T_2) = \frac{||T_1 \wedge T_2||_{\mathfrak{J}}}{||T_2||_{\mathfrak{J}}} \tag{6}$$

4.3 Probabilistic, Mereological Type Theory with Records (probM-TTR)

The simple enrichment we make to prob-TTR is to expand the domain of the basic type Ind from individuals to individuals and mereological sums thereof.[8] That is to say that we replace the basic type for individuals with the type of 'stuff' which we express as the basic type $^{*}Ind$. A learner's task will be to establish what, if anything, the individuals denoted by a particular predicate are. For example, given a world full of stuff, a learner of the predicate *cat* must learn which portions of stuff are individual cats. The type of individual for a predicate P will be represented Ind_P, so the type of single cat individuals will be Ind_{cat}.

Following Krifka [10,11], we distinguish between a qualitative and a quantitative criterion for applying nominal predicates.[9] Qualitative criteria may include perceptual properties such as color, shape, size and perceptual individuability (for example, grains of sand are harder to perceive and differentiate than grains of rice), but also functional properties. For simplicity, here we simply refer to this cluster of properties for a predicate P as the predicate P_{Qual}. This simple looking predicate should actually represent an entire frame that details, for example, functional and perceptual aspects of denotations relevant for forming predicate judgements. We will elaborate on the details of these frames in further work. This qualitative frame then acts as an argument for a 'quantitative' function $f_{P_{quant}} : (RecType \rightarrow NatNum)$. This is a function which outputs a natural number as a quantity value for some stuff with some combination of the relevant P qualities.

[8] This could equally be achieved using sets. For a set of formal atoms $\{a, b, c\}$, the domain of Ind entities would be $\{a, b, c, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

[9] As pointed out by a reviewer, a related concept is discussed by Geach [8]. However, Geach's criteria of identity is not identical with, for example, Krifka's *Natural Unit* function.

$$\begin{bmatrix} s_{p_{stuff}} & : \begin{bmatrix} x & : {}^*Ind \\ s_{p_{qual}} & : P_{Qual}(x) \end{bmatrix} \\ f_{p_{quant}} & : (\begin{bmatrix} x & : {}^*Ind \\ s_{p_{qual}} & : P_{Qual}(x) \end{bmatrix} \rightarrow \mathbb{N}) \\ i & : \mathbb{N} \\ s_{p_{quant}} & : f_{p_{quant}}(s_{p_{stuff}}) = i \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} s_{rice_{stuff}} & : \begin{bmatrix} x & : {}^*Ind \\ s_{rice_{qual}} & : rice_{Qual}(x) \end{bmatrix} \\ f_{rice_{quant}} & : (\begin{bmatrix} x & : {}^*Ind \\ s_{rice_{qual}} & : rice_{Qual}(x) \end{bmatrix} \rightarrow \mathbb{N}) \\ s_{rice_{quant}} & : f_{rice_{quant}}(s_{rice_{stuff}}) = 1 \end{bmatrix} \tag{8}$$

Examples of how we represent the qualitative frame and the quantitative function are given as a schema in (7) and for the predicate $rice(x)$ in (8). In both, the first field labels a type of situation in which some stuff has the relevant P-qualities/*rice*-qualities. The second field specifies a function from the quality record type to a natural number. The fourth field in (7) and the third field in (8) show the output to this function. In (8), this has been specified as 1. For this special case, this will be the type for single rice grains since the perceptually salient partition of rice is into grains. In this special case, we adopt an abbreviation convention in which (8) is rewritten as $[x : Ind_{rice}]$.

4.4 Prototypical Count Nouns

We can now specify the lexical entry for a concrete noun. Landman [12] specifies lexical entries as pairs of sets ⟨denotation, counting base⟩. We emulate this idea with frames and also adopt the terminology of Landman [13] of *body* for the regular denotation of a predicate, and *base* for the counting base. It should be emphasised, however, that the precise meaning of body and base for us differs from Landman's proposal. For a predicate such as $cat(x)$, we get:

$$\lambda r : [x : {}^*Ind]. \begin{bmatrix} s_{\text{body}} & : [\, s_{cat} : cat(r.x) \,] \\ s_{\text{base}} & : [\, r.x : Ind_{cat} \,] \end{bmatrix} \tag{9}$$

Entities of the type for the label s_{body} are in the denotation of the number neutral cat-property. Entities of the type for the label s_{base} provide the potentially countable entities for the number neutral property (the single cats).

This pair of types balances the pressures of individuation and reliability. Picking out single cats from the type of stuff is highly informative, since there is very little uncertainty as to which set of entities should be judged as cat individuals. The individuation schema provided by Ind_{cat} is also a highly reliable indication that one may apply the predicate $cat(x)$. If something is a cat individual in one context, it will rarely if ever be the case that one cannot apply the predicate $cat(x)$ to this individual across contexts. To see why the two pressure of individuation and reliability are both satisfied in this case, consider an alternative individuation schema that would be roughly as informative, for example, one which selected with a high probability, all cat pairs (every sum of two single cats). Unlike the good case, this schema would not be reliable, since it would, for example, incorrectly exclude single cats from being judged as cats.

In Sects. 5 and 6 we will consider two reasons when or why the type labelled s_{base} (the Ind_P type) is unavailable as a counting base for other nouns.

5 Counting-Context Sensitivity, Overlap, and Disjointness

In standard mereological approaches, overlap (not-disjoint) is a higher-order property of sets. Within our type theoretic paradigm, we will define it as a higher order type (a type of types). In the case of concrete nouns this will be a type of type of individuals. Other than this difference in approach, disjointness may be defined in a relatively standard way. However, one further added complexity is how the probabilistic aspect of our formalism interacts with the mereology. We introduce a (possibly context sensitive) probability threshold θ above which agents make judgements. A type is disjoint if all entities judged with sufficient certainty to be of that type are disjoint. For those types which have no clear instances, disjointness is undefined (one should not make a judgement either way with respect to disjointness). The intuitive idea here is that one cannot judge something to be disjoint or overlapping with respect to, say, a predicate, if one is not at all certain what falls under the predicate.

Definition 1. *A type T is disjoint relative to a probability threshold θ ($Disj_\theta$):*

IF	*there is at least some a such that $p(a : T) \geq \theta$,*
THEN	*$T : Disj_\theta$ iff, for all a, b such that $p(a : T) \geq \theta$ and $p(b : T) \geq \theta$,*
	if $a \neq b$, then $a \cap b = \varnothing$,
ELSE	*Undefined.*

We follow Landman [12] in making the grammatical counting function sensitive to disjointness. We also assume that the function applies to the type in a lexical entry labelled s_{base} (what is counted are the entities of the type in the counting base). Hence, for a counting function f_{count} and probability threshold θ, we propose a type restriction:

$$f_{count,\theta} : (RecType : Disj_\theta \rightarrow NatNum) \tag{10}$$

This type restriction means that the counting function is only defined for types that are disjoint (relative to some probability threshold).

For prototypical count nouns such as *cat*, *woman*, and *chair*, the types for the counting base are Ind_{cat}, Ind_{woman}, and Ind_{chair}, respectively. These are not are not overlapping. Thus they are defined for grammatical counting.

There are two classes of data that we need to explain, namely, the mass/count variation in *collective artifacts* and in *homogenous objects*. We do this by showing how context sensitivity with respect to individuation schemas results in a tension between the pressures of individuation and reliability.

5.1 Collective Artifacts

For mass nouns such as *furniture*, *kitchenware*, *fencing*, and count nouns such as *huonekalu-t* ('furniture', Finnish), *Küchengerät-e* ('kitchenware', German), and *fence*, the story is a little more complex. In Sutton and Filip [19] we suggested a treatment for neat mass nouns (*furniture*, *kitchenware*) and their count-counterparts. Here, using our more developed formal apparatus, we extend this analysis to context-sensitive semantically atomic nouns analyzed in Rothstein [20] (*fence*, *hedge*), and their mass-counterparts (*fencing*, *hedging*).

As we argued in Sects. 2.2 and 2.3, for both of these groups of nouns, the difference between mass and count encoding can be seen as involving either the non-resolution of overlap at a general context ('counting as one simultaneously and in the same context'), or as the resolution of overlap at a specific context. One aspect of Rothstein's [20] and Landman's [12] work that we suggested could be further developed is an account of what counting contexts are. Here we further develop the inchoate suggestion made in Sutton and Filip [19] that counting contexts can be modeled as schemas of individuation (formally modeled as quantitative functions). Furthermore, that, under pressure from individuation, variation in how we interact with the denotations of such nouns leads us to develop distinct individuation schemas (quantitative functions) and thereby distinct Ind_P types. We will give two examples: *furniture*$_{-\mathrm{C}}$ vs. *huonekalu-t*$_{+\mathrm{C}}$ ('furniture', Finnish), and *fencing*$_{-\mathrm{C}}$ vs. *fence*$_{+\mathrm{C}}$.

furniture$_{-\mathrm{C}}$ vs. *huonekalu-t*$_{+\mathrm{C}}$: Informally speaking, when learning what counts as 'one' with respect to *furniture* (or what counts as 'one' with respect to *huonekalu*), one is faced with inconsistent evidence. For example, vanity tables seem to be single items of furniture, but so do the framed mirrors that can be part of them. This creates a categorization problem, since BOTH the part and the whole should not be counted as one (even if both seemingly do count as one). This variation creates a conflict. A single individuation schema, represented as one quantitative function, would not be a reliable indicator of what counts as one item of furniture across contexts, since, for example, a single schema might correctly exclude counting the mirror in the vanity context, but incorrectly exclude counting such mirrors in other contexts. To remedy this, one must adopt different schemas in different contexts meaning that no one schema is wholly reliable. Hence, prioritizing the pressure towards individuation gives rise to unreliability.

To accommodate the pressure towards reliability, one could form a generalized individuation schema (formed from all admissible quantitative functions). This generalized schema would be a reliable indicator, since at every context, what counts as one would be included by at least one of the individuation schemas. In terms of the probabilistic semantics, this would mean that the conditional probability of correctly applying *furniture*, given the individuation schema would be very high. However, the generalized schema would no longer individuate since it would include as in counting as 'one' all entities that could count as one irrespective of whether they overlap (it would include the vanity table and the mirror that is a part of it). No longer individuating, in information theoretic terms, means carrying a lower informational value than a expression that transmits a single individuation schema, since the more general schema is equivocal between all admissible specific schemas. Hence, prioritizing the pressure towards reliability gives rise to less individuation.

For lexical items in this class, languages may, seemingly as a matter of convention, take one of two paths: prioritize individuation (at the expense of reliability), but allow the individuation schema to vary across situations; or prioritize reliability (at the expense of individuation), and form a generalized schema to cover all situations. We now formally outline how these two paths may be represented, then we show how the choice of path leads to a difference in mass/count encoding.

Formally speaking, for each noun where a clash of pressures arises, multiple quantitative functions are inferred by a learner. For example, with *furniture*, one function will map the type of situation which includes a vanity to the value 1 (the vanity as a whole counts as one). A different function which will map this same type of situation onto the value 2 (for the table and the mirror to be counted separately). In the later case, the same function would map the type of situation containing just the table (without mirror), or just the mirror (without table) to the value 1. Since our terminology $Ind_{furniture}$ is just shorthand for the type of situation where some entity receives a quantitative function value of 1, we can describe there being *two* functions in terms of an agent tracking two $Ind_{furniture}$ types. Call these $Ind_{furniture,1}$ and $Ind_{furniture,2}$. When more than one Ind_P type is being tracked, there are two strategies available for classifying individual P-items:

1. Prioritize individuation. For the case in hand, *furniture*, one could either apply only one type in any given instance. However, as noted above neither $Ind_{furniture,1}$ nor $Ind_{furniture,2}$ is reliable. To remedy this, one could make the choice of individuation schema context sensitive, namely to sometimes apply $Ind_{furniture,1}$ and sometimes applying $Ind_{furniture,2}$.
2. Prioritize reliability. To do this one need merely form a more generalized type to cover all cases. This would obviate the need to add in context sensitivity. In TTR, a more generalized type can be formed via a disjunction (or join) between types as shown in (11).

$$Ind_{P,gen} = Ind_{P,1} \vee Ind_{P,2} \vee ... \vee\ Ind_{P,n} \tag{11}$$

However, now the generalized schema does not fully individuate since it equivocates between whether a sum counts as one or more than one item of furniture.

The availability of a 'choice' of which pressure to prioritize explains mass/count variation via a difference in lexical entries for mass nouns such as *furniture* (12) as opposed to cross linguistic count-counterparts such as the Finnish *huonekalu* ('item of furniture') (13).

$$[\![\text{furniture}]\!] = \lambda r : \left[\, x : {}^{*}Ind \,\right] . \begin{bmatrix} s_{\text{body}} : \left[\, s_{furn} : furn(r.x) \,\right] \\ s_{\text{base}} : \left[\, r.x : Ind_{furn,gen} \,\right] \end{bmatrix} \tag{12}$$

$$[\![\text{huonekalu}]\!] = \lambda r : \left[\, x : {}^{*}Ind \,\right] . \begin{bmatrix} s_{\text{body}} : \left[\, s_{furn} : furn(r.x) \,\right] \\ s_{\text{base}} : \left[\, r.x : Ind_{furn,i} \,\right] \end{bmatrix} \tag{13}$$

The reason these entries lead to the mass encoding of *furniture*, but the count encoding of *huonekalu* is due to the semantic qualities of the type for the label s_{base} in each case. In (12), the type $Ind_{furn,gen}$ is not disjoint. This is because, for example, both a dressing table (including mirror) and a dressing table (excluding mirror) will be of this type. Other examples of overlap include tables that are pushed together (are they one or many tables?), and chairs with cushions (should the chairs be counted separately from the cushions or together?). Non-disjoint types are not defined for the counting function (10), and so *furniture* is mass. In contrast, because, in (13), *huonekalu* is encoded to select a specific quantitative function (determined, for example, by the context of use), each type $Ind_{furn,i}$ is disjoint. As such, *huonekalu* will be defined for counting. That said, from context to context, the counting result may vary. In some contexts, the dressing table (including mirror) will count as one *huonekalu*, in others it may count as two.

This pattern in which counting results may differ from context to context should sound familiar from the case of *fence*. Recall Rothstein's example of a square field enclosed by fencing. Whether we count this as one fence around the field, or two, three or four may depend on the context. We are able to use exactly the same tools as we use for *furniture* versus *huonekalu* to model this. The entry for *fence* is given in (14) and the entry for *fencing* is given in (15).

$$[\![\text{fence}]\!] = \lambda r : \left[\, x : {}^{*}Ind \,\right] . \begin{bmatrix} s_{\text{body}} : \left[\, s_{fence} : fence(r.x) \,\right] \\ s_{\text{base}} : \left[\, r.x : Ind_{fence,i} \,\right] \end{bmatrix} \tag{14}$$

$$[\![\text{fencing}]\!] = \lambda r : \left[\, x : {}^{*}Ind \,\right] . \begin{bmatrix} s_{\text{body}} : \left[\, s_{fence} : fence(r.x) \,\right] \\ s_{\text{base}} : \left[\, r.x : Ind_{fence,gen} \,\right] \end{bmatrix} \tag{15}$$

The reason these entries lead to the count encoding of *fence*, and the mass encoding of *fencing* parallels that of the previous case. Given that, at any context, the entry for *fence* selects a single quantitative function, the type $Ind_{fence,i}$ is disjoint, and so defined for counting, even if the exact result of counting the same portion of fencing may result in different answers across contexts. In contrast, *fencing* does not distinguish between contexts and is defined in terms of more generalized join type $Ind_{fence,gen}$ that is not disjoint. The reason that it is not disjoint is that, for example, in Rothstein's square field example, the sum of four fence sides is of type $Ind_{fence,gen}$, but so are the four fence-sides taken individually. Non-disjoint types are undefined for countability, and so *fencing* is mass.

Furthermore, these different conceptions are driven by which pressure is prioritized. If one prioritizes individuation, then the pressure is to find a single counting schema (at least in a context) from the possible schemas. However, in order to be reliable, the schema one uses must be context sensitive. This means that at each context, one has a non-equivocating individuation schema from the set of possible schemas. Choosing a single one (at a context) is maximally informative, thus the pressure of individuation is satisfied. On the other hand, one can prioritize reliability and adopt a generalized schema that ($Ind_{fence,gen}$) that is a reliable indicator of when to apply the number neutral predicate $fence$. However, this generalized schema does not fully satisfy the pressure of individuation since it equivocates between specific schemas.

In this section we have argued that counting-context sensitivity gives rise to a competition between the pressures of individuation and reliability. Prioritizing one of these pressures over the other seems to be a matter of convention. Prioritizing individuation yields count encoding. Prioritizing reliability yields mass encoding. With this form of context-sensitivity, we cannot yet explain count/mass variation in granular nouns such as *lentil*, *rice* which we have assumed have disjoint Ind_P types (the types for single rice grains and single lentils). Nor can we, at this point, say anything about substance mass nouns such as *mud* and *air*. For this, we will need to appeal to another form of context-sensitivity, one related to vagueness. In Sect. 6, we will argue that vagueness can also lead to a clash between the pressures of individuation and reliability and so also to variation in mass/count encoding.

6 Contextual Variation and Vagueness

The conception of vagueness we adopt is based loosely on Sutton ([17,18]). On this conception, vagueness is represented as a form of metalinguistic uncertainty that arises, in part, from inconsistent learning data. For example, for color predicates, we have good evidence for judging canonical cases of green as 'green', and likewise for blue. Towards the blurred boundary between green and blue, we either have a dearth of evidence for making 'blue'/'green' judgements, or we have conflicting information (sometimes a shade will be described as 'blue', sometimes not). Either way, we infer a distribution that describes a gradual trailing off of the probability of a competent speaker making a 'blue' judgement as the shade of the object in question becomes ever greener, *mutatis mutandis* for 'green'.

Following Chierchia ([2]) we argue that a similar mechanism affects the semantic representations of some nouns, however, that the graded increase in uncertainty varies with the output of the quantitative function. This mechanism is again a form of context sensitivity. The variation in what counts as, for example, *rice*, across contexts yields metalinguistic uncertainty (vagueness) with respect to what quantity of rice-stuff is sufficient to classify that stuff as *rice*.

6.1 Granular Nouns

The context-sensitivity of granular and substance nouns differs from that of collective nouns such as *kitchenware* and *furniture*. As Chierchia [2] observes, our judgements about whether granular and other substances are in the denotation of a given predicate vary depending on their amount in a given context. For example, whether we are willing to accept that we have mud on our shoes varies with context. In clean-room manufacturing or scientific contexts, even small specks of mud count as *mud*, because the tolerance for even tiny quantities of mud is near zero. In contexts like entering the apartment after a walk, our tolerance for mud is much higher, and in contexts like entering the garden shed it is even higher. For nouns such as *rice* or *lentils* one could truly say that we do not have any rice/lentils for dinner when only a few grains/lentils remain in the packet, but equally truly say that some rice/lentils fell on the floor during a meal even though the number of grains/lentils may be identical in both cases. Context matters. However, from a probabilistic learning perspective, these cases provide inconsistent data with respect to the categorical application of classifiers such as *mud*, *rice*, and *lentils*. The rational response for a learner (aside from seeking aspects of the contexts to explain this variation) is to lower the confidence with which she would apply the predicate for the specific amount of mud/rice/lentils in question. We model this as a Bayesian update given the judgement set. The judgement set consists of situations (which can be understood as contexts from a situation theoretic point of view) and probabilistic type judgements made about those situations (contexts). In other words, the agent calculates the probability of applying e.g. the *rice* conditional with respect to the context with some quantity of stuff with the appropriate rice qualities. This is represented in (16) for some quantity value of 10. The value 0.5 would reflect the borderline case where the agent has as much reason to classify some quantity (of grains) of rice as *rice* as she has reason to judge them not to be rice.

$$p_{A,\mathfrak{J}}(r:\begin{bmatrix} x & : {}^*Ind \\ s_{rice} & : rice(x) \end{bmatrix} \mid r: \begin{bmatrix} s_{rice_{stuff}} & : \begin{bmatrix} x & : {}^*Ind \\ s_{rice_{qual}} & : rice_{Qual}(x) \end{bmatrix} \\ f_{rice_{quant}} & : (\begin{bmatrix} x & : {}^*Ind \\ s_{rice_{qual}} & : rice_{Qual}(x) \end{bmatrix} \to \mathbb{N}) \\ i & : \mathbb{N} \\ s_{rice_{quant}} & : f_{rice_{quant}}(s_{rice_{stuff}}) = 10 \end{bmatrix}) = 0.5 \quad (16)$$

For nouns such as *rice*, numerical values need not be taken to align perfectly with numbers of grains. For higher values, the output of the function could just as easily indicate some range of numbers of grains as some specific number. Either way, uncertainty about whether to apply the rice predicate will increase with smaller quantitative function values. This means a gradual increase of uncertainty about applying the predicate as quantities of rice get smaller. The idea that this represents is simply that one is safer, across contexts, using *rice* to describe larger quantities (a bowlful, a whole packet) than much smaller quantities (a grain, a few grains). The uncertainty involved in using the predicate across these cases reflects this.

Unlike with nouns such as *furniture* and *kitchenware* as well as with *fence* and *fencing*, this uncertainty is not about what counts as one (leading to a proliferation in individuation functions), but uncertainty about how much rice is enough to safely form a *rice* judgement. Yet, similarly to the *furniture*, *kitchenware*, *fence* and *fencing* cases, mass/count encoding of granular nouns can be seen as arising from the competition between the pressures of reliability and individuation.

The pressure of individuation pushes in one direction, namely that, for nouns such as *rice* and *lentils*, the types for the counting bases of the nouns should be the types Ind_{rice} and Ind_{lentil}, respectively. For *furniture*- and *fence*-like nouns, there were multiple competing equally informative individuation schemas (e.g. one which counts the table and mirror as two and another schema that counts the table and mirror as one, a vanity). However, for granular-like nouns, there is really only one plausible individuation schema, namely, that which counts grains, flakes etc.[10] However, the gradation in probability values in the representation of nouns such as *rice*, and *lentils* means that, types for lower quantity values such as 1 (represented as types Ind_{rice}, Ind_{lentil}) are not *reliable* indicators of when to apply *rice* or *lentils*. In other words, prioritizing individuation leads to a fall in reliability. This is because single grains of rice or single lentils will not qualify as *rice* or *lentils*, respectively, *reliably* in all contexts. A strategy of prioritizing individuation will simply enter the Ind_P type as the counting base. The lexical entry for granular nouns could resemble far more closely the one for *cat* in (9). This is what we suggest occurs for nouns such as the English *lentil* as in (19). Individuation is prioritized since types such as Ind_{lentil} are disjoint, but reliability is forfeit since this type is not a wholly reliable indication of when one may apply the predicate *lentil(x)*.

The pressure of reliability pushes in the opposite direction to the pressure of individuation for granular nouns. Prioritizing reliability militates against taking, for example, the type for single grains of rice (Ind_{rice}) or single lentils ($Ind_{lentils}$) as a counting base. Recall that reliability entails finding a counting base such that the probability of (correctly) applying a predicate is high given that some entity is of that type specified in the base. One way to boost this probability and so prioritize reliability, as we find with the English *rice*, is to lexically encoding the counting base not with the type Ind_{rice}, but with the less specific predicate *rice* as in (17). Reliability is maximized here since, trivially, $p_{A,\mathfrak{J}}(a : T|a : T) = 1$, and so the type labelled s_{base} in (17) is a perfect predictor of the type labelled s_{body}. On this strategy, individuation is forfeit, since those entities which perceptually saliently count as one (such as individual rice grains), are not clear cases of the predicate *rice* across contexts.

In summary, for nouns such as *rice* and *lentils*, the context sensitivity that gives rise to graded probability judgements for entities in terms of applying a predicate, given some qualitative properties *and* a quantitative function value, in turn, creates a conflict between the pressures of individuation and consistency. The result is to prioritize one pressure. If one prioritizes reliability, the base does not individuate. Examples are given in (17) for the English *rice* and in (18) for the Bulgarian mass noun *lešta* ('lentil'). If one prioritizes individuation, the base is simply the relevant Ind_P type. An example of this is given in (19) for the English 'lentil'.

[10] This may not be universally true. For example, grains that come in easily separable halves might have two viable schemas, one which counts halves and one which counts wholes.

$$[\![\text{rice}]\!] = \lambda r : \left[x : {}^*Ind \right] . \begin{bmatrix} s_{\text{body}} : \left[s_{rice} : rice(r.x) \right] \\ s_{\text{base}} \ : \left[s_{rice} : rice(r.x) \right] \end{bmatrix} \tag{17}$$

$$[\![\text{lešta}]\!] = \lambda \text{r} : \left[x : {}^*Ind \right] . \begin{bmatrix} s_{\text{body}} : \left[s_{lentil} : lentil(r.x) \right] \\ s_{\text{base}} \ : \left[r.x : lentil(r.x) \right] \end{bmatrix} \tag{18}$$

$$[\![\text{lentil}]\!] = \lambda r : \left[x : {}^*Ind \right] . \begin{bmatrix} s_{\text{body}} : \left[s_{lentil} : lentil(r.x) \right] \\ s_{\text{base}} \ : \left[r.x : Ind_{\text{lentil}} \right] \end{bmatrix} \tag{19}$$

The difference between (17) and (18) on the one hand and (19) on the other is in the type for the label s_{base}. In (19), the type Ind_{lentil} is a disjoint type and so is suitable for counting. Hence *lentil* is count. In (17), the type for the labels s_{body} and s_{base} are the same. Depending on the probability threshold, this type contains parts of grains, grains, or collections of grains of rice and sums thereof. As such, the type labelled s_{base} is not disjoint, and so is not defined for grammatical counting.

6.2 Substance Nouns

As we stated above, substance nouns like *mud* are vague in the same way as granular nouns in that what counts as *mud* varies from context to context, thus generating an inconsistent set of evidence for what counts as *mud*. We may assume, therefore, that the same ways of balancing the pressures of reliability and individuation that we employed for vague granular nouns like *rice* and *lentil* could be adopted for substance nouns, namely one of the two entries (20) or (21).

$$[\![\text{mud}]\!] = \lambda r : \left[x : {}^*Ind \right] . \begin{bmatrix} s_{\text{body}} : \left[s_{mud} : mud(r.x) \right] \\ s_{\text{base}} \ : \left[r.x : Ind_{\text{mud}} \right] \end{bmatrix} \tag{20}$$

$$[\![\text{mud}]\!] = \lambda r : \left[x : {}^*Ind \right] . \begin{bmatrix} s_{\text{body}} : \left[s_{mud} : mud(r.x) \right] \\ s_{\text{base}} \ : \left[s_{mud} : mud(r.x) \right] \end{bmatrix} \tag{21}$$

Prioritizing reliability yields the entry in (21) which would lead to the mass encoding of *mud* for the same reason as we got a mass encoding for *rice* in English. The type for the label s_{base} is not disjoint.

In contrast to *lentil*, however, the entry in (20) will not yield count encoding. For object count nouns, collective artifacts, and granular nouns (where the granules are not too small) there is relatively clear perceptual and/or functional based evidence for establishing what counts as 'one' item in the denotation of the relevant noun. In probM-TTR terms that means that for such a predicate P, there are at least some objects a, such that an agent is able to judge that $a : Ind_P$ with a reasonably high probability. This is not the case for substance, liquid and gas nouns. Unlike nouns like *cat* and *rice*, the denotations of these nouns are such that there is little, perceptually speaking, to aid in the identification of salient individuated units. Unlike nouns such as *chair* and *furniture*, nor do the denotations of substance nouns typically get partitioned in terms of function. This distinction in itself can be viewed as a further form of vagueness: what the perceptually/functionally salient entities of substance noun denotations are is highly uncertain.

In terms of reliability and individuation, this, in contrast to the granular case, means that types such as Ind_{mud} fail to carry a sufficiently high informational content (fail to specify a sufficiently specific portion of mud such that portion would count as *one unit of mud*). Furthermore, unless a language imports a significant amount of context-sensitivity in what counts as an individuated mud unit (as could be argued is the case in languages such as Yudja), the pressure of individuation cannot be satisfied. We therefore would expect (21) and not (20) to be the lexical entry for *mud*. Put another way, unless made radically dependent on the context of application, the type Ind_{mud} is simply not useful since it is neither a good indicator for the applicability of *mud* (not consistent), nor does it convey a high enough informational content (does not individuate).

In probM-TTR terms that means that for such a substance/liquid predicate P, there are no objects a, such that an agent is able to judge that $a : Ind_P$ with a high probability. With respect to the disjointness (Definition 1), this means that types such as Ind_{mud} are undefined for disjointness. Since the counting function requires a disjoint type as input, this means that substance nouns such as *mud* will be encoded as mass, even if their lexical entries are of a similar form to (20).

7 Conclusions and Summary

We hypothesized that there are two competing pressures on natural language predicates: (i) to individuate (recast partly in information-theoretic terms as being informationally rich); (ii) to find a reliable criterion for counting (a criterion which reliably predicts the type for the whole extension of P, modelled as a high conditional probability that something is of the body type, given that it is of the base type).

Inductive evidence for this hypothesis is provided by the predictions it makes with respect to the variation in the mass/count encoding. We show that the ways in which these two pressures can (or cannot) be satisfied in dependence on the different types of context-sensitivity represented in our formal model, predict the expected range of constraints on the variation in the mass/count encoding. In addition, this allows us to cover a broader range of data than other leading accounts.

Prototypical object nouns: The types that pick out the individuable entities in the denotations of prototypical object nouns are also highly consistent indicators of when to apply the nouns. The pressures on individuation and reliability work in the same direction, i.e., they converge on the count encoding. We, therefore, have no reason to expect much, if any, mass encoding, cross- and intralinguistically.[11]

[11] One possible counter example to this is Brazilian Portuguese which seems to encode mass readings of most or even all object count nouns when used in the bare singular. For example, the bare singular 'How much book...?' can get a non-coerced measure (weight) reading. See [16].

Collective and homogeneous object nouns: Context-sensitivity with these nouns affects the reliability with which individual types apply. For example, across contexts, a sum can count as one fence, one item of kitchenware or two fences, two items of kitchenware. This means that any particular individuation schema will inconsistently determine the extension. To prioritize individuation, multiple individuation schemas, each indexed to a context, can be used. This yields count nouns such as *fence*, and *Küchengeräte* ('kitchenware' German). Alternatively, to prioritize reliability, all individuation schemas can be merged together. This yields a non-disjoint schema and so mass nouns such as *fencing* and *kitchenware*.

Granular nouns: Context-sensitivity with granular noun denotations has an effect on what quantities of the relevant stuff are needed to qualify for that stuff to fall under a given noun denotation. Granular nouns tend to be easily perceptually individuable (in terms of salient individual grains), but given that single grains are not always enough to qualify as falling under a given noun denotation across all contexts, the type for single grains, that prioritizes individuation, is inconsistent as a basis for applying a noun. Prioritizing individuation yields a count noun encoding, which is commonly presupposed by pluralization, e.g. *lentils*, *kaurahiutale-et* ('oatmeal' Finnish), *oats*. On the other hand, prioritizing reliability yields a non-disjoint individuation schema, and so leads to a mass noun encoding, as in *oatmeal*, *kaura* ('oats', Finnish), *čočka* ('lentils', Czech).

Substance nouns: Similarly as with granular noun denotations, context-sensitivity has an effect on amounts of quantities (e.g., of substances, liquids, and gases) reaching a certain threshold to qualify as falling under a given noun (e.g., *mud*, *blood*, and *air*). However, the perceptual qualities of the denotations of these nouns do not easily enable the prioritization of individuation that could be achieved for count granular nouns.[12] If individuation cannot be prioritized, then reliability will be prioritized, therefore, we expect a heavy tendency towards mass encoding for these nouns.

Our formal account can capture these competing pressures either in terms of how sharply and specifically (as opposed to generally and vaguely) types relate to entities in the world. Our link to learning models also allows us to describe how (un)reliability can arise out of a process of classifier learning. Together, this means that we are not only able to formally represent noun meanings and countability, but we have also outlined the general mechanisms that give rise to the variation in the mass/count encoding.

References

1. Chierchia, G.: Plurality of nouns and the notion of semantic parameter. In: Rothstein, S. (ed.) Events and Grammar, vol. 70, pp. 53–103. Springer, Heidelberg (1998)
2. Chierchia, G.: Mass nouns, vagueness and semantic variation. Synthese **174**, 99–149 (2010)

[12] However, see the caveat about Yudja in Sect. 6.2.

3. Cooper, R.: Frames in formal semantics. In: Loftsson, H., Rögnvaldsson, E., Helgadóttir, S. (eds.) NLP 2010. LNCS, vol. 6233, pp. 103–114. Springer, Heidelberg (2010). doi:10.1007/978-3-642-14770-8_13
4. Cooper, R., Dobnik, S., Lappin, S., Larsson, S.: A probabilistic rich type theory for semantic interpretation. In: Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics, pp. 72–79 (2014)
5. Cooper, R., Dobnik, S., Larsson, S., Lappin, S.: Probabilistic type theory and natural language semantics. LILT **10**(4), 1–43 (2015)
6. Cooper, R.: Type theory and semantics in flux. In: Kempson, R., Fernando, T., Asher, N. (eds.) Philosophy of Linguistics, pp. 271–323. Elsevier, Amsterdam (2012)
7. Fillmore, C.J.: Frame semantics and the nature of language. Ann. N. Y. Acad. Sci. **280**(1), 20–32 (1976)
8. Geach, P.T.: Reference and Generality: An Examination of Some Medieval and Modern Theories. Cornell University Press, Ithaca (1962)
9. Kirby, S., Hurford, J.: The emergence of linguistic structure: an overview of the iterated learning model. In: Laurea, A.C., Laurea, D.P. (eds.) Simulating the Evolution of Language, pp. 121–148. Springer, London (2002)
10. Krifka, M.: Nominal reference, temporal constitution and quantification in event semantics. In: Bartsch, R., van Benthem, J.F.A.K., van Emde Boas, P. (eds.) Semantics and Contextual Expression, pp. 75–115. Foris Publications, Providence (1989)
11. Krifka, M.: Common nouns: a contrastive analysis of English and Chinese. In: Carlson, G., Pelletier, F.J. (eds.) The Generic Book, pp. 398–411. Chicago University Press, Chicago (1995)
12. Landman, F.: Count nouns - mass nouns - neat nouns - mess nouns. Baltic Int. Yearb. Cogn. **6**, 1–67 (2011)
13. Landman, F.: Iceberg semantics for count nouns and mass nouns: the evidence from portions (2015, unpublished handout)
14. Lima, S.: All notional mass nouns are count nouns in Yudja. Proc. SALT **24**, 534–554 (2014)
15. Piantadosi, S.T., Tily, H., Gibson, E.: The communicative function of ambiguity in language. Cognition **122**, 280–291 (2011)
16. de Oliveira, R.P., Rothstein, S.: Bare singular noun phrases are mass in Brazilian Portuguese. Lingua **121**, 2153–2175 (2011)
17. Sutton, P.R.: Vagueness, communication, and semantic information. Ph.D. thesis. King's College London (2013)
18. Sutton, P.R.: Towards a semantics for vague adjectives. In: Zeevat, H., Schmitz, H.-C. (eds.) Bayesian Natural Language Semantics and Pragmatics, vol. 2, pp. 221–246. Springer, Heidelberg (2015)
19. Sutton, P.R., Filip, H.: Vagueness, overlap, and countability. In: Proceedings of Sinn und Bedeutung 20 (2016, forthcoming)
20. Rothstein, S.: Counting and the mass/count distinction. J. Semant. **27**, 343–397 (2010)

Semantic Dependency Graphs

Henk Zeevat(✉)

Institute for Logic, Language and Computation,
University of Amsterdam and SFB991 Heinrich-Heine-Universität,
Düsseldorf, Germany
henk.zeevat@uva.nl

Abstract. Dependency Grammar has been taken as a formalism for syntactic representation, comparable to close competitors such as phrase structure grammar or categorial grammar. This paper argues that in fact the dependency graphs (DGs) should—like semantic frames—be seen as a semantic formalism like e.g. FOL, Montague's IL or Discourse Representation Structures. For this, arrows must have semantically interpretable labels and two additional kinds of arrows need to be added: scope arrows and anaphoric arrows.

1 Introduction

Dependency Grammar (Tesnière 1959; Baum 1976; Hudson 1984, 2007) has been taken too often as a formalism for syntactic representation, comparable to close competitors such as phrase structure grammar or categorial grammar. This paper argues that in fact the dependency graphs (DGs) should, like semantic frames (Barsalou 1992; Löbner 2014; Löbner 2015; Petersen 2007), be seen as a semantic formalism like e.g. FOL, Montague's Intensional Logic (Montague 1973) or Discourse Representation Structures (Kamp and Reyle 1993). The view is not neutral with respect to the dependency graphs that must be employed. Arrows must have semantically interpretable labels and two additional kinds of arrows need to be added, scope arrows and anaphoric arrows. The extra arrows break a standard property of dependency graphs, namely having at most one incoming arrow per node.

In Zeevat (2014), a dependency parser is defined for semantic parsing, i.e. for the task of linking arguments to their predicates, of modifiers to what they modify, of pronouns to their antecedents, of quantifiers to their restrictors and scopes. The result is a special kind of dependency structure which is then transformed into a variable-free prolog notation, where references to objects are always made by the term that originally introduced them. This paper explores the alternative strategy of directly defining satisfaction and update on the dependency structures in a way that generalises to other ways of obtaining dependency structures.

The paper however makes no assumptions about how one should arrive at these structures, but takes as an obvious starting point that recent advances in deriving dependency structure within computational linguistics[1] make it more

[1] Since Lin (1998) parser correctness is measured on the correctness of the derived DGs, leading to many algorithms that map trees to dependency structures.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 171–184, 2017.
DOI: 10.1007/978-3-662-54332-0_10

and more plausible that one can arrive at such structures using a variety of methods and knowledge sources.

The typical dependency graphs obtained in these enterprises (Lin 1998; Nivre and Scholz 2004) fall short of what one needs for semantic interpretation, but can be seen as underdetermined semantic structures, in which the arrows stand in need of further labeling or to which further arrows need to be added.

Formally, a dependency graph $< N, LA >$ is a set of labeled nodes $n =< i, l > \in N$ where l is drawn from a lexicon L and i is an index[2] and labeled arrows LA of the form $< n_1, n_2, l >$ with n_1 and n_2 drawn from N and l a label drawn from a different set of labels AL. $< N, LA >$ is proper iff the nodes are connected by arrows and the arrows connect nodes and there is a single root node r that has no incoming arrows and all other nodes can be reached from r by following arrows. A node k can be reached in this sense if there is a number $n > 0$ such that there is a node j that can be reached in $n - 1$ steps from r and an arrow $< j, k, v >$.

Having a root is a feature of dependency grammar that does not seem to fit well with the kind of semantic graphs that are needed for lexical semantics. Petersen (2007) uses the example of the noun *father* whose intuitive root note should paradoxically refer to a person with an incoming *father*-arrow from a child of that person. Such problems do not arise in syntax. The assumption of roots is harmless however since in the contexts the more complex semantic frames—like the ones used by Petersen—can be reconstructed as combinations of rooted DGs.

The lexicon is divided into heads, operators and marks. The arrow labels (AL) can be technical (*scope, restrictor, anap* or marking *mark*) or be the names of functional relations between objects in the domain. The heads are lexical words, operators functional words like the negation *not* or the quantifier *each*, while marks are a remainder category containing articles, verbal markings like *to* and others.

Scope arrows run from heads to operators. Mark arrows do not run to heads but to mark labels like *the, to* or *ing* and originate in heads. *Anap*-arrows link heads with heads. Non-technical and non-marking arrows also connect heads with heads.

A sequence of dependency graphs is a context. Contexts optionally include a set of *anap*-labeled arrows $< n, m, anap >$ where n and m are nodes from different DGs and n is taken from a DG that is later in the sequence than the DG in which m occurs.[3] There are arguments for also allowing other arrows in the contexts. In (1) *he* should be connected by a contextual *anap*-arrow to its antecedent *John*.

[2] Needed to keep nodes with the same labels separate. This is not necessary when one thinks of DGs as graphical objects where such nodes can be distinguished by their spatial position. The assumption made is that different nodes have different indices, even in contexts.

[3] An alternative is to make contexts into graphs. Then these *anap*-arrows are normal graph arrows. The current presentation seems marginally more perspicuous, since it unburdens the graph notation by some set-theoretic notation.

(1) John came in. He smiled.

In the corresponding Italian example (2) however, *he* is omitted and one could have a contextual *agent*-arrow directly going to *Giovanni* instead.

(2) Entrava Giovanni. Sorrideva.

Other constraints may include the presence of certain attributes for certain heads (e.g. an agent arrow for a head like *cough*), at least one incoming arrow for an operator and the requirement that a marker node (a non-head and non-operator node) can only be the target of a mark arrow.

A different way of interpreting this paper is as a specific proposal for assigning truth-conditions to semantic frames, in combination with specific proposals for representing context-dependency, anaphora and scope within such semantic frames.

The existence of this alternative interpretation can be taken as a point of departure for various programmes to de-emphasize the distinction between semantic and syntactic representations.

Two important issues in this connection are the equivalence between different parses and lexical disambiguation. Arguably the natural equivalence class between parses is leading to the same class of fully disambiguated readings. The theory developed in this paper would be a way of filling in this notion which is often not well-motivated, since the DGs are not interpreted. Lexical disambiguation is arguably the most complex problem in determining the truth-conditional meaning of sentences[4] and impossible without computing the contribution of the context. Arguably, the reinterpretation of the dependency structure as a logical representation language can be a great help in dealing with lexical ambiguity, since it immediately gives access to consistency.

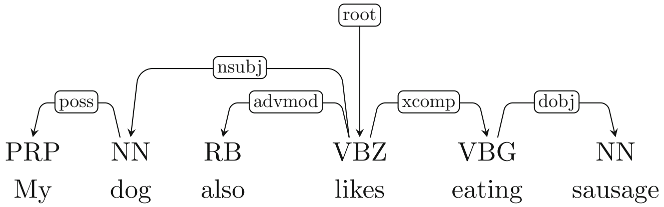
Zeevat (2014) makes the additional claim that formalisms like the dependency graphs discussed in this paper provide a formalisation of the pre-Fregean psychologistic logic based on mental representations, since the formalism makes sense in a systematic way of the idea that representations are both propositional and denoting, while maintaining a strict connection between the object of a representation and the way in which that object is given. It is also argued in Zeevat (2014) that formalisations of classical mental representations have advantages over modern representational formalisms such as DRT in which the link between the "discourse referents" and the way in which they were introduced is severed. For example, the formalism developed in this paper allows a uniform analysis of definiteness that incorporates familiarity, functionality and Russellian unique description, as a logical distinction between head nodes and a notion of intensional identity reminiscent of Carnap's between representations, that can be employed in the semantics of propositional attitudes.

[4] The number of readings of sentences can be estimated as m^n where n is the number of words and for a language like English m is roughly 5 of which 2.5 is due to lexical ambiguity alone.

2 What Are Dependency Graphs?

The following picture gives a typical example of a dependency graph as found in the literature[5].

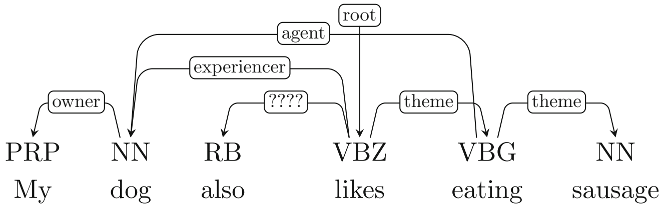
(3)



It does as such not meet the additional criteria that need to be imposed for semantic interpretation. The label *poss* must be interpreted by the attribute *owner* (the referent of a possessive pronoun can bear a number of relations to the referent of the head on which it depends, compare e.g. *his friends, his book, his soup*), *nsubj* should be replaced by *experiencer, advmod* is currently out of reach[6], *xcomp* should be *theme* and an *agent* link from *eating* to *dog* must be added. Also *dobj* should read *theme*.

The emendations would give the following new graph.

(4)



We can now interpret the dependency graph as saying (5).

(5) there is a dog
there is the speaker
the speaker owns the dog
there is a liking state
the experiencer of the liking state is the dog
the theme of the liking state is an activity of eating
the theme of the eating activity is sausage
the agent of the eating activity is the dog

[5] Taken from the tikz-dependency package documentation.

[6] *Also* has a syntactic associate X. The presupposition associated with *also* is that the clause already holds for some Y distinct from X. This can be checked if Y is connected by an *anap*-arrow to *also*. Such a treatment is however difficult, because the associates can be syntactically complex which needs a treatment of *anap*-arrows which allows for complex antecedents.

And the basic idea of the interpretation is straightforward, as in (6).

1. nodes with a lexical label denote an object that meets the label
2. if an arrow with a label L goes from A to B then the denotation f of the label L applied to the denotation of A is the denotation of B.

As a whole, the dependency graph can be understood as expressing its satisfiability: one should be able to find denotations for the lexical nodes that meet conditions (1) and (2).

Applied to *my dog likes the sausage*, assuming a denotation function *den* this gives the constraints in (6).

(6) $den(owner)(den(< i, dog >) = den(< l, me >)$
$den(< i, dog >) \in (den(me) = den(dog)$
$den(< j, like >) \in den(like)$
$den(< k, sausage >) \in den(sausage)$
$den(experiencer)(den(< j, like >) = den(< i, dog >)$

As a whole, the dependency graph can be understood as expressing its satisfiability: one should be able to find denotations for the lexical nodes that meet conditions (1) and (2). But such a way of satisfying also assigns a denotation to the root node (and to the other head nodes), in the example an event of liking, that—because it is functionally related to the denotation of the other head nodes—is the object denoted by the whole DG. Both the root node and the head nodes therefore both denote an object and express a proposition.

It makes sense to let the determiner *every* to be, like other determiners, a marker of the noun it depends on, while at the same time it must take in scope and restrictor arrows as an operator. Other hybrid cases are nodes that are both lexical and operators (*nobody, everywhere, belief*).

Operators can be defined as the targets of scope and restrictor arrows, and markers as the targets of mark arrows.

In this section, a run of the mill DG was turned into a DG that can be semantically interpreted in an intuitive way. What is needed for the map from ordinary DG to semantically interpretable DG is lexical knowledge to find out which thematic role is expressed by which syntactic function and further disambiguation. In the next section, the full requirements on semantically interpretable DGs will be presented.

3 What Should Be a Semantic Dependency Graph?

1. **Attributes**

The labels on the non-technical arrows should be interpretable as attributes in the‘sense of frame semantics, i.e. as partial functions over the domain of the model. The subject of a verb is not a good attribute since it defies a definition in purely semantic terms. The thematic roles that are employed in this paper also can be criticized from this point of view. Dowty (1991) decomposes the notion of *agent* and *theme* into proto-agent properties and proto-patient properties only some of which apply in particular uses that are described as *agents* and *patients*. This shows that *agent* and *patient* cannot be taken as semantic primitives. The ideal of doing away entirely with thematic roles cannot be realised *in abstracto* and requires detailed lexical semantics for verbs and nouns. Since that is beyond the scope of this paper, it is therefore unavoidable to stick with thematic roles and pretend contrafactually that they correspond with natural semantic functions. In this sense the proposal made in this paper is only an approximation.

2. **Anaphora**

Aanphora in the broad sense also contains ellipsis of various kinds, presupposition resolution and accommodation, the treatment of different missing arguments and contextual restrictions on the interpretation of predicates. These phenomena should be represented in the dependency graph and this is a non-trivial matter for most of these cases. The only device that is considered here are *anap*-arrows.

$X \rightarrow anap \rightarrow Y$

These can link heads to heads, even across the boundary of a DG to its context and are interpreted by referential identity.

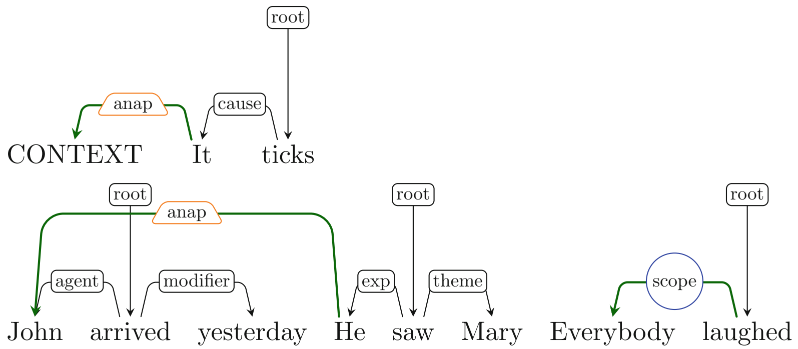
3. **Scope**

Operators have scope and English is notorious for undermarking the scope of operators. The DG should represent scopes, but also take account of the typological possibility that scope is marked syntactically (e.g. the split negations of French and Afrikaans are often interpreted as marking scope).

The proposal is to let the elements in the scope of an operator connect to the operator by a scope or restrictor arrow. The operators themselves can be treated as markers of various categories, e.g. *not* as a verbal marker, *no* as a nominal marker. For an operator like the verb *believe*, it seems most proper to treat it as both a head and an operator. That means that it can be the target of modifiers and also have scope.

Some of these new representational devices are illustrated in (7).

(7)



4 Dependency Graphs as Semantic Structures

Anaphora brings in the context in an essential way. Very few utterances have a truth-value without anaphoric links to the linguistic and non-linguistic context. For interpreting DGs it is therefore necessary to place them in contexts.

(8) **Definition**
A context is a finite sequence of DGs which may come with a set of *anap*-arrows of the form $< n \rightarrow anap \rightarrow m$ with a non-empty class of models where n and m are nodes of different DGs in the context and m comes from a DG that precedes the DG that contains m.

As sequences, contexts C and C' can be concatenated to $C \circ C'$ and a context C can be split up as follows: $C =< DG|C_{rest} >$. $<>$ is the empty sequence. Contexts are necessary for evaluating trans-sentential anaphora. We will adopt an incremental regime in which we define truth in a model given a precontext C $(C, M \models DG)$ where C is already true on the model and where the DG is added as the last element of the context.

(9) **Definition**
A *model* is a structure $M = (D_M, o_M, w_M, i_M, b_M)$ where

1. o_M: a partial map that assigns objects in w_M to a set of head nodes.
2. w_M: a complete map that assigns sets of objects from D_M to the head node labels L (and thereby to the nodes bearing the label).
3. i_M: a complete map that assigns partial functions over D to the arrow labels in AL.
4. b: assigns belief states (contexts) to a subset of objects from D_M.

The one innovation in the definition of a model is to give two values to head nodes: they define both a set of objects and a referent. This corresponds with the ambiguity of the notion of mental representation in the philosophical tradition as representing an object and as the proposition that there is an object that meets the representation. This ambiguity is eliminated in the treatment below: the node *man* is satisfied iff it denotes a man, but satisfaction and denotation are not the same thing.

Some auxiliary definitions.

A *main node* of a DG is a head node or operator node that is not the source of a scope or restrictor arrow.

Main nodes are the objects in the DG that have a referent and correspond with the discourse referents in DRT.

The *extension* (M, f) of a model M. If f is a function from a set of nodes to D_M, $(M, f) = (D_M, o_M \cup f, w_M, i_M, b_M)$ if $o_M \cup f$ is a function.

(M, f) provides a way to extend models of a context to a model of an extended context.

A context is true on a model M iff each of its DGs is true on the model given its precontext and each of its *anap*-links connects head nodes with an identical value under o_M.

But the models should be incrementable and this has to be spelled out in detail. So more precisely, $M \models C$ iff $<>, M, (M, g) \models C$. $C, M, M' \models D$ is defined in two steps.

(10) 1. $C, M, M \models <>$
2. $C, M, M_{out} \models < DG|C_{rest} >$ iff there is an f from the main nodes of DG to D_M such that $C, (M, f) \models DG$ and $C \circ < DG >, (M, f), M_{out} \models C_{rest}$

The definition of truth for a DG in a given context is given in (11).

(11) $C, (M, f) \models DG$ iff
$M \models C$ and $dom(o_M)$ is the set of main head nodes of C,
f is a function from the main head nodes of DG to D_M and for each of its main nodes m
$(M, f) \models m$

And finally, we need to define what it means for a model to satisfy a node of the DG. All the action is in this definition: the constraints deriving from the node labels, from the attribute labels and from the marking devices are treated here. In addition, each operator needs to be defined by a separate clause. Below, only negation, universal quantification and belief are treated, but it would not be difficult to come up with clauses for more operators. The markers need to be treated as extra conditions on the nodes m that they mark. This is dealt with below when the mark is an operator, but not in the other cases. A mark *definite* is a constraint on the interpretation of the head node: it should have, possibly through the addition of an *anap*-arrow, a definite interpretation. Conversely, a node marked as indefinite cannot have a definite interpretation. From the point of view of the model theory, both types of marking have no proper contribution. They invoke or prevent anaphora.

(12) If m is an unmarked head node
$C, M \models m$ iff
$o_M(m) \in w_M(m)$ and for each of the non-scope arrows $m \rightarrow l \rightarrow k$ of m
$i_M(l)(o_M(m)) = o_M(k)$

If m is a head node marked by x
$C, M \models m$ iff
$o_M(m) \in w_M(m)$, o_M meets x and for each of the non-scope arrows $m \rightarrow l \rightarrow k$
$i_M(l)(o_M(m)) = o_M(k)$

If m is a *negation* node
$C, M \models m$ iff
for the context $C_{scope} = \{k : k \rightarrow scope \rightarrow m\}$ there is no function f from the head nodes of C_{scope} to D_M such that $C, M, (M, f) \models C_{scope}$

If m is an *all* node,
$C, M \models m$ iff
$C_{restrictor} = \{k : k \rightarrow restrictor \rightarrow m\}$ and $C_{scope} = \{k : k \rightarrow scope \rightarrow m\}$ and for all f from the main head nodes of $C_{restrictor}$ to D_M such that $C, M, (M, f) \models C_{restrictor}$ there is a g from the main head nodes of C_{scope} to D_M such that $C, (M, f), (M, f \cup g) \models C_{scope}$.

If m is a *belief* node,
$C, M \models m$ iff $m \rightarrow experiencer \rightarrow k$ and $o_M(k) = d$ and $b_d = D$, there is a subcontext C_0 of C and other activated information such that $C_0 \cup D$ is a consistent context and $C_0 \cup D \models C_{scope} \subseteq D$

This last definition requires the definition in (13) of $C \models D \subseteq E$. Note that nodes n_1 and n_2 can be equivalent in a context C in two ways. In the place, when n_1 and n_2 are head nodes, it can be the case that in all models of C, n_1 and n_2 have the same denotation. In the other case, n_1 and $n2$ are operator nodes with scopes and restrictors that can be equivalent in the sense of having the same truth-value.

(13) $C \models E \subseteq F$ iff
there is an injection i from the main nodes of E to the main nodes of F such that for all $x \in dom(i)$ either x is a head node and $C \models ix = x$ or x is an operator node and ix is an operator node with the same operator and $C \models ix \leftrightarrow x$

Let's illustrate these definitions with some examples. The basic case is the interpretation of head nodes. Here models assign both a class (wH) and a referent (oH) to the head. The condition imposed by the head node is that $oH \in wH$. If the head node has an *anap*-arrow, a further constraint is that $oH = oX$ where X is the target of the *anap*-arrow. If H has other non-scope arrows with a label l, these give further constraints on oH, namely that it is mapped by the attribute $i(l)$ to oX where X is the target of the arrow. Heads finally may be marked, e.g. as indefinite. This will impose the further constraint that $\{oH\} \neq wH$.[7]

Consider the following example (14) which can be taken as the representation of "cat runs" in a language like Russian, Japanese or Latin in which definiteness and indefiniteness is not obligatorily marked.

(14) $H = run$
$H \rightarrow agent \rightarrow J$
$J = cat$

This leads to the constraints on M stated in (15).

(15) oH a running event
wH the class of running events

[7] While this constraint seems correct for indefinite NPs, it is unlikely to exhaust the contribution of indefiniteness marking.

oJ a cat
wJ the cats
the agent of oH is oJ

The second example (16) illustrates the indefinite marker.

(16) A man sleeps.
$o\ sleep \in w\ sleep$
$o\ man \in w\ man$
$i(theme)(o\ sleep) = o\ man$
$w\ man \neq \{o\ man\}$

Intersective modification can be captured by an attribute *mod* that is interpreted as identity. This is illustrated in (17). We assume here that *yesterday* is the set of events that happened yesterday. A more complete analysis would capture the deictic identification of the day before the contextually given *now* and derive the set of yesterday's events by applying an operation: yesterday is a day as well as the set of events that happened in the course of yesterday and the set of states that held yesterday.

(17) John slept yesterday.

slept → *theme* → John
slept → *mod* → yesterday

$o\ sleep \in w\ sleep$
$o\ john \in w\ john$
$i\ theme(o\ sleep) = o\ john$
$o\ yesterday \in w\ yesterday$
$o\ yesterday = o\ sleep$

Anaphoric links (*anap*-arrows) are interpreted in exactly the same way as intersective modification, by identifying referents. In (18), it is assumed that there is an *anap*-arrow to a H in an earlier element of the context.

(18) The man slept

$sleep\ \rightarrow theme \rightarrow man$
$man\ \rightarrow mark \rightarrow the$
$man\ \rightarrow anap \rightarrow H$
$i\ theme(o\ sleep) = o\ man$
$o\ man = o\ H$
$o\ man \in w\ man$
$o\ sleep = w\ sleep$

Notice that there is no separate interpretation of definiteness marking: it is already interpreted by the *anap*-arrow and should in a proper treatment of the interpretation be regarded as the cause of the *anap*-arrow. Here the view of Zeevat (2014) is followed that definiteness is a logical notion. Translated to the current context, that view can be described as the view that a head node can—given its dependents, the context and any model of that context—only receive a single value. This captures the three possibilities that have been defended as analyses of the linguistic feature of definiteness. Russell analysed definite descriptions—following Frege—as definitions. In Hawkins' familiarity theory they are anaphora. In the view of Löbner, they are attribute values for a given object. The speaker indicates by means of the definite marker that the noun—possibly with its dependents—should be interpreted as definite. The hearer has to check and disambiguate, possibly by creating an *anap*-arrow (the Hawkins case) or by creating an attribute arrow from the noun to a salient object in the context (the Löbner case). The Russell-Frege case is the one where checking is directly successful: the noun with its dependents picks out a unique object. The extra *anap*-link or the attribute link turns the noun (with its old and new dependents) into a definite concept.[8] Such additions however are part of the inference of the DG and therefore fall outside the scope of this paper. (19) is an example of the inference of an extra attribute arrow on the basis of definiteness marking.

(19) John is having trouble. The father thinks that he can solve the problem.
Inferred: john $\rightarrow$ *father* $\rightarrow$ father

4.1 Negation and Logic

The negation rule repeated here as (20) takes up the DRT-style negation that incorporates quantification. The main head nodes are the discourse referents, so that the negation and its scope expresses a formula of the form $\neg\exists x_1 \ldots x_n\varphi$. As is well known, the negation suffices for the expressive power of first order logic and can be used to define the DRT implication (here called *all*) and the basic quantifiers. Later occurrences of the same discourse referents need to be represented by *anap*- and *mod*-arrows. A restriction are *anap*-arrows into the scope of operators. They only make sense under special circumstances.[9]

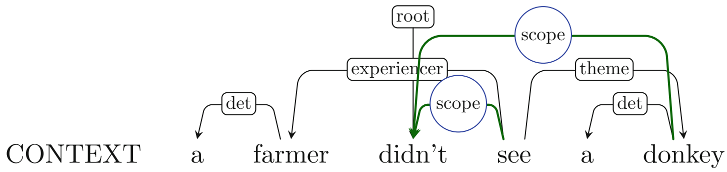
(20) If m is a *negation* node
$C, M \models m$ iff
for the context $C_{scope} = \{k : k \rightarrow \textit{scope} \rightarrow m\}$ there is no function f from the head nodes of C_{scope} to D_M such that $(C, M \models C_{scope}$

[8] A comprehensive discussion of definiteness within a related framework is in Chap. 5 of Zeevat (2014).

[9] The restriction captures the accessibility relation in Discourse Representation Theory: a pronoun cannot be bound by a bound variable. There are cases where the accessibility does not seem to operate. A famous case is: *A wolf might come in. It would eat you first.* A proper treatment of these cases is outside the scope of this paper.

An example is given in (21).

(21) A farmer didn't see a donkey



4.2 Attitudes

The motivation for the somewhat complicated semantics for belief sentences repeated in (22) is twofold.

(22) If m is a *belief* node,
$C, M \models m$ iff $m \to experiencer \to k$ and $o_M(k) = d$ and $b_d = D$, there is a subcontext C_0 of C and other activated information such that $C_0 \cup D$ is a consistent context and $C_0 \cup D \models C_{scope} \subseteq D$

$C \models E \subseteq F$ iff
there is an injection i from the main nodes of E to the main nodes of F such that for all $x \in dom(i)$ either x is a head node and $C \models ix = x$ or x is an operator node and ix is an operator node with the same operator and $C \models ix \leftrightarrow x$

In the first place, there are cases like (23) where John who does not know your brother (or knows about him or you) could never report his belief in this way.

(23) John believes that your brother is wounded.

The context must in that case have enough information about your brother and John's beliefs that makes it the case that "your brother" and "that guy" are the same in all models of the contextual information. The injection required asks that the main head nodes in the complement can be identified with the head nodes in the DG that represents the subject's belief state.[10]

Second, in Edelberg's famous story (Edelberg 1992), Arsky and Barsky are detectives investigating what they think is the murder of Smith, who in fact died from accidental causes. Arsky and Barsky operate without communication and after a time, Arsky but not Barsky has formed the belief that the murderer also killed Jones, in fact another case of an accidental death. It then holds that (24),

[10] This means there is a problem with the operators. A semantics of this kind requires that operator nodes also denote. A solution to this problem is in preparation.

(24) Barsky believes that someone murdered Smith and Arsky believes he killed Jones.

but not that (25).

(25) Arsky believes that someone killed Jones and Barsky believes that he killed Smith.

The *anap*-arrows in these cases go from the pronoun to the someone in the complement of the first belief-clause. The semantics provided can deal with the case only if it is possible to add that complement to the context under which the second belief is evaluated. This is needed for any case of intentional identity. So "he killed Jones" should be part of Arsky's beliefs where we can use "someone killed Smith" and consistent parts of the general context C to show that "he killed Jones" is part of those beliefs. That is unproblematic. The *anap*-arrow makes "he" into the murderer of Smith, the guy Arsky believes killed Jones.

The same procedure does not work for (0). The *anap*-arrow makes "he" into the murderer of Jones, while the constructed context lacks any means to identify the murderer of Jones with the murderer of Smith.

5 Conclusion

This paper explored the road of leaving the dependency structure intact rather than reconstructing mental representations as in Zeevat (2014) and it seems safe to conclude that this is another way of giving a variable-free version of DRT which moreover has the advantage of being very close to a very reasonable syntactic representation. The treatment of definites and beliefs sketched above is not possible in DRT where discourse referents and their introducing lexical elements are separated, in line with Frege's rejection of psychologism.

It is unclear to me whether the version in Zeevat (2014) has any advantages over the current version in terms of semantic potential. Anaphora is still a neglected theme in frame semantics and there are obviously two ways to go: build ever larger frames by unifying nodes in different frames or adding the *anap*-arrows of this paper. The brief discussion of the Edelberg asymmetry can be used as an argument for the *anap*-arrows.

Obvious further work should be directed towards incorporating plural NPs, tense and other semantic phenomena. Essential is also the incorporation of lexical semantics, including the proposal for disambiguation in Zeevat et al. (2015). It is only if that connection is made that the current proposal can be properly evaluated.

It cannot be stressed enough that enterprises such as word sense disambiguation, the detection of semantic roles, anaphora resolution, stochastic parsing, stochastic models of pragmatics are all partial answers in the battle to overcome the underdetermination of meaning by form, something that the human interpreter in a human conversation seems to do effortlessly. In this sense the current proposal is one of a common scoreboard where the results of different techniques can be gathered and which is close enough to syntax and morphology on the one hand to allow evaluation of syntactic and morphological criteria and on the hand semantically fully interpreted so that proof-theoretic and model-theoretic techniques can be directly applied.

References

Barsalou, L.: Frames, concepts and conceptual fields. In: Lehrer, A., Kittay, E.F. (eds.) Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organisation, pp. 21–74. Lawrence Erlbaum (1992)

Baum, R.: Dependenzgrammatik. Niemeyer, Tübingen (1976)

Dowty, D.: Thematic proto-roles and argument selection. Language **67**(3), 547–619 (1991)

Edelberg, W.: Intentional identity and the attitudes. Linguist. Philos. **15**(6), 561–596 (1992)

Hudson, R.: Word Grammar. Basil Blackwell, Oxford (1984)

Hudson, R.: Language Networks: The New Word Grammar. OUP, Oxford (2007)

Kamp, H., Reyle, U.: From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer Academic Publishers, Dordrecht (1993)

Lin, D.: A dependency-based method for evaluating broad-coverage parsers. Nat. Lang. Eng. **4**, 97–114 (1998)

Lobner, S.: Evidence for frames from human language. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Frames and Concept Types. Studies in Linguistics and Philosophy, vol. 94, pp. 23–67. Springer, New York (2014)

Lobner, S.: Functional concepts and frames. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Meaning, Frames, and Conceptual Representation. Studies in Language and Cognition, vol. 2, pp. 13–42. Düsseldorf University Press, Düsseldorf (2015)

Montague, R.: The proper treatment of quantification in ordinary english. In: Hintikka, J., Moravcsik, J., Suppes, P. (eds.) Approaches to Natural Language, pp. 221–242. Reidel (1973)

Nivre, J., Scholz, M.: Deterministic dependency parsing of English text. In: Proceedings of the 20th International Conference on Computational Linguistics, pp. 64–69. ACL (2004)

Petersen, W.: Representation of concepts as frames. In: Skilters, J., et al. (eds.) Complex Cognition and Qualitative Science. The Baltic International Handbook of Cognition, Logic and Communication, vol. 2, pp. 151–170. University of Latvia (2007)

Tesnière, L.: Éléments de Syntaxe Structurale. Klincksieck, Paris (1959)

Zeevat, H.: Language Production and Interpretation: Linguistics Meets Cognition. Jacob Brill, Leiden (2014)

Zeevat, H., Grimm, S., Hogeweg, L., Lestrade, S., Smith, E.A.: Representing the lexicon: identifying meaning in use via overspecification. In: Balogh, K., Petersen, W. (eds.) Proceedings of Workshop Bridging Formal and Conceptual Semantics (BRIDGE-14) (2015)

Reflexive and Reciprocal Determiners

Richard Zuber(✉)

CNRS, Laboratoire de Linguistique Formelle, Paris, France
Richard.Zuber@linguist.univ-paris-diderot.fr

Abstract. Reflexive (*most..., including himself* and reciprocal (*no... except each other*) determiners are anaphoric determiners. They form arguments of transitive verbs which cannot occur in subject position of sentences. Various logical properties (invariance, conservativity, a-conservativity, a-intersectivity) of functions denoted by these determiners are studied. These properties account for their anaphoricity and show formal differences between anaphoric and ordinary determiners.

1 Introduction

According to the well-established terminology, ("ordinary") determiners are functional expressions which take one or more common nouns (CNs) as arguments and give a noun phrase, (NP), as result. For instance *every, most, five, no except two* and *more... than...* are determiners. Syntactically NPs are arguments of intransitive, transitive or ditransitive verb phrases (VPs), that is they can occur as subjects, direct or indirect objects. There are, however, expressions which are arguments of verbs but which cannot occur in all argumental position of the verb, and thus, which are not, strictly speaking NPs:

(1) a. Leo and Lea kissed each other.
    b. * Each other kissed Leo and Lea.

(2) a. Leo and Lea washed themselves.
    b. *(They)selves washed Leo and Lea.

The reciprocal *each other* is an argument of the verb *kiss* in (1) where it occurs as a direct object. As shown in (1b) this reciprocal cannot occur in the subject position. Similarly, the reflexive *themselves* occurs in the object position in (2a) but it does not have the (corresponding) nominative form which could occur in the subject position.

Reciprocals and reflexives belong to the class of generalised NPs (GNPs) that is these nominal expression which typically fulfil the function of arguments of the main clause and thus can serve as arguments of (transitive) VPs. Obviously "ordinary" NPs are also GNPs. However, reciprocals and reflexives are proper (genuine) GNPs because, contrary to "ordinary" NPs, proper GNPs cannot occur in all argumental positions of a transitive VP, in particular they cannot occur in the subject positions, not even in the subject positions of simple

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 185–201, 2017.
DOI: 10.1007/978-3-662-54332-0_11

intransitive sentences. Typical examples of such GNPs are the reflexive pronouns *himself/herself/themselves* and the reciprocal pronoun *each other*. These can be Booleanly combined with other GNPs, proper or "ordinary", to give complex GNPs such as *each other but not themselves, himself and most students, ten students including each other and themselves*, etc. Here are some examples of sentences containing Booleanly complex GNPs:

(3) a. Leo admires himself and most linguists.
b. * Himself and most linguists admire Leo.

(4) a. Leo and Lea admire themselves and each other.
b. *Themselves and each other admire Leo and Lea.

In this paper I do not study GNPs in general, even if some differences between ordinary NPs and genuine GNPs will be indicated in Sect. 4. I will study here, in a preliminary way, functional expressions forming some GNPs. Functional expressions forming ordinary NPs, that is *(nominal) determiners* forming a DP (or a NP) from a CN have been extensively studied. Formal properties of (full) reciprocals and reflexives are studied in Zuber (2016). In this paper I analyse formal properties of (1) *reflexive determiners* (RefDets) that is functional expressions which take a CN as argument and form a reflexive GNP (like for instance *most..., including himself and Lea*) and (2) reciprocal determiners (RecDets), that is functional expressions which take a CN as argument and give a reciprocal GNP as result (like for instance *no... except each other and themselves*). Both these classes of functional expressions form generalised determiners (GDets). In addition, as will be shown below, GNPs formed by GDets considered here are anaphors. In that sense they are different from other GDets forming GNPs such as *the same* or *a different number of*, which do not form anaphors when applied to a CN.

I will be specifically interested in logical and semantic properties of functions denoted by RefDets and by RecDets. These properties will indicate formal similarities and differences between "ordinary" determiners (those forming "ordinary" DPs with a CN) and GDets considered here. They will also indicate differences and similarities between reflexives and reciprocals. Two kinds of such properties wiil be discussed: those related to the anaphoricity of reflexive and reciprocal determiners and those related to the conservativity of functions they denote. Concerning conservativity, two, logically related, types of it will be discussed, one of which is characteristic for anaphoric determiners.

In the next section we indicate in some detail the data we will be concerned with. Then formal tools from the extended Generalised Quantifier Theory are recalled. In Sect. 4 the semantics of various RefDets and RecDets is provided and Sect. 5 discusses formal properties which show differences and similarities between functions denoted by anaphoric determiners and quantifiers denoted by ordinary determiners.

2 Some Data

RefDets and RecDetds have been only scarcely discussed even if much more have been written about RefDets. Both these classes can be divided into possessive and non-possessive GDets. Some (but not all) languages have “marked” or morphologically simple possessive RefDets. The possessive anaphoric pronoun SVOJ in Slavic languages (as opposed to EGO) or *hans* in Norwegian (both meaning roughly *his/her own*) are probably well-known (Zuber 2009). The Polish pronoun *swój* can in addition combine with virtually any other “ordinary” determiner to give a series of complex possessive RefDets which in English corresponds to the series like *all of his own, most of his own, ten of his own*, etc.

Concerning possessive RecDets we have the possessive form *each other's* and various Boolean combination of it with “ordinary” (non anaphoric) possessives determiners. Thus *each other's but not Bill's ..., everybody's, including each other's...* are possessive RecDets as in the following examples:

(5) a. Leo and Lea help each other's but not Bill's (students).
 b. Leo and Lea help each other's and their own (students).

Interestingly, in Polish, the possessive RefDet SVOJ can, in many situations have the meaning corresponding to possessive RecDet *each other's*.

Non-possessive RefDets and RecDets are obtained from specific “ordinary” determiners. One can distinguish two classes of such RefDets: those obtained from, roughly speaking, inclusive determiners, and those obtained from exclusive determiners (Zuber 1998, Zuber 2010b). Inclusive determiners are discontinuous determiners of the form *Det,..., including EXP* (where *Det* is an ordinary simple determiner denoting a monotone increasing (on the second argument) type $\langle 1, 1\rangle$ quantifier) and exclusive determiners are determiners of the form *every/no... except EXP* The expression *EXP* is the complement of *including* or of *except*. Both these classes of determiners form a NP when applied to a CN.

By replacing the complement *EXP* of *including* by an expression which denotes a **PI** function (see below) we get inclusive anaphoric RefDets. Thus inclusive anaphoric RefDets are expressions of the form *Det, including himself/herself* or of the form *Det, including NP and himself/herself*. For example the following expressions are RefDets: *most...including herself, most...including some Albanians and himself, ten...including herself and two Japanese*, etc. The last determiner occurs in (6a). Observe that (6a) means (6b) and apparently cannot mean (6c). This fact is related to the anaphoricity of the determiner involved in (6a):

(6) a. Lea admires ten students, including herself and two Japanese.
 b. Lea admires ten students including herself and two Japanese students.
 c. Lea admires ten students including herself and two Japanese which are not students.

There are also "negative" inclusive determiners from which we can obtain Ref Dets and RecDets. In the following examples *no..., not even himself* is such a RefDet and *no... not even each other* is such a RecDet:

(7) a. Leo admires no linguist, not even himself.
b. Leo and Leo admire no linguist, not even each other

We will not analyse here such "negative" inclusive anaphoric determiners.

Exclusive ordinary determiners are determiners such as *every...except Leo, every... but two, no...except Japanese, no...except Albanian and Sue*, etc. By replacing in them the complement of *except* by a reflexive GNP (that is an expression whose denotation satisfies **PI** and does not satisfy **EC**) we can form RefDets like the following: *every... except himself, no...except Leo and herself*. The following sentence contains such a RefDet:

(8) Leo and Lea hate every linguist except themselves.

Not surprizingly, non-possessive RecDets can also be formed from the inclusive and exclusive "ordinary" determiners by putting as the complement of *including* or of *except* a reciprocal GNP. Thus in (9a) and (9b) we have RecDets based on inclusive "ordinary" determiners and in (10a), (10b) and (10c) - RecDets based on exclusive determiners:

(9) a. Leo and Lea hate most vegetarians, including each other.
b. Most teachers admire some Japanese, including each other and themselves.

(10) a. Leo and Lea admire no philosopher except each other and Plato.
b. Three linguists admire every linguist except each other.
c. Two monks admire no philosopher, except each other and themselves.

This way of constructing non-possessive RefDets and RecDets from the ordinary inclusive and exclusive determiners is productive in many languages.

Let us see now some differences between possessive and non-possessive anaphoric determiners in their relation to the class of "ordinary" determiners. Up to now we have considered only unary determiners. Natural languages have also n-ary determiners (Keenan and Moss 1985). For instance (11a) can naturally mean what (11b) means in which case *most...and...* should be considered as binary determiner. In other words the admiration of Leo concerns two groups of people: a group of linguists and a group of philosophers. Similarly in (12) we have a binary determiner *more...than...*:

(11) a. Leo admires most linguists and philosophers.
b. Leo admires most linguists and most philosophers.

(12) Lea knows more linguists than philosophers.

One observes that possessive RefDets and RecDets can take many CNs as arguments as seen in (13) and (14):

(13) Leo burnt more of his own paintings than letters.

(14) Leo and Bill like each other's books and articles.

It does not seem that there are non-possessive RefDets or non-possessive RecDets taking many nominal arguments: in (15) and in (16) only one group of people is involved, those who are linguists and philosophers "at the same time":

(15) Leo and Lea admire most linguists and philosophers, including themselves.

(16) Leo and Lea admire all linguists and philosophers, except each other.

In addition to *except* and *including* some other connectors can be used to form non-possessive anaphoric determiners. This is the case with *apart from* and, possibly, *in addition to*. Constructions with such connectors will be ignored in what follows.

In the next section we give the semantics for various types of anaphoric determiners presented above. Even if it is possible to extend various definitions given in the preceeding section to n-ary determiners, we will consider only the semantics of unary determiners. Furthermore, we will not provide the semantic description of possessive anaphoric determiners. Semantic properties of some possessive determiners are discussed in Zuber (2009).

3 Formal Preliminaries

We will consider binary relations and functions over a universe E, assumed to be finite throughout this paper. $D(R)$ denotes the domain of the relation R. The relation I is the identity relation: $I = \{\langle x, y\rangle : x = y\}$. If R is a binary relation and X a set then $R/X = R \cap (X \times X)$. The binary relation R^S is the greatest symmetric part of the relation R, that is $R^S = R \cap R^{-1}$. A symmetric relation R is cross-product iff $R = A \times A$ or $R = (A \times A) \cap I'$ for some $A \subseteq E$. If R is a symmetric relation then $\Pi(R)$ is the least fine partition of R such that every of its blocks is a cross-product relation and every two blocks have incompatible domain: if $B_1 \in \Pi(R)$ and $B_2 \in \Pi(R)$ then $D(B_1) \cap D(B_2) = \emptyset$. A partition is 1. *atomic* iff every of its blocks is a singleton; 2. *simgular* iff it contains only one block (which is not a singleton); 3. *non-trivial* iff it is neither atomic nor singular.

If a function takes only a binary relation as argument, its type is noted $\langle 2 : \tau\rangle$, where τ is the type of the output; if a function takes a set and a binary relation as arguments, its type is noted $\langle 1, 2 : \tau\rangle$. If $\tau = 1$ then the output of the function is a set of individuals and thus its type is $\langle 2 : 1\rangle$ or $\langle 1, 2 : 1\rangle$. The function $SELF$, denoted by the reflexive *himself* defined as $SELF(R) = \{x : \langle x, x\rangle \in R\}$, is of type $\langle 2 : 1\rangle$ and the function denoted by the anaphoric determiner *every...but himself* is of type $\langle 1, 2 : 1\rangle$. We will consider here also the case when

τ corresponds to a set of type $\langle 1 \rangle$ quantifiers and thus τ equals, in Montagovian notation, $\langle\langle\langle e,t\rangle t\rangle t\rangle$. The type of such functions will be noted either $\langle 2 : \langle 1 \rangle\rangle$ - functions from binary relations to sets of type $\langle 1 \rangle$ quantifiers)) or $\langle 1, 2 : \langle 1 \rangle\rangle$ - functions from sets and binary relations to sets of type $\langle 1 \rangle$ quantifiers.

Basic type $\langle 1 \rangle$ quantifiers are functions from sets to truth-values. In this case they are denotations of subject NPs. However, NPs can also occur in the direct object positions and in this case their denotations do not take sets (denotations of VPs) as arguments but denotations of TVPs (relations) as arguments. To account for this eventuality one extends the domain of application of basic type $\langle 1 \rangle$ quantifiers so that they apply to n-ary relations and have as output an (n–1)-ary relation. Since we are basically interested in binary relations, the domain of application of basic type $\langle 1 \rangle$ quantifiers will be extended by adding to their domain the set of binary relations. When a quantifier Q acts as a "direct object" we get its *accusative case extension* Q_{acc} (Keenan and Westerstahl 1997):

Definition 1. *For each type* $\langle 1 \rangle$ *quantifier* Q, $Q_{acc}R = \{a : Q(aR) = 1\}$, *where* $aR = \{y : \langle a, y\rangle \in R\}$.

A type $\langle 1 \rangle$ quantifier Q is *positive* iff $Q(\emptyset) = 0$ and Q is *atomic* iff it contains exactly one element, that is if $Q = \{A\}$ for some $A \subseteq E$. We will call a type $\langle 1 \rangle$ quantifier Q *natural* iff either Q is positive and $E \in Q$ or Q is not positive and $E \notin Q$; Q is plural, $Q \in PL$, iff if $X \in Q$ then $|X| \geq 2$.

A special class of type $\langle 1 \rangle$ quantifiers is formed by *individuals*: I_a is an individual (generated by $a \in E$) iff $I_a = \{X : a \in X\}$. More generally, $Ft(A)$, the *(principal) filter generated by a set.*A, is defined as $Ft(A) = \{X : X \subseteq E \wedge A \subseteq X\}$. Principal filters generated by singletons are called *ultrafilters*. Thus individuals are ultrafilters. They are denotations of proper names. NPs of the form *Every CN* denote principal filters generated by the denotation of *CN*. Meets of two principal filters are principal filters: $Ft(A) \cap Ft(B) = Ft(A \cup B)$. Thus conjunctions (supposed to denote meets) of proper names denote principal filters generated by the union of referents of the proper names.

We will use also the property of *living on* (cf. Barwise and Cooper 1981). The basic type $\langle 1 \rangle$ quantifier lives on a set A (where $A \subseteq E$) iff for all $X \subseteq E$, $Q(X) = Q(X \cap A)$. If E is finite then there is always a smallest set on which a quantifier Q lives. If A is a set on which Q lives we will write $Li(Q, A)$ and the smallest set on which Q lives will be noted $SLi(Q)$.

A related notion is the notion of a witness set of the quantifier Q, relative to the set A on which Q lives:

Definition 2. $W \in Wt_Q(A)$ *iff* $W \in Q \wedge W \subseteq A \wedge Li(Q, A)$.

Thus $Wt_Q(A)$ is the class of witness sets of Q relative to the set A on which Q lives.

Observe that any principal filter lives on the set by which it is generated, and, moreover, this set is its witness set. Atomic quantifiers live on the universe E only and weakly live on their unique elements.

"Ordinary" determiners denote functions from sets to type $\langle 1 \rangle$ quantifiers. They are thus type $\langle 1, 1 \rangle$ quantifiers.

Accusative extensions of type $\langle 1 \rangle$ quantifiers are specific type $\langle 2 : 1 \rangle$ functions. They satisfy the invariance property called *accusative extension condition* **EC** (Keenan and Westerstahl 1997):

Definition 3. *A type $\langle 2 : 1 \rangle$ function F satisfies **EC** iff for R and S binary relations, and $a, b \in E$, if $aR = bS$ then $a \in F(R)$ iff $b \in F(S)$.*

Observe that if F satisfies **EC** then for all $X \subseteq E$ either $F(E \times X) = \emptyset$ or $F(E \times X) = E$. Given that $SELF(E \times A) = A$ the function $SELF$ does not satisfy **EC**. The function $SELF$ satisfies the following weaker predicate invariance condition **PI** (Keenan 2007):

Definition 4. *A type $\langle 2 : 1 \rangle$ function F is predicate invariant (**PI**) iff for R and S binary relations, and $a \in E$, if $aR = aS$ then $a \in F(R)$ iff $a \in F(S)$.*

This condition is also satisfied for instance by the function $ONLY\text{-}SELF$ defined as follows: $ONLY\text{-}SELF(R) = \{x : xR = \{x\}\}$. Given that $ONLY\text{-}SELF(E \times \{a\}) = \{a\}$, the function $ONLY\text{-}SELF$ does not satisfy **EC**.

The following proposition indicates another way to define **PI**:

Proposition 1. *A type $\langle 2 : 1 \rangle$ function F is predicate invariant iff for any $x \in E$ and any binary relation R, $x \in F(R)$ iff $x \in F(\{x\} \times xR)$.*

The conditions **EC** and **PI** concern type $\langle 2 : 1 \rangle$ functions, considered here as being denoted by "full" verbal arguments or GNPs. Such verbal arguments can be syntactically complex in the sense that they are formed by the application of *generalised determiners* (GDets) to CNs. For instance the GDet *every...except himself* can apply to the CN *student* to give a genuine GNP *every student except himself*. In this case GDets denote type $\langle 1, 2 : 1 \rangle$ functions. Such functions also are constrained by invariance conditions. Thus:

Definition 5. *A type $\langle 1, 2 : 1 \rangle$ function F satisfies **D1EC** iff for R and S binary relations, $X \subseteq E$ and $a, b \in E$, if $aR \cap X = bS \cap X$ then $a \in F(X, R)$ iff $b \in F(X, S)$.*

Observe that if $F(X, R)$ satisfies **D1EC** then for all $X, A \subseteq E$ either $F(X, E \times A) = \emptyset$ or $F(X, E \times A) = E$. Denotations of ordinary determiners occurring in NPs which are in the direct object position satisfy **D1EC**. More precisely, if D is a type $\langle 1, 1 \rangle$ (conservative) quantifier, then the function $F(X, R) = D(X)_{acc}(R)$ satisfies **D1EC**. Indeed, in this case $F(X, R) = \{y : D(X)(yR \cap X) = 1\}$ and $F(X, S) = \{y : D(X)(yS \cap X) = 1\}$. So if $aR \cap X = bS \cap X$ then $a \in F(X, R)$ iff $b \in F(X, S)$.

Functions denoted by properly anaphoric determiners (ones which form GNPs denoting functions satisfying **PI** but failing **EC**) do not satisfy **D1EC**. For instance the function $F(X, R) = \{y : X \cap yR = \{y\}\}$ denoted by the anaphoric determiner *no... except himself/herself* does not satisfy **D1EC**. To see this observe that for $A = \{a\}$ and X such that $a \in X$ one has $F(X, E \times A) = \{a\}$ and thus $F(X, E \times X) \neq \emptyset$ and $F(X, E \times X) \neq E$.

Type $\langle 1, 2 : 1 \rangle$ functions denoted by anaphoric determiners do not satisfy **D1EC**. They satisfy the following weaker condition (Zuber 2010b):

Definition 6. *A type* $\langle 1, 2 : 1 \rangle$ *function* F *satisfies* ***D1PI*** *(*predicate invariance for unary determiners*) iff for* R *and* S *binary relations* $X \subseteq E$*, and* $x \in E$*, if* $xR \cap X = xS \cap X$ *then* $x \in F(X, R)$ *iff* $x \in F(X, S)$*.*

The following proposition indicates an equivalent way to define **D1PI**:

Proposition 2. *A type* $\langle 1, 2 : 1 \rangle$ *function* F *satisfies* ***D1PI*** *iff for any* $x \in E$*,* $X \subseteq E$*, any binary relation* R *one has* $x \in F(X, R)$ *iff* $x \in F(X, (\{x\} \times X) \cap R)$*.*

The above invariance principles concern type $\langle 2 : 1 \rangle$ and type $\langle 1, 2 : 1 \rangle$ functions. We need to present similar "higher order" invariance principles for type $\langle 2 : \langle 1 \rangle \rangle$ and type $\langle 1, 2 : \langle 1 \rangle \rangle$ functions that is functions having as output a set of type $\langle 1 \rangle$ quantifiers. This is necessary because, as we will see, some type $\langle 1, 2 : \langle 1 \rangle \rangle$ functions are denotations of RecDets.

One can distinguish various kinds of type $\langle 2 : \langle 1 \rangle \rangle$ and type $\langle 1, 2 : \langle 1 \rangle \rangle$ functions. Observe first that any type $\langle 2 : 1 \rangle$ function whose output is denoted by a VP can be lifted to a type $\langle 2 : \langle 1 \rangle \rangle$ (type $\langle \langle \langle e, t \rangle t \rangle t \rangle$ in Montague notation) function. This is in particular the case with the accusative extensions of a type $\langle 1 \rangle$ quantifier. For instance the accusative extension of a type $\langle 1 \rangle$ quantifier can be lifted to type $\langle 2 : \langle 1 \rangle \rangle$ function in the way indicated in (17). Such functions will be called *accusative lifts*. More generally, if F is a type $\langle 2 : 1 \rangle$ function, its lift F^L, a type $\langle 2 : \langle 1 \rangle \rangle$ function, is defined in (18):

(17) $Q^L_{acc}(R) = \{Z : Z(Q_{acc}(R)) = 1\}$.

(18) $F^L(R) = \{Z : Z(F(R)) = 1\}$.

The variable Z above runs over the set of type $\langle 1 \rangle$ quantifiers.

For type $\langle 2 : \langle 1 \rangle \rangle$ functions which are lifts of type $\langle 2 : 1 \rangle$ functions we have:

Proposition 3. *If a type* $\langle 2 : \langle 1 \rangle \rangle$ *function* F *is a lift of a type* $\langle 2 : 1 \rangle$ *function then for any type* $\langle 1 \rangle$ *quantifiers* Q_1 *and* Q_2 *and any binary relation* R*, if* $Q_1 \in F(R)$ *and* $Q_2 \in F(R)$ *then* $(Q_1 \wedge Q_2) \in F(R)$*.*

For type $\langle 2 : \langle 1 \rangle \rangle$ functions which are accusative lifts we have:

Proposition 4. *Let* F *be a type* $\langle 2 : \langle 1 \rangle \rangle$ *function which is an accusative lift. Then for any* $A, B \subseteq E$*, any binary relation* R*,* $Ft(A) \in F(R)$ *and* $Ft(B) \in F(R)$ *iff* $Ft(A \cup B) \in F(R)$*.*

Accusative lifts satisfy the following higher order extension condition **HEC** (Zuber 2014):

Definition 7. *A type $\langle 2 : \langle 1 \rangle\rangle$ function F satisfies **HEC** (higher order extension condition) iff for any natural type $\langle 1 \rangle$ quantifiers Q_1 and Q_2 with the same polarity, any $A, B \subseteq E$, any binary relations R, S, if $Li(Q_1, A)$, $Li(Q_2, B)$ and $\forall_{a \in A} \forall_{b \in B}(aR = bS)$ then $Q_1 \in F(R)$ iff $Q_2 \in F(S)$.*

Functions satisfying **HEC** have the following property:

Proposition 5. *Let F satisfies **HEC** and let $R = E \times C$, for $C \subseteq E$ arbitrary. Then for any $X \subseteq E$ either $Ft(X) \in F(R)$ or for any X, $Ft(X) \notin F(R)$*

Thus a function satisfying **HEC** condition and whose argument is the cross-product relation of the form $E \times A$, has in its output either all principal filters or no principal filter.

It follows from Proposition 5 that lifts of genuine predicate invariant functions do not satisfy **HEC**. They satisfy the following weaker condition (Zuber 2014):

Definition 8. *A type $\langle 2 : \langle 1 \rangle\rangle$ function F satisfies **HPI** (higher order predicate invariance) iff for type $\langle 1 \rangle$ quantifier Q, any $A \subseteq E$, any binary relations R, S, if $Li(Q, A)$ and $\forall_{a \in A}(aR = aS)$ then $Q \in F(R)$ iff $Q \in F(S)$.*

An equivalent way to define **HPI** is given in Proposition 6:

Proposition 6. *Function F satisfies **HPI** iff if $Li(Q, A)$ then $Q \in F(R)$ iff $Q \in F((A \times E) \cap R)$*

The above definitions of **HEC** and of **HPI** easily extend to type $\langle 1, 2 : \langle 1 \rangle\rangle$ functions, which are, as we will see, denotations of RecDets:

Definition 9. *A type $\langle 1, 2 : \langle 1 \rangle\rangle$ function F satisfies **D1HEC** (higher order extension condition for unary determiners) iff for any natural type $\langle 1 \rangle$ quantifiers Q_1 and Q_2 with the same polarity, any $A, B \subseteq E$, any binary relations R, S, if $Li(Q_1, A)$, $Li(Q_2, B)$ and $\forall_{a \in A} \forall_{b \in B}(aR \cap X = bS \cap X)$ then $Q_1 \in F(X, R)$ iff $Q_2 \in F(X, S)$.*

Definition 10. *A type $\langle 1, 2 : \langle 1 \rangle\rangle$ function F satisfies **D1HPI** (higher order predicate invariance for unary determiners) iff for any type $\langle 1 \rangle$ quantifier Q, any $A \subseteq E$, any binary relations R, S, if $Li(Q, A)$ and $\forall_{a \in A}(aR \cap X = aS \cap X)$ then $Q \in F(X, R)$ iff $Q \in F(X, S)$.*

The condition **D1HPI** can also be characterised as in:

Proposition 7. *$F(X, R)$ satisfies **D1HPI** iff if Q lives on A then $Q \in F(X, R)$ iff $Q \in F(X, (A \times X) \cap R)$*

The second series of properties of functions we will discuss concerns conservativity. Recall first the constraint of conservativity for type $\langle 1, 1 \rangle$ quantifiers:

Definition 11. *$F \in CONS$ iff $F(X, Y) = F(X, X \cap Y)$ for any $X, Y \subseteq E$*

Conservative quantifiers have two important sub-classes: intersective and co-intersective quantifiers (Keenan 1993): a type $\langle 1,1\rangle$ quantifier F is intersective (resp. co-intersective) iff $F(X_1,Y_1) = F(X_2,Y_2)$ whenever $X_1 \cap Y_1 = X_2 \cap Y_2$ (resp. $X_1 \cap Y_1' = X_2 \cap Y_2'$).

All the above properties of quantifiers can be generalised so that they apply to simple and higher order functions (Zuber 2010a):

Definition 12. *A function F of type $\langle 1,2:\tau\rangle$ is conservative iff $F(X,R) = F(X,(E\times X)\cap R)$.*

Definition 13. *A type $\langle 1,2:\tau\rangle$ function is intersective iff $F(X_1,R_1) = F(X_2,R_2)$ whenever $(E\times X_1)\cap R_1 = (E\times X_2)\cap R_2$.*

Definition 14. *A type $\langle 1,2:\tau\rangle$ function is co-intersective iff $F(X_1,R_1) = F(X_2,R_2)$ whenever $(E\times X_1)\cap R_1' = (E\times X_2)\cap R_2'$.*

As in the case of type $\langle 1,1\rangle$ quantifiers it is possible to give other, equivalent, definitions of intersectivity for type $\langle 1,2:\tau\rangle$ functions:

Proposition 8. *F is intersective iff $F(X,R) = F(E,(E\times X)\cap R)$.*

One can notice that intersective and co-intersective functions are conservative. Furthermore, the type $\langle 1,2:1\rangle$ function $F(X,R) = D(X)_{acc}(R)$ and the type $\langle 1,2:\langle 1\rangle\rangle$ function $F(X,R) = D(X)^{L}_{acc}(R)$ are intersective if D is an intersective type $\langle 1,1\rangle$ quantifier. In Sect. 5 we will additionally define stronger properties of conservativity, intersectivity and co-intersectivity, properties which are displayed by anaphoric but not by ordinary determiners.

Interestingly for functions satisfying **D1PI** or **D1HPI** we have:

Proposition 9. *Any function satisfying **D1PI** or **D1HPI** is conservative.*

Observe that most of the above definitions do not depend on the type τ and thus they apply to type $\langle 1,2:1\rangle$ and type $\langle 1.2:\langle 1\rangle\rangle$ functions.

4 Semantics of Anaphoric Determiners

For simplicity we will consider that reciprocals formed from RecDets give rise only to full (logical) reciprocity. This means, informally, that given a group of participants in an action described by a transitive verb which can be interpreted as involving reciprocity, all members of the group are in this relation with each other. Indeed, it seems that contrary to the interpretation of the full reciprocal *each other* complex reciprocals cannot easily get a weaker interpretation of reciprocity (cf. Dalrymple *et al.* 1998).

As we have seen, we are considering sentences of the form given in (19) - for RefDets and in (20) - for RecDets:

(19) $NP\ TVP\ RefDet(CN)$

(20) $NP\ TVP\ RecDet(CN)$

In order to present semantics and some formal properties of RefDets and RecDets the first thing we have to do is to determine their grammatical category and the type of functions they denote. This problem is solved for RefDets: since they form reflexive GNPs by applying to a CN and reflexive GNPs denote type $\langle 2:1\rangle$ functions, RefDets denote a type $\langle 1,2:1\rangle$ function. Reciprocal GNPs and RecDets differ in many respects from reflexive GNPs and RefDets respectively. Both these classes also differ from ordinary dets and ordinary NPs. We have already seen some syntactic differences. To see semantic differences between genuine (anaphoric) GNPs and ordinary NPs consider the following examples:

(21) a. Leo and Lea hug each other.
b. Bill and Sue hug each other.

(22) Leo, Lea, Bill and Sue hug each other.

Clearly (21a) in conjunction with (21b) does not entail (22). Thus, given Proposition 3, functions denoted by reciprocal GNPs are not lifts of type $\langle 2:1\rangle$ functions and the conjunction *and* is not understood pointwise. Hence, to avoid the type mismatch and get the right interpretations we will consider that the GNPs *each other* denotes a type $\langle 2:\langle 1\rangle\rangle$ function and RecDets denote type $\langle 1,2:\langle 1\rangle\rangle$ functions.

We can now look at the semantics of anaphoric determiners. We consider first the class of inclusive anaphoric determiners. As we have seen, a frequent form of inclusive RefDets is given in (23), (where *Det* is an ordinary determiners denoting a monotonic (on the second place) type $\langle 1,1\rangle$ quantifier), *CONJ* is a binary operator. The part *CONJ NP* can be omitted. An example of the determiner of the form (23) is given in (6a). Some other examples are given in (24a) and (24b). As these examples show the Boolean operator *CONJ* needs not to be a "simple conjunction":

(23) *Det...including himself CONJ NP*

(24) a. Dan kissed most students including himself, Leo and Lea.
b. Dan hates most monks including himself but not most Japanese (monks).
c. Dan hates ten logicians including himself or Leo.

The functions denoted by RefDets of the from (23) is given in (25), where D is the denotation of *Det*, $\otimes$ - the denotation of *CONJ* and *NP* denotes Q:

(25) $F(X,R) = \{y : y \in X \wedge \langle y,y\rangle \in R \wedge y \in D(X)_{acc}(R) \otimes y \in Q_{acc}(R) \wedge SLi(Q,A) \subseteq X\}$

To give the semantics of anaphoric RecDets we will use the partition $\Pi(R^S/X)$. Our definitions will be definitions "be cases" which are determined by the fact that the partition $\Pi(R^S/X)$ is atomic, singular or non-trivial. Thus (27) gives the semantics for RecDets of the form (26), where the *Ft(G)NP* is a NP denoting the principal filter generated by the set G and $EXT(X) = \{X\}$:

(26) *Det... including each other CONJ Ft(G)NP*

(27) (i) $F(X,R) = \emptyset$ if $R^S/X = \emptyset$ or $\Pi(R^S/X)$ is atomic
(ii) $F(X,R) = \{Q : Q \in PL \wedge Li(Q,X) \wedge EXT(D(B)) \subseteq Q \otimes Q \in Ft(X \cap G)^L_{acc}(R)\}$ if $\Pi(R^S/X)$ is singular and B is its only block.
(iii) $F(X,R) = \{Q : Q \in PL \wedge Li(Q,X) \wedge \exists_B(B \in \Pi(R^S/X) \wedge Q(D(B)) = 1) \otimes Q \in Ft(X \cap G)^L_{acc}(R)\}$ if $\Pi(R^S/X)$ is non-trivial.

Clause (i) takes into account the fact that NPs like *nobody, no two individuals, no three students*, etc. cannot occur in the subject position of sentences of the form (26). When the partition has only one block B (clause (ii)) then this block is a product relation and only members of the domain of B are in the mutual relation determined by R.

Let us see now functions denoted by exclusive Ref Dets and exclusive RecDets. Various results concerning exclusive RefDets are given in Zuber (2010b). Exclusive determiners denote intersective or co-intersective type $\langle 1,1\rangle$ quantifiers. Such quantifiers form atomic Boolean algebras whose atoms are uniquely determined by sets. More precisely atoms of the intersective algebra are functions At_A such that $At_A(X)(Y) = 1$ iff $X \cap Y = A$ and atoms of the co-intersective algebra are functions At_B such that $At_B(X)(Y) = 1$ iff $X \cap Y' = B$, $(A, B, X, Y \subseteq E)$.

Atoms of intersective and co-intersective algebras are denoted precisely by exclusive dets which have as the complement of *except* a conjonction of proper names. Thus, roughly speaking, exclusive determiners with *No* denote atoms of the intersective algebra and exclusive determiners with *Every* denote atoms of the co-intersective algebra. For instance the determiner *no...except Leo* denotes the atomic intersective quantifier determined by the singleton $\{L\}$ whose only element is Leo and the determiner *every...except Leo and Lea* denotes the atom of co-intersective functions determined by the set composed of Leo and Lea.

Consider now some examples of type $\langle 1,2:1\rangle$ functions and RefDets denoting them (cf. Zuber 2010b). Let At_A be the (intersective or co-intersective) atom determined by the set A. The type $\langle 1.2:1\rangle$ function F_{At_A} given in (28) is an anaphoric function based on the atomic quantifier At_A. Furthermore, if At_A is intersective then F_{At_A} is intersective and if At_A is co-intersective then F_{At_A} is co-intersective:

(28) $F_{At_A}(X,R) = \{x : x \notin A \wedge At_{A\cup\{x\}}(X)(xR) = 1\}$

Let us see some functions which are instances of (28) for illustration. Take the type $\langle 1,1\rangle$ quantifier NO. It is the atomic intersective quantifier determined by the empty set. Thus $A = \emptyset$, $At_\emptyset = NO$ and consequently, given the values of NO, the anaphoric function F_{NO} based on NO is given in (29):

(29) $F_{NO}(X, R) = \{x : X \cap xR = \{x\}\}$

The function in (29) is the denotation of the RefDet *no...except himself/herself.*

If $At_A = EVERY\text{-}BUT\text{-}\{L\}$ (where $EVERY\text{-}BUT\text{-}\{\text{L}\}(X, Y) = 1$ iff $X \cap Y' = \{L\}$) then the anaphoric function based on $EVERY\text{-}BUT\text{-}\{L\}$ is given in (30). This function is the denotation of the anaphoric determiner *every... except Leo and himself* (if *Leo* refers to L) which occurs in (31):

(30) $F_{EVERY-BUT-\{L\}}(X, R) = \{x : X \cap xR' = \{x, L\}\}$

(31) Dan admires every linguist except Leo and himself.

Thus (28) gives us a class of functions which are denotable by RefDets.

Let us see now the functions denoted by some exclusive RecDets. To do this we will also use the partition $\Pi(R^S/X)$. In (32) we have the function denoted by the reciprocal determiner *no...except each other*:

(32) (i) $F(X, R) = \{Q : Q \in PL \wedge \neg TWO(E) \subseteq Q\}$ if $R^S/X = \emptyset$ or $\Pi(R^S/X)$ is atomic
(ii) $F(X, R) = \{Q : Q \in PL \wedge D(B) \times D'(B) \cap R = \emptyset \wedge B \cap I' = B \wedge EXT((D(B)) \subseteq Q\}$ if $\Pi(R^S/X)$ has B as its only block.
(iii) $F(X, R) = \{Q : Q \in PL \wedge \exists_B (B \in \Pi(R^S/X)) \exists_W (W \in Wt_Q(SLi(Q) \wedge (W \times W) \cap I') = B \wedge D(B) \times D'(B) \cap R = \emptyset\}$ if $\Pi(R^S/X)$ is non-trivial.

To illustrate (32) let $R = \{\langle a, b\rangle, \langle b, a\rangle, \langle a, c\rangle, \langle c, d\rangle, \langle d, c\rangle\}$ and $E = X = \{a, b, c, d\}$. In this case $R^S/X = \{B_1, B_2\}$, where $B_1 = \{\langle a, b\rangle, \langle b, a\rangle\}$ and $B_2 = \{\langle c, d\rangle, \langle d, c\rangle\}$ and thus the clause (iii) applies. Consequently $(I_a \wedge I_b) \notin F(X, R)$ - because $\langle a, c\rangle \in R$, and $(I_c \wedge I_d) \in F(X, R)$. If $R = (A \times A) \cap I'$, where $A = X = \{a, b, c\}$ then $\Pi(R^S/X$ is singular with $B = R$ and $D(B) = A$. Hence, given clause (ii) $EXT(A) \in F(X, R)$, $I_a \wedge I_b \wedge I_c) \in F(X, R)$, $I_b \wedge I_c) \in F(X, R)$. In addition, for instance $Q = \neg(I_c \wedge I_d) \in F(X, R)$ because $EXT(A) \subseteq Q$.

To obtain the function denoted by *every... except each other* observe the following equivalence (supposing that *like* is the negation of *dislike*):

(33) Leo and Lea like every student except each other.

(34) Leo and Lea dislike no student except each other.

We can thus consider that the function $G(X, R)$ denoted by *every...except each other* can be obtained from the function $F(X, R)$ denoted by *no... except each other* by changing the relational argument into its Boolean complement: $G(X, R) = F(X, R')$.

5 Formal Properties

The functions described in the previous section are anaphoric in the sense that they satisfy predicate invariance conditions **D1PI** or **D1HPI** and do not satisfy the weaker conditions **D1EC** or **D1HEC**. This is easy to see for functions in (25), (29) and (30). To show that functions denoted by RecDets do not satisfy **D1HEC** we can use Proposition 10, analogous to Proposition 5:

Proposition 10. *Let F satisfies* ***D1HEC*** *and let $R = E \times C$, for $C \subseteq E$ arbitrary. Then for any $A \subseteq E$ either $Ft(A) \in F(X, R)$ or for any X, $Ft(A) \notin F(X, R)$*

Using Proposition 10 one can show that function in (32) and the function denoted by *every..., except each* are anaphoric.

Examples of RefDets discussed above suggest that functions they denote satisfy a constraint stronger than conservativity. Observe that the anaphoric functions given in (25), (28) and (29) all have the property given in (35):

(35) $F(X, R) \subseteq X$.

This is also true of denotations of anaphoric determiners formed with *self* and other connectives than *except* or *including*. It is easy to see that the determiner like *five..., in addition to Lea and himself* also denotes a function which satisfies the condition given in (35). We see for instance that in (6a) Lea is a student and in (8) Leo and Lea are linguists.

Interestingly, the anaphoric condition **D1PI** and the condition given in (35) entail a specific version of conservativity, *anaphoric conservativity* (or a-conservativity), specific to non possessive anaphoric determiners. It is defined as follows:

Definition 15. *A type $\langle 1, 2 : \tau \rangle$ function F is* a-conservative *iff $F(X, R) = F(X, (X \times X) \cap R)$.*

The following proposition makes clearer what a-conservativity is:

Proposition 11. *A type $\langle 1, 2 : \tau \rangle$ function F is* a-conservative *iff for any $X \subseteq E$ and any binary relations R_1 and R_2 if $(X \times X) \cap R_1 = (X \times X) \cap R_2$ then $F(X, R_1) = F(X, R_2)$.*

Thus, informally, second, relational arguments of an a-conservative function give rise to different values of the function only if they differ by a specific symmetric part formed from the first argument of the function.

Any a-conservative function is conservative. Ordinary determiners in object position in general do not denote a-conservativce functions: if D is a (conservative) type $\langle 1, 1 \rangle$ quantifier, then the type $\langle 1, 2 : 1 \rangle$ function $F(R, X) = D(X)_{acc}(R)$ is not a-conservative. For instance if $D = ALL$ and $R = E \times A$ then $F(X, R) = ALL(X)_{acc}(E \times A) = E$ if $X \subseteq A$ but in this case $F(X, (X \times X) \cap R) = ALL(X)_{acc}((X \times X) \cap (E \times A) = X$. Thus $F(X, R) \neq F(X, (X \times X) \cap R)$ which means that $F(X, R) = ALL(X)_{acc}(R)$ is not a-conservative (though it is conservative).

Concerning RefDets and a-conservativity we have:

Proposition 12. *A type $\langle 1, 2 : 1 \rangle$ function F satisfying* ***D1PI*** *such that $F(X, R) \subseteq X$ is a-conservative.*

Thus the functions denoted by (non-possessive) reflexive anaphoric determiners are a-conservative.

When one looks at type $\langle 1, 2 : \langle 1 \rangle\rangle$ functions $F(X, R)$, denotations of non-possessive RecDets, one observes that they have the property given in (36):

(36) If $Q \in F(X, R)$, then Q lives on X.

For instance in (9a) Leo and Lea are vegetarians and thus the quantifier denoted by *Leo and Lea* weakly lives on the set $VEGETARIAN$. Similarly, in (9b) most teachers are Japanese and thus the quantifier $MOST(TEACHER)$ weakly lives on the set $JAPANESE$.

Properties indicated in (35) and (36) are related to the meaning of the connectors *including* and *except.* occurring in non-possessive anaphoric determiners. Possessive anaphoric determiners do not have these properties.

For functions denoted by non-possessive RecDets which satisfy the condition in (36) we have:

Proposition 13. *Any type $\langle 1, 2 : \langle 1 \rangle\rangle$ conservative functions satisfying* ***D1HPI*** *and the condition in (36) is a-conservative.*

We can thus suppose that *self* and *each other* type anaphoric determiners denote a-conservative functions.

More can be said with respect to the class of functions denoted by anaphoric exclusive determiners. Since they are related either to "ordinary" intersective determiners (like *no... except Leo*) or to "ordinary' co-intersective determiners (like *every... except Lea*) they are provably either intersective or co-intersective (in the sense of definitions D13 and D14 respectively). The function in (32) is intersective and the function denoted by *every..., except each other* is co-intersective.

In addition, given that the functions we consider satisfy predicate invariance and condition like (35) or (36), they have a stronger property than just intersectivity or co-intersectivity: they are a-intersective or a-co-intersective:

Definition 16. *A type $\langle 1, 2 : \tau\rangle$ function F is a-intersective iff $F(X_1, R_1) = F(X_2, R_2)$ whenever $(X_1 \times X_1) \cap R_1 = (X_2 \times X_2) \cap R_2$.*

Definition 17. *A type $\langle 1, 2 : \tau\rangle$ function F is a-co-intersective iff $F(X_1, R_1) = F(X_2, R_2)$ whenever $(X_1 \times X_1) \cap R_1' = (X_2 \times X_2) \cap R_2'$.*

The following proposition gives another characterisation of the a-intersectivity and a-co-intersectivity:

Proposition 14. *A type $\langle 1, 2 : \tau\rangle$ function F is a-intersective iff $F(X, R) = F(E, (X \times X) \cap R)$.*

Proposition 15. *A type $\langle 1, 2 : \tau\rangle$ function F is a-co-intersective iff $F(X, R) = F(E, ((X \times X)') \cup R)$.*

Functions which are a-intersective or a-co-intersective are a-conservative. Functions in (29) and in (32) are a-intersective and functions in (30) and the one denoted by *every ..., except each other* are a-cointersective.

6 Conclusive Remarks

Any discussion of the meaning of (full) reflexives and reciprocals necessitates the use of simple logical tools from the theory of relations. In this paper such tools, in addition to the generalised quantifier theory, have been used to discuss logical properties of anaphoric determiners, that is functional expressions which apply to CNs and form reflexive or reciprocals. Syntactically, anaphoric determiners are discontinuous formatives which contain as their parts "ordinary" determiners and anaphoric pronouns like *himself* or *each other*. This fact entails the proposal made here concerning the logical type of functions denoted by anaphoric determiners: these functions take two arguments: the first argument is a set, because they are denoted by determiners and the second argument is a binary relation because they form simple nominal anaphors. Formal properties of such anaphoric determiners are inherited from the properties of their parts: they are conservative (intersective, co-intersective) because the "ordinary" determiners that compose them are conservative (intersective, co-intersective) and they are predicate invariant because anaphoric pronouns that compose them are predicate invariant. Their anaphoricity is characterised in addition by a-conservativity (a-intersectivity, a-co-intersectivity), a property which is not displayed by "ordinary" determiners.

The results presented in this paper show that though the existence of anaphoric determiners extends the expressive power of NLs because the functions they denote lie outside the class of generalised quantifiers classically defined, these functions resemble quantifiers denoted by "ordinary" nominal determiners in certain important ways.

References

Barwise, J., Cooper, R.: Generalised quantifiers and natural language. Linguist. Philos. **4**, 159–219 (1981)

Dalrymple, M. et al.: Reciprocal expressions and the concept of reciprocity. Linguist. Philos. **21**, 151–210 (1998)

Keenan, E.L.: On the denotations of anaphors. Res. Lang. Comput. **5**(1), 5–17 (2007)

Keenan, E.L., Westerståhl, D.: Generalized quantifiers in linguistics and logic. In: van Benthem, J., ter Meulen, A. (eds.) Handbook of Logic and Language, pp. 837–893. Elsevier, Amsterdam (1997)

Keenan, E.L.: Natural language, sortal reducibility and generalised quantifiers. J. Symbolic logic **58**(1), 314–325 (1993)

Keenan, E.L., Moss, L.: Generalized quantifiers and the expressive power of natural language. In: van Benthem, J., ter Meulen, A. (eds.) Generalized Quantifiers, Foris, Dordrecht, pp. 73–124 (1985)

Zuber, R.: On the semantics of exclusion and inclusion phrases. In: Strolovitch, D., Lawson, A. (eds.) Proceedings of SALT 9, pp. 267–283. LSA Publications (1998)

Zuber, R.: Semantics of Slavic anaphoric determiners. In: Cormany, Ed., et al. (eds.) Proceedings of SALT 19, pp. 464–477 (2009)

Zuber, R.: Generalising conservativity. In: Dawar, A., Queiroz, R. (eds.) WoLLIC 2010. LNCS (LNAI), vol. 6188, pp. 247–258. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13824-9_20

Zuber, R.: Semantic constraints on anaphoric determiners. Res. Lang. Comput. **8**, 255–271 (2010b)
Zuber, R.: Generalising predicate and argument invariance. In: Asher, N., Soloviev, S. (eds.) LACL 2014. LNCS, vol. 8535, pp. 163–176. Springer, Heidelberg (2014)
Zuber, R.: Anaphors and quantifiers. In: Väänänen, J., Hirvonen, Å., de Queiroz, R. (eds.) WoLLIC 2016. LNCS, vol. 9803, pp. 432–445. Springer, Heidelberg (2016). doi:10.1007/978-3-662-52921-8_26

Logic and Computation

The Topology of Full and Weak Belief

Alexandru Baltag[1], Nick Bezhanishvili[1], Aybüke Özgün[1,2(✉)], and Sonja Smets[1]

[1] ILLC, University of Amsterdam, Amsterdam, The Netherlands
thealexandrubaltag@gmail.com, ozgunaybuke@gmail.com
{n.bezhanishvili,s.j.l.smets}@uva.nl
[2] LORIA, CNRS - Université de Lorraine, Nancy, France

Abstract. We introduce a new topological semantics for belief logics in which the belief modality is interpreted as the interior of the closure of the interior operator. We show that the system **wKD45**, a weakened version of **KD45**, is sound and complete with respect to the class of all topological spaces. While generalizing the topological belief semantics proposed in [1,2] to all spaces, we model conditional beliefs and updates and give complete axiomatizations of the corresponding logics with respect to the class of all topological spaces.

Keywords: Topological models · Epistemic and doxastic logic · Updates · Conditional beliefs · (hereditarily) Extremally disconnected spaces

1 Introduction

Understanding the relation between knowledge and belief is an issue of central importance in formal epistemology. Especially after the birth of the knowledge-first epistemology in [36], the question of what exactly distinguishes an item of belief from an item of knowledge and how one can be defined in terms of the other has become even more pertinent. This problem has been tackled from two rather opposite perspectives in the literature. On the one hand, there has been proposals in the line of justified true belief account of knowledge (JTB), accepting the conceptual priority of belief over knowledge. According to this approach, one starts with a weak notion of belief (which is at least justified and true) and tries to reach knowledge by making the chosen notion of belief stronger in such a way that the defined notion of knowledge would no longer be subject to Gettier-type counterexamples [15]. Among this category, we can mention the conception of knowledge as *correctly justified belief*: *not only the content of belief has to*

A. Özgün—Acknowledges support from European Research Council grant EPS 313360.

S. Smets—Contribution to this paper has received funding from the European Research Council under the European Community's 7th Framework Programme/ERC Grant agreement no. 283963.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 205–228, 2017.
DOI: 10.1007/978-3-662-54332-0_12

be true, but its justification has to be correct. This approach can be formalized via topologies under the *interior-based semantics* (see, e.g., Sect. 2.2). Other responses falling under the first category include the *defeasibility analysis of knowledge* [20,21], the *sensitivity account* [24], the *contextualist account* [12] and the *safety account* [29][1].

The second perspective, on the other hand, challenges the 'conceptual priority of belief over knowledge' [36] and reverts the relation by giving priority to knowledge. When knowledge has priority, other attitudes (e.g. beliefs) should be explainable or definable in terms of it. One of the few philosophers who has worked out a formal system that ties in with this second approach is Stalnaker. In [30], Stalnaker uses a relational semantics for knowledge based on *reflexive, transitive and directed* Kripke models. In his work, he analyses the relation between knowledge and belief and builds a combined modal system for these notions with the axioms extracted from this analysis. He intends to capture a strong notion of belief based on the conception of 'subjective certainty'

$$B\varphi \rightarrow BK\varphi$$

meaning that *believing implies believing that one knows* [30, p. 179]. Stalnaker refers to this concept as 'strong belief', but following our previous work [1,2] we prefer to call it *full belief.* In fact, the above axiom holds biconditionally in his system and belief therefore becomes *subjectively indistinguishable from knowledge: an agent (fully) believes φ iff she (fully) believes that she knows φ* [1,2]. Moreover, Stalnaker argues that the 'true' logic of knowledge is **S4.2** and that (full) belief can be defined as the *epistemic possibility of knowledge.* More precisely,

$$B\varphi = \neg K \neg K\varphi$$

meaning that *an agent believes φ iff she does't know that she does't know φ*[2]. He moreover states that his system embeds the logic of belief **KD45** when B is defined as $\langle K\rangle K$[3] (and K is an **S4.2** modality).

In [1,2] Stalnaker's semantics was generalized from a relational setting to a topological setting. In particular, a topological semantics was given for full belief extending the interior semantics for knowledge with a semantic clause for the belief modality via the *closure of the interior* operator and it was shown that the proposed semantics on extremally disconnected spaces constitutes the canonical (most general) semantics for Stalnaker's axiom. In this way, Stalnaker's formalization was generalized by making it independent from its relational semantics. [1,2] focused on the unimodal cases for knowledge and belief and proved that while the knowledge logic of extremally disconnected spaces under the interior-based semantics is indeed **S4.2**, its belief logic under the proposed topological semantics is **KD45**. In this paper (Sect. 3), we give a brief overview of the work

[1] For an overview of responses to the Gettier challenge and a detailed discussion, we refer the reader to [18,27].

[2] For a more detailed discussion on Stalnaker's approach, we refer the reader to [2].

[3] $\langle K\rangle$ denotes the dual of K, i.e., $\neg K\neg\varphi := \langle K\rangle\varphi$.

done in [1,2]. We refer to [1,2] for a more detailed discussion. This framework, however, comes with a problem when extended to a dynamic setting by adding update modalities in order to capture the action of learning (conditioning with) new 'hard' (true) information P, as also elaborated in [2]. Conditioning with new 'hard' (true) information P is commonly modelled by deleting the 'non-P' worlds from the initial model. Its natural topological analogue, as recognized in [5,6,38] (among others) and also applied in [2], is a topological update operator, using the restriction of the original topology to the subspace induced by the set P. In order for this interpretation to be successfully implemented, the subspace induced by the new information P should possess the same structural properties as the initial topology that renders the axioms of the underlying knowledge/belief system sound. More precisely, we demand the subspace induced by the new information P to be in the class of structures with respect to which the (static) knowledge/belief logics in questions are sound and complete. However, extremally disconnectedness is not a hereditary property. In other words, it is not guaranteed that an arbitrary subspace of a given extremally disconnected space is extremally disconnected. Therefore, the aforementioned topological interpretation of conditioning with true, hard information cannot be implemented on extremally disconnected spaces. In [2], we present a solution for this problem by modelling updates on the topological spaces whose every subspace is extremally disconnected, i.e., by modelling updates on hereditarily extremally disconnected spaces.

In this paper, we propose another solution for this problem via arbitrary topological spaces. More precisely, we do it by introducing a topological semantics for belief based on *all* topological spaces in terms of the *interior of the closure of the interior operator*. It is important that this semantics coincides with the topological belief semantics introduced in [1,2] on extremally disconnected space, thus, we here generalize the semantics proposed in [1,2] to all topological spaces. Further, while the complete logic of knowledge is actually **S4** (McKinsey and Tarski [23]), we show that the complete logic of belief is a weaker system than **KD45**, namely the logic **wKD45**. The latter result follows by translating **S4** fully and faithfully into **wKD45**. The restriction of this translation to **S4.2** coincides with Stannaker's embedding of **KD45** into **S4.2**. We also formalize a notion of conditional belief $B^{\varphi}\psi$ by *relativizing* the semantic clause for simple belief modality to the extension of the learnt formula φ. We moreover formalize updates $\langle !\varphi \rangle \psi$ again as a topological update operator using the *restriction* of the initial topology to its subspace induced by the new information φ and show that we no longer encounter the problem about updates risen in the case of extremally disconnected spaces: *updates on all topological spaces are 'well-behaved'.*

We note that the interior of the closure of the interior operator is also interesting from a purely topological point of view. This operation can be seen as a *regularization* of an open set. Geometrically this operation 'patches up cracks' of an open region (see Sect. 4.1 for more details on this as well as for an epistemic interpretation of this operation). Furthermore, from a purely syntactical point of view, part of our work can be seen as studying the $B := K\langle K \rangle K$-fragment of the system **S4** for K and providing a complete axiomatization for this modality (which is interpreted as belief (B) in this particular setting). Given that our work is inspired by Stalnaker's [30], one natural question to ask is why we are interested in the $K\langle K\rangle K$-fragment of **S4** rather than the $\langle K\rangle K$-fragment as a belief system. In fact, the latter approach, namely logics of belief as *epistemic possibility of knowledge*, stemming from knowledge modalities of different strength, has been of interest in recent years. Klein et al. [19] investigate this fragment when K is not positively introspective, more precisely, when K is of type **KT.2**. To the best of our knowledge, finding a complete axiomatization of the $\langle K\rangle K$-fragment of **S4** is still an open and interesting question from a proof theoretical perspective. However, we know that it is neither a normal modal logic nor does it include the (D)-axiom. It therefore does not form a 'good' logic of belief in this particular setting with highly idealized agents. The $K\langle K\rangle K$-fragment on the other hand is equivalent to the $\langle K\rangle K$-fragment when K is of **S4.2** type. Moreover, it is the only non-empty, positive modality that is normal in **S4** and not equivalent to the knowledge modality K (see e.g., [10, Ex. 3.14, p. 102]). Hence, it is the only alternative for Stalnaker's belief as *subjective certainty* that can satisfy most of the standard axioms of belief.

The paper is structured as follows. In Sect. 2 we introduce the topological preliminaries used in this paper and present the interior-based topological semantics as well as its connection to the standard Kripke semantics and to the topological interpretation of knowledge. Section 3 gives a brief overview of the previous related work. Sections 4 and 5 constitute the main parts of this paper: while the former presents a topological semantics for belief based on all topological spaces, the latter is concerned with the topological interpretation of conditional beliefs and updates. In Sect. 6 we conclude by giving a summary of our results and pointing out a number of directions for future research.

2 Background

2.1 Topological Preliminaries

In this section, we introduce the basic topological concepts that will be used throughout this paper. For more detailed discussion we refer the reader to [13,14].

A ***topological space*** is a pair (X, τ), where X is a non-empty set and τ is a family of subsets of X containing X and $\emptyset$ and is closed under finite intersections and arbitrary unions. The set X is called a ***space***. The subsets of X belonging to τ are called ***open sets*** (or ***opens***) in the space; the family τ of open subsets of X is called a ***topology*** on X. Complements of open sets are called ***closed sets***. An open set containing $x \in X$ is called an ***open neighbourhood*** of x. The ***interior*** $\text{Int}(A)$ of a set $A \subseteq X$ is the largest open set contained in A whereas the ***closure*** $\text{Cl}(A)$ of A is the least closed set containing A. It is easy to see that Cl is the De Morgan dual of Int (and vice versa) and can be written as $\text{Cl}(A) = X \backslash \text{Int}(X \backslash A)$. Moreover, the set of boundary points of a set $A \subseteq X$, denoted by $\text{Bd}(A)$, is defined as $\text{Bd}(A) = \text{Cl}(A) \backslash \text{Int}(A)$.

2.2 The Interior Semantics for Modal (Epistemic) Logic

In this section we provide the formal background for the aforementioned interior-based topological semantics for modal (epistemic) logic that originated in the work of McKinsey and Tarski [23]. Moreover, we present important completeness results concerning logics of knowledge **S4**, **S4.2** and **S4.3** based on the interior semantics, explain the connection between the interior and standard Kripke semantics, and focus on the topological (evidence-based) interpretation of knowledge.

Syntax. We consider the standard unimodal (epistemic) language $\mathcal{L}_K$ with a countable set of propositional letters Prop, Boolean operators $\neg$ and $\wedge$ and a modal operator K. Formulas of $\mathcal{L}_K$ are defined as usual by the following grammar

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi$$

where $p \in$ Prop. Abbreviations for the connectives $\vee$, $\rightarrow$, $\leftrightarrow$ are standard. Moreover, the existential modal operator $\langle K\rangle$ and $\bot$ are defined as $\langle K\rangle\varphi := \neg K\neg\varphi$ and $\bot := p \wedge \neg p$,

Semantics. Given a topological space (X, τ), we define a ***topological model*** (or simply a ***topo-model***) as $\mathcal{M} = (X, \tau, \nu)$ where $\nu : \text{Prop} \rightarrow \mathcal{P}(X)$ is a valuation function.

Definition 1. *Given a topo-model $\mathcal{M} = (X, \tau, \nu)$, we define the **interior semantics** for the language $\mathcal{L}_K$ recursively as:*

$$\begin{array}{lll} \mathcal{M}, x \models p & \textit{iff} & x \in \nu(p) \\ \mathcal{M}, x \models \neg\varphi & \textit{iff} & \textit{not } \mathcal{M}, x \models \varphi \\ \mathcal{M}, x \models \varphi \wedge \psi & \textit{iff} & \mathcal{M}, x \models \varphi \textit{ and } \mathcal{M}, x \models \psi \\ \mathcal{M}, x \models K\varphi & \textit{iff} & (\exists U \in \tau)(x \in U \wedge \forall y \in U,\ \mathcal{M}, y \models \varphi) \end{array}$$

where $p \in$ Prop[4].

We let $[\![\varphi]\!]^{\mathcal{M}} = \{x \in X \mid \mathcal{M}, x \models \varphi\}$ denote the ***extension*** of a modal formula φ in a topo-model $\mathcal{M}$, i.e., the *extension* of a formula φ in a topo-model $\mathcal{M}$ is defined as the set of points in $\mathcal{M}$ satisfying φ. We skip the index when it is clear in which model we are working. It is now easy to see that $[\![K\varphi]\!] = \text{Int}([\![\varphi]\!])$ and $[\![\langle K\rangle\varphi]\!] = \text{Cl}([\![\varphi]\!])$. We use this extensional notation throughout the paper as it makes clear the fact that the modalities, K and $\langle K\rangle$, are interpreted in terms of specific and *natural* topological operators. More precisely, K and $\langle K\rangle$ are modelled as the *interior* and the *closure* operators, respectively.

We say that φ is true in a topo-model $\mathcal{M} = (X, \tau, \nu)$ if $[\![\varphi]\!]^{\mathcal{M}} = X$, and that φ is valid in (X, τ) if $[\![\varphi]\!]^{\mathcal{M}} = X$ for all topo-models $\mathcal{M}$ based on (X, τ), and finally we say that φ is valid in a class of topological spaces if φ is valid in every member of the class [33]. Soundness and completeness with respect to the interior semantics are defined as usual.

4 Originally, McKinsey and Tarski [23] introduce the interior semantics for the basic modal language. Since we talk about this semantics in the context of *knowledge*, we use the basic *epistemic* language.

Theorem 1 **([23]).** **S4** *is sound and complete with respect to the class of all topological spaces under the interior semantics.*

Topological interpretation of knowledge: open sets as pieces of evidences. One of the reasons as to why the interior operator is interpreted as knowledge is that the Kuratowski properties (see, e.g., [13,14]) of the interior operator amount to **S4** axioms written in topological terms. This implies that (as we can also read from Theorem 1), *topologically*, knowledge is *Truthful*

$$K\varphi \to \varphi,$$

Positively Introspective

$$K\varphi \to KK\varphi,$$

but *not necessarily* Negatively Introspective

$$\neg K\varphi \to K\neg K\varphi.$$

From a philosophical point of view, the principle of Negative Introspection is arguably the most controversial axiom regarding the characterization of knowledge. It leads to some undesirable consequences, such as Voorbraak's paradox (see e.g., [1,35]), and is rejected by some prominent people in the field such as Hintikka [17], Lenzen [22], Stalnaker [30] (among others).

Another argument in favour of *knowledge as the interior operator* conception is of a more 'semantic' nature: the interior semantics provides a deeper insight into the evidence-based interpretation of knowledge. We can interpret opens in a topological model as 'pieces of evidence' and, in particular, open neighborhoods of a state x as the pieces of *true* (*sound, correct*) evidence that are observable by the agent at state x. If an open set U is included in the extension of a proposition φ in a topo-model $\mathcal{M}$, i.e. if $U \subseteq [\![\varphi]\!]^{\mathcal{M}}$, we say that *the piece of evidence* U *entails* (*supports, justifies*) *the proposition* φ. Recall that, for any topo-model $\mathcal{M} = (X, \tau, \nu)$, any $x \in X$ and any $\varphi \in \mathcal{L}_K$, we have

$$x \in [\![K\varphi]\!]^{\mathcal{M}} \text{ iff } (\exists U \in \tau)(x \in U \wedge U \subseteq [\![\varphi]\!]^{\mathcal{M}}).$$

Thus, taking open sets as pieces of evidence and in fact open neighbourhoods of a point x as *true* pieces of evidence (that the agent can observe at x), we obtain the following evidence-based interpretation for knowledge: *the agent knows* φ *iff she has a true piece of evidence* U *that justifies* φ. In other words, *knowing* φ is the same as *having a correct justification for* φ. The necessary and sufficient conditions for one's belief to qualify as knowledge consist in it being not only truthful, but also in having a correct (evidential) justification. Therefore, the interior semantics implements the widespread intuitive response to Gettier's challenge: knowledge is *correctly* justified belief (rather than being simply true justified belief) [1,2].

Connection between Kripke frames and topological spaces. The interior semantics is closely related to the standard Kripke semantics of **S4** (and of its normal extensions): every reflexive and transitive Kripke frame corresponds to a special kind of (namely, Alexandroff) topological spaces.

Let us now fix some notation and terminology. We denote a ***Kripke frame*** by $\mathcal{F} = (X, R)$, a ***Kripke model*** by $M = (X, R, \nu)$ and $\|\varphi\|^M$ denotes the *extension* of a formula φ in a Kripke model $M = (X, R, \nu)$[5]. A topological space (X, τ) is called ***Alexandroff*** if τ is closed under arbitrary intersections, i.e., $\bigcap \mathcal{A} \in \tau$ for any $\mathcal{A} \subseteq \tau$. Equivalently, a topological space (X, τ) is Alexandroff iff every point in X has a least neighborhood. As mentioned, there is a one-to-one correspondence between reflexive and transitive Kripke frames and Alexandroff spaces. More precisely, given a reflexive and transitive Kripke frame $\mathcal{F} = (X, R)$, we can construct a topological space, indeed an Alexandroff space, $\mathcal{X} = (X, \tau_R)$ by defining τ_R to be the set of all upsets[6] of $\mathcal{F}$. Moreover, the evaluation of modal formulas in a reflexive and transitive Kripke model coincides with their evaluation in the corresponding (Alexandroff) topological space (see e.g., [26, p. 306]). As a result of this connection, the Kripke completeness of the normal extensions of **S4** implies topological completeness under the interior semantics (see, e.g., [33]).

Normal extensions of S4: the logics S4.2 and S4.3. There are two other knowledge systems, namely **S4.2** and **S4.3** , that are of particular interest for us. Both **S4.2** and **S4.3** are strengthenings of **S4** which are defined as

$$\mathbf{S4.2} := \mathbf{S4} + \langle K \rangle K\varphi \to K\langle K \rangle \varphi, \text{ and}$$
$$\mathbf{S4.3} := \mathbf{S4} + K(K\varphi \to \psi) \vee K(K\psi \to \varphi)$$

where $\mathbf{L} + \varphi$ denotes the smallest logic containing $\mathbf{L}$ and φ.

We recall that a topological space (X, τ) is ***extremally disconnected*** if the closure of every open subset of X is open and it is ***hereditarily extremally disconnected*** if every subspace of (X, τ) is extremally disconnected. We here would like to remind that extremally disconnectedness is, in general, not a hereditary property[7].

Theorem 2 (cf. [33]). S4.2 *is sound and complete with respect to the class of extremally disconnected spaces under the interior semantics.*

Theorem 3 ([2,7]). S4.3 *is sound and complete with respect to the class of hereditarily extremally disconnected spaces under the interior semantics.*

We note that the completeness parts of Theorems 2 and 3 follow from the Kripke completeness of **S4.2** and **S4.3** (which is a direct consequence of the

[5] The reader who is not familiar with the standard Kripke semantics is referred to [8,11] for an extensive introduction to the topic.

[6] A set $A \subseteq X$ is called an *upset* of (X, R) if for each $x, y \in X$, xRy and $x \in A$ imply $y \in A$.

[7] A topological property is said to be *hereditary* if for any topological space (X, τ) that has the property, every subspace of (X, τ) also has it [14, p. 68].

Sahlqvist theorem) and the fact that Alexandroff spaces corresponding to transitive, reflexive and directed Kripke frames (**S4.2**-frames) are extremally disconnected and Alexandroff spaces corresponding to reflexive and transitive Kripke frames with no branching to the right (**S4.3**-frames) are hereditarily extremally disconnected. The soundness with respect to the topological semantics, however, needs some argumentation. The detailed proofs can be found in [33, p. 253] and [7, Proposition 3.1]. The logical counterpart of the fact that extremally disconnected spaces (**S4.2**-spaces) are not closed under subspaces is that **S4.2** is not a subframe logic [10, Sect. 9.4]. The logical counterpart of the fact that hereditarily extremally disconnected spaces (**S4.3**-spaces) are extremally disconnected spaces closed under subspaces is that the subframe closure of **S4.2** is **S4.3**, see [37, Sect. 4.7]. For examples of extremally disconnected and hereditarily extremally disconnected spaces, we refer to [2,7,28].

3 The Topology of Full Belief: Overview of [1]

3.1 Stalnaker's Combined Logic of Knowledge and Belief

In his paper [30], Stalnaker focuses on the properties of knowledge and belief and the relation between the two and he approaches the problem of understanding the concrete relation between knowledge and belief from an unusual perspective. Unlike most research in the formal epistemology literature, he starts with a chosen notion of knowledge and weakens it to obtain belief. He bases his analysis on a conception of belief as 'subjective certainty': *from the point of the agent in question, her belief is subjectively indistinguishable from her knowledge* [1]. In this section, we briefly recall Stalnaker's proposal of the 'true' logic of knowledge and belief. Throughout this paper, following [1,2,25], we will refer to Stalnaker's notion as 'full belief'.

The *bimodal language* $\mathcal{L}_{KB}$ of *knowledge and (full) belief* is obtained by extending $\mathcal{L}_K$ by a belief modality B:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi \mid B\varphi.$$

We define the doxastic possibility modality $\langle B\rangle\varphi$ by $\neg B\neg\varphi$. We call Stalnaker's system, given in the following Table 1, **KB**.

We refer to [1,2,25] for a discussion on the axioms of **KB** and continue with some conclusions of philosophical importance derived by Stalnaker in [30] and stated in the following proposition:

Proposition 1 (Stalnaker [30]). *The following equivalence is provable in the system* **KB***:*

$$B\varphi \leftrightarrow \langle K\rangle K\varphi. \tag{1}$$

Moreover, the axioms (K) $B(\varphi \to \psi) \to (B\varphi \to B\psi)$*,* (D) $B\varphi \to \langle B\rangle\varphi$*,* (4) $B\varphi \to BB\varphi$*,* (5) $\neg B\varphi \to B\neg B\varphi$ *of the system* **KD45** *and the* (.2)*-axiom* $\langle K\rangle K\varphi \to K\langle K\rangle\varphi$ *of the system* **S4.2** *are provable in* **KB***.*

Proposition 1 thus shows that full belief is definable in terms of knowledge as 'epistemic possibility of knowledge' via equivalence (1), the 'true' logic of belief is **KD45** and the 'true' logic of knowledge is **S4.2** (see [2] for the proof).

Table 1. Stalnaker's System **KB**

	Stalnaker's axioms	
(K)	$K(\varphi \to \psi) \to (K\varphi \to K\psi)$	Knowledge is additive
(T)	$K\varphi \to \varphi$	Knowledge implies truth
(KK)	$K\varphi \to KK\varphi$	Positive introspection for K
(CB)	$B\varphi \to \neg B\neg\varphi$	Consistency of belief
(PI)	$B\varphi \to KB\varphi$	(Strong) positive introspection of B
(NI)	$\neg B\varphi \to K\neg B\varphi$	(Strong) negative introspection of B
(KB)	$K\varphi \to B\varphi$	Knowledge implies Belief
(FB)	$B\varphi \to BK\varphi$	Full Belief
	Inference rules	
(MP)	From φ and $\varphi \to \psi$ infer ψ	Modus Ponens
(K-Nec)	From φ infer $K\varphi$	Necessitation

3.2 The Topological Semantics of Full Belief

In [1,2,25], a topological semantics for full belief and knowledge is proposed by extending the interior semantics for knowledge with a semantic clause for belief. The belief modality B is interpreted as the *closure of the interior operator* on *extremally disconnected spaces*. Several topological soundness and completeness results for both bimodal and unimodal cases, in particular for **KB** and **KD45**, with respect to the proposed semantics are also proved. We now briefly overview the topological semantics for full belief introduced in [1,2,25] and state the completeness results. The proofs can be found in [2,25].

Definition 2 (Topological Semantics for Full Belief and Knowledge). *Given a topo-model $\mathcal{M} = (X, \tau, \nu)$, the semantics for the formulas in $\mathcal{L}_{KB}$ is defined for Boolean cases and $K\varphi$ the same way as in the interior semantics. The semantics for $B\varphi$ is defined as*

$$[\![B\varphi]\!]^{\mathcal{M}} = \mathrm{Cl}(\mathrm{Int}([\![\varphi]\!]^{\mathcal{M}})).$$

Truth and validity of a formula, soundness and completeness are defined the same way as in the interior semantics.

Proposition 2. *A topological space validates all the axioms and rules of Stalnaker's system* **KB** *(under the semantics given above) iff it is extremally disconnected.*

Theorem 4. *The sound and complete logic of knowledge and belief on extremally disconnected spaces is given by Stalnaker's system* **KB**.

Besides, as far as full belief is concerned, the above topological semantics constitutes the *most general extensional semantics* for Stalnaker's system **KB** [1,2,25]. Moreover, Stalnaker's combined logic of knowledge and belief yields the system **S4.2** as the unimodal logic of knowledge and the system **KD45** as the unimodal logic of belief (see Proposition 1). It has been already proven that **S4.2** is complete with respect to the class of extremally disconnected spaces under the interior semantics. This raises the question of topological soundness and completeness for **KD45** under the proposed semantics for belief in terms of the *closure and the interior operator*:

Theorem 5 ([1,2,25]). **KD45** *is sound and complete with respect to the class of extremally disconnected spaces under the topological belief semantics.*

Theorem 5 therefore shows that the belief logic of extremally disconnected spaces is **KD45** when B is interpreted as the closure of the interior operator. These results on extremally disconnected spaces, however, encounter problems when extended to a dynamic setting by adding update modalities formalized as model restriction by means of subspaces.

Topological Semantics for Update Modalities. We now consider the language $\mathcal{L}_{!KB}$ obtained by adding to the language $\mathcal{L}_{KB}$ (existential) dynamic update modalities $\langle !\varphi\rangle\psi$ meaning that φ *is true and after the agent learns* φ, ψ *becomes true*. As also observed in [5,6,38], the topological analogue of updates corresponds to taking the restriction of a topology τ on X to a subset $P \subseteq X$, i.e., it corresponds to the restriction of the original topology to its subspace induced by the new, true information P.

Given a topological space (X, τ) and a non-empty set $P \subseteq X$, a space (P, τ_P) is called a ***subspace*** of (X, τ) where $\tau_P = \{U \cap P : U \in \tau\}$.

For a topo-model (X, τ, ν) and $\varphi \in \mathcal{L}_{!KB}$, we denote by $\mathcal{M}_\varphi$ the ***restricted model*** $\mathcal{M}_\varphi = (\llbracket\varphi\rrbracket, \tau_{\llbracket\varphi\rrbracket}, \nu_{\llbracket\varphi\rrbracket})$ where $\llbracket\varphi\rrbracket = \llbracket\varphi\rrbracket^{\mathcal{M}}$ and $\nu_{\llbracket\varphi\rrbracket}(p) = \nu(p) \cap \llbracket\varphi\rrbracket$ for any $p \in$ Prop. Then, the semantics for the dynamic language $\mathcal{L}_{!KB}$ is obtained by extending the semantics for $\mathcal{L}_{KB}$ with:

$$\llbracket\langle !\varphi\rangle\psi\rrbracket^{\mathcal{M}} = \llbracket\psi\rrbracket^{\mathcal{M}_\varphi}.$$

To explain the problem: Given that the underlying *static* logic of knowledge and belief is the logic of extremally disconnected spaces (see e.g., Theorems 2, 4 and 5) and extremally disconnectedness is not inherited by arbitrary subspaces, we cannot guarantee that the restricted model induced by an arbitrary formula φ remains extremally disconnected. Under the topological belief semantics, both the (K)-axiom (also known as the axiom of *Normality*) $B(\varphi \wedge \psi) \leftrightarrow (B\varphi \wedge B\psi)$ and the (D)-axiom (also named as the *Consistency of Belief*) $B\varphi \rightarrow \langle B\rangle\varphi$ characterize extremally disconnected spaces [2,25]. Therefore, if the restricted model is not extremally disconnected, the agent comes to have inconsistent beliefs after an update with true information: the formula $B\varphi \wedge B\neg\varphi$ is satisfiable in a non-extremally disconnected topo-model. For an example illustrating this problem, we refer to [2, p. 21].

One possible solution for this problem is a further limitation on the class of spaces we work with: we can restrict our attention to *hereditarily extremally disconnected spaces*, thereby, we guarantee that no model restriction leads to inconsistent beliefs. As the logic of hereditarily extremally disconnected spaces under the interior semantics is **S4.3**, the underlying static logic, in this case, would consist in **S4.3** as the logic of knowledge but again **KD45** as the logic of belief. In [2], we examine this solution. In this paper, we present another solution which approaches the issue from the opposite direction: we propose to work with all topological spaces instead of working with a restricted class. This solution, unsurprisingly, leads to a weakening of the underlying static logic of knowledge and belief. As we already mentioned earlier, it is a classic result that the knowledge logic of all topological spaces is **S4** and here we will explore the (weak) belief logic of all topological spaces under the topological belief semantics.

4 The Topology of Weak Belief

4.1 Topological Semantics of Weak Belief

Recall that given an extremally disconnected space (X, τ), we have

$$\mathrm{Cl}(\mathrm{Int}(A)) = \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}(A)))$$

for any $A \subseteq X$. Hence, given a topo-model $\mathcal{M} = (X, \tau, \nu)$, the semantic clause for the belief modality can be written in the following equivalent forms

$$[\![B\varphi]\!]^{\mathcal{M}} \overset{(1)}{=} \mathrm{Cl}(\mathrm{Int}([\![\varphi]\!]^{\mathcal{M}})) \overset{(2)}{=} \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}([\![\varphi]\!]^{\mathcal{M}})))$$

if (X, τ) is an extremally disconnected space. However, $\mathrm{Cl}(\mathrm{Int}(A)) = \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}(A)))$ is not always the case for all topological spaces and all $A \subseteq X$; the equation demands the restriction to extremally disconnected spaces. Besides, if we evaluate B as the closure of the interior operator on *all* topological spaces, we obtain that neither the (K)-axiom nor the (D)-axiom is sound. Syntactically speaking, B defined as $\langle K\rangle K$ does not yield a 'good' logic of belief when K is an **S4**-type modality: $\langle K\rangle K$ is neither normal nor does satisfy the (D)-axiom. Moreover, purely **S4**-type knowledge could not have been what Stalnaker had in mind while considering B as $\langle K\rangle K$ since this would violate his principles (CB) and (PI). Moreover, given that K is interpreted as the interior operator on topological spaces, equation (1) makes the schema $B\varphi \leftrightarrow \langle K\rangle K\varphi$ and equation (2) makes the schema $B\varphi \leftrightarrow K\langle K\rangle K\varphi$ valid on all topological spaces. While **S4.2** $\vdash \langle K\rangle K\varphi \leftrightarrow K\langle K\rangle K\varphi$, we have **S4** $\nvdash \langle K\rangle K\varphi \leftrightarrow K\langle K\rangle K\varphi$ and B as $K\langle K\rangle K$ is the only alternative holding the property of being equivalent to $\langle K\rangle K$ in **S4.2** and being not equivalent to $\langle K\rangle K$ in **S4**. Moreover, $K\langle K\rangle K$ is the only non-empty and positive modality that is normal and is *not* equivalent to knowledge in **S4** [10, Ex. 3.14, p. 102]. Therefore, a notion of belief that works well on all topological spaces and coincides with Stalnaker's belief as subjective certainty on extremally disconnected spaces demands the alternative interpretation of belief in terms of the interior of the closure of the interior operator.

We thus concentrate on the latter equation: we interpret B as the *interior of the closure of the interior* operator on all topological spaces.

Semantics. Let $\mathcal{M} = (X, \tau, \nu)$ be a topo-model. The semantic clauses for the propositional variables and the Boolean connectives are the same as in the interior semantics. For the modal operator B, we put

$$[\![B\varphi]\!]^{\mathcal{M}} \quad = \quad \text{Int}(\text{Cl}(\text{Int}([\![\varphi]\!]^{\mathcal{M}})))$$

and the semantic clause for $\langle B \rangle$ is easily obtained as

$$[\![\langle B\rangle\varphi]\!]^{\mathcal{M}} \quad = \quad \text{Cl}(\text{Int}(\text{Cl}([\![\varphi]\!]^{\mathcal{M}}))).$$

Validity of a formula is defined as usual. We call this semantics **w**-*topological belief semantics* referring to the system **wKD45** for which we will prove soundness and completeness. This way we distinguish it from the *topological belief semantics* presented in Sect. 3.2 w.r.t. to which we proved the soundness and completeness of the system **KD45**.

Our new topological interpretation of belief also comes with intrinsic philosophical motivation that fits well with the topologically defined notions of *closeness* and *small/negligible sets*. To elaborate, it is well-known that the closure operator represents a topological conception of 'closeneess'. Intuitively speaking, we can read $x \in \text{Cl}(A)$ as *x is very close to the set A*, i.e., *it cannot be sharply distinguished from the elements of A via an open set*. Therefore, recalling that K is interpreted as the interior modality, according to the semantics for (full) belief in terms of the closure and the interior operator introduced in Sect. 3.2, 'the agent fully believes φ at a state x iff she cannot sharply distinguish x from the worlds in which she has knowledge of φ' [2, p. 24]. Therefore, *full belief is very close to knowledge* when 'close' is interpreted topologically [2]. This interpretation in fact captures the notion of belief as *subjective certainty*. Our new interpretation of belief in terms of the interior of the closure of the interior operator $[\![B\varphi]\!] = \text{Int}(\text{Cl}([\![K\varphi]\!]))$, on the other hand, makes the connection between these two notions even stronger: the belief operator interpreted this way comes even *closer* to knowledge, yet does not coincide with it. According to the new interpretation of belief, the agent believes φ at a state x iff there exists an open neighbourhood U of x such that $U \subseteq \text{Cl}([\![K\varphi]\!])$. This implies, since $[\![K\varphi]\!]$ is open, that $x \in U \subseteq [\![K\varphi]\!] \cup \text{Bd}([\![K\varphi]\!])$, where $\text{Bd}([\![K\varphi]\!])$ is the set of boundary points of $[\![K\varphi]\!]$. As $U \subseteq [\![K\varphi]\!] \cup \text{Bd}[\![K\varphi]\!]$, the set $U \cap [\![\neg K\varphi]\!] = U \cap \text{Bd}([\![K\varphi]\!])$ and it is possibly non-empty. Thus, it is still not guaranteed that the agent can distinguish the states in which she knows φ from the ones in which she does not. However,

$$U \cap [\![\neg K\varphi]\!] = U \cap \text{Bd}([\![K\varphi]\!]) \subseteq \text{Bd}([\![K\varphi]\!])$$

and, since $[\![K\varphi]\!] = \text{Int}[\![\varphi]\!]$ is an open set, $\text{Bd}([\![K\varphi]\!])$ is nowhere dense[8]. Therefore, $U \cap [\![\neg K\varphi]\!]$ is also nowhere dense. As nowhere dense sets constitute one of the topological notions of 'small, negligible sets' and $U \subseteq [\![K\varphi]\!] \cup \text{Bd}([\![K\varphi]\!])$, we can

[8] A subset $A \subseteq X$ is called *nowhere dense* in (X, τ) if $\text{Int}(\text{Cl}(A)) = \emptyset$.

say that *the agent believes φ at x iff she can almost sharply distinguish x from the states in which she does not know φ*. The part of U that is consistent with $\neg K\varphi$ is topologically negligibly small. We therefore further argue that this is the "closest-to-knowledge" notion of belief that can be defined in terms of the topological tools at hand and that is not identical with the notion of knowledge taken as the primitive operator.

Topologically, our new belief operator behaves like a 'regularization' operator for the opens in a topology. Given a topological space (X, τ), we can define $B : \tau \to \tau$ such that $B(U) = \mathrm{Int}(\mathrm{Cl}(U))$. Therefore, B takes an open set and makes it *regular open*[9]. In fact, for any open set $U \in \tau$, the set $\mathrm{Int}(\mathrm{Cl}(U))$ is the smallest regular open such that $U \subseteq \mathrm{Int}(\mathrm{Cl}(U))$[10]. Therefore, this operator extends an open set in a minimal way by gluing its holes and cracks together. To illustrate, consider the natural topology on the real line $(\mathbb{R}, \tau)$ and let $P = [-2, 3) \cup (3, 5) \cup \{7\}$ (see Fig. 1).

Fig. 1. $(\mathbb{R}, \tau)$

We have $\mathrm{Int}(P) = (-2, 3) \cup (3, 5)$ and $\mathrm{Int}(P)$ is not regular open. However, $B(\mathrm{Int}(P)) = \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}(P))) = (-2, 5)$, which is the smallest regular open containing $\mathrm{Int}(P)$. Similarly, on the Euclidean plane, the belief operator patches up the cracks of an open set (see Fig. 2).

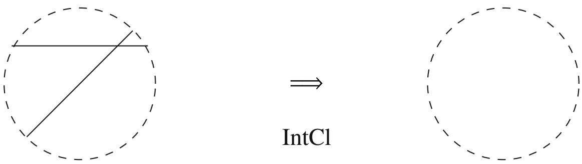


Fig. 2. From U to $\mathrm{Int}(\mathrm{Cl}(U))$

4.2 The Axiomatization of wKD45

We define the logic **wKD45** as

$$\mathbf{wKD45} = \mathbf{K} + (B\varphi \to \langle B\rangle\varphi) + (B\varphi \to BB\varphi) + (B\langle B\rangle B\varphi \to B\varphi)$$

[9] A subset $A \subseteq X$ of a topological space (X, τ) satisfying the condition $A = \mathrm{Int}(\mathrm{Cl}(A)$ is called *regular open* [14].

[10] In fact, for any $A \subseteq X$, the set $\mathrm{Int}(\mathrm{Cl}(A))$ is regular open, however, it is not always the case that $A \subseteq \mathrm{Int}(\mathrm{Cl}(A))$.

and call it *weak* **KD45**. This logic is weaker than **KD45** since it is obtained by replacing the (5)-axiom with the axiom $B\langle B\rangle B\varphi \to B\varphi$, and while $B\langle B\rangle B\varphi \to B\varphi$ is a theorem of **KD45**, the (5)-axiom is not a theorem of **wKD45**. More precisely, **KD45** $\vdash B\langle B\rangle B\varphi \to B\varphi$ but **wKD45** $\nvdash \langle B\rangle\varphi \to B\langle B\rangle\varphi$. We find it hard to give a direct and clear interpretation for this axiom as is given for the axiom of Negative Introspection, since it is too complex in the sense that it includes three consecutive modalities. However, we can interpret it on the basis of the axioms that we have already given an interpretation, in particular, based on the interpretation of Negative Introspection. It is easier to see the correspondence if we state the weak axiom in the following equivalent form:

$$\neg B\varphi \to \langle B\rangle B\neg B\varphi.$$

Recall that the principle of Negative Introspection says that *if an agent does not believe φ, then she believes that she does not believe φ*. On the other hand, taking the reading of Negative Introspection as the reference point, a direct doxastic reading for this axiom is *if the agent does not believe φ, then it is doxastically possible to her that she believes that she does not believe φ*. Therefore, in this section, we work with consistent, positively introspective yet not fully negatively introspective belief. This weakened system **wKD45** stands between **KD4** and **KD45**. While the latter is commonly used as the standard logic for belief, the former has also been studied as a belief system [16,31,32].

4.3 Soundness and Completeness of wKD45

In this section, we prove that **wKD45** for B is sound and complete with respect to the class of all topological spaces. Soundness proof can be presented in a standard way by checking the validity of the axioms and inference rules of **wKD45** with respect to the **w**-topological belief semantics. We leave this proof to the reader and argue for soundness in a different way. For completeness, we follow a technique which allows us to reduce the completeness problem of **wKD45** to the topological completeness of **S4** in the interior semantics. We do this so by defining a translation $(.)^{\circledast}$ from the doxastic language $\mathcal{L}_B$ to the epistemic language $\mathcal{L}_K$ such that for any $\varphi \in \mathcal{L}_B$, we obtain

$$\mathbf{S4} \vdash \varphi^{\circledast} \text{ iff } \mathbf{wKD45} \vdash \varphi.$$

Although we only need the direction from left-to-right for completeness, the other direction comes almost for free and we use this direction to argue for soundness. The above implication can be seen as the key intermediate result for the topological completeness proof of **wKD45**. In order to reach this result, we also make use of soundness and completeness of **S4** and **wKD45** in the standard Kripke semantics. Moreover, we believe that the full and faithful translation of **wKD45** into **S4** given by $(.)^{\circledast}$ is also of interest from a purely modal logical perspective. It implies that the $K\langle K\rangle K$-fragment of **S4** is **wKD45**. In the same way, a full and faithful translation of **KD45** into **S4.2** given by the $\langle K\rangle K$-modality implies that the $\langle K\rangle K$-fragment of **S4.2** is **KD45** [2]. To the best of our knowledge it is still an (interesting) open question how to axiomatize the (non-normal) modal logic **L** which is the $\langle K \rangle K$-fragment of **S4**.

Throughout this section, we use the notation $[\varphi]^{\mathcal{M}}$ for the extension of a formula $\varphi \in \mathcal{L}_K$ w.r.t. the *interior semantics* in order to make clear in which semantics we work. We reserve the notation $[\![\varphi]\!]^{\mathcal{M}}$ for the extensions of the formulas $\varphi \in \mathcal{L}_B$ w.r.t. the **w**-topological belief semantics. We skip the index when confusion is unlikely to occur.

Definition 3 (Translation $(.)^{\circledast} : \mathcal{L}_B \to \mathcal{L}_K$). *For any $\varphi \in \mathcal{L}_B$, the translation $(\varphi)^{\circledast}$ of φ into $\mathcal{L}_K$ is defined recursively as follows:*

1. $(p)^{\circledast} = p$, *where* $p \in \mathrm{Prop}$
2. $(\neg\varphi)^{\circledast} = \neg\varphi^{\circledast}$
3. $(\varphi \wedge \psi)^{\circledast} = \varphi^{\circledast} \wedge \psi^{\circledast}$
4. $(B\varphi)^{\circledast} = K\langle K \rangle K\varphi^{\circledast}$

Note that $(\langle B \rangle\varphi)^{\circledast} = \langle K \rangle K \langle K \rangle \varphi^{\circledast}$.

Proposition 3. *For any topo-model $\mathcal{M} = (X, \tau, \nu)$ and for any formula $\varphi \in \mathcal{L}_B$ we have*

$$[\![\varphi]\!]^{\mathcal{M}} = [\varphi^{\circledast}]^{\mathcal{M}}.$$

Proof. We prove the lemma by induction on the complexity of φ. The cases for the propositional variables and Booleans are straightforward. Now let $\varphi = B\psi$, then

$$\begin{aligned}
[\![\varphi]\!]^{\mathcal{M}} &= [\![B\psi]\!]^{\mathcal{M}} \\
&= \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}([\![\psi]\!]^{\mathcal{M}}))) && \text{(by the } \mathbf{w}\text{-topological belief semantics for } \mathcal{L}_B) \\
&= \mathrm{Int}(\mathrm{Cl}(\mathrm{Int}([\psi^{\circledast}]^{\mathcal{M}}))) && \text{(by I.H.)} \\
&= [K\langle K \rangle K\psi^{\circledast}]^{\mathcal{M}} && \text{(by the interior semantics for } \mathcal{L}_K) \\
&= [(B\psi)^{\circledast}]^{\mathcal{M}} && \text{(by the translation } ^{\circledast}) \\
&= [\varphi^{\circledast}]^{\mathcal{M}}.
\end{aligned}$$

We now recall some frame conditions concerning the relational completeness of the respective systems.

Let (X, R) be a transitive Kripke frame. Recall that a *cluster* is an equivalence class wrt the equivalence relation $\sim$ defined by $x \sim y$ if xRy and yRx for each $x, y \in X$. We denote the set of final clusters of (X, R) by $\mathfrak{C}_R$. A transitive Kripke frame (X, R) having at least one final cluster is called ***weakly cofinal*** if for each $x \in X$ there is a $C \in \mathfrak{C}_R$ such that for all $y \in C$ we have xRy. In fact, every finite reflexive and transitive frame is weakly cofinal. We call a weakly cofinal frame a ***weak brush*** if $X \setminus \bigcup \mathfrak{C}_R$ is an irreflexive anti-chain, i.e., for each $x, y \in X \setminus \bigcup \mathfrak{C}_R$ we have $\neg(xRy)$. A weak brush with a singleton $X \setminus \bigcup \mathfrak{C}_R$ is called a ***weak pin***[11]. By definition, every weak brush and every weak pin is transitive and also serial. A transitive frame (X, R) is called ***rooted***, if there is

[11] Brushes and pins are introduced in [25] and a similar terminology is used in this paper.

an $x \in X$, called a root, such that for each $y \in X$ with $x \neq y$ we have xRy. Finally, we say that a transitive frame (X, R) is of depth n if there is a chain of points $x_1Rx_2R \ldots Rx_n$ such that $\neg(x_{i+1}Rx_i)$ for any $i \leq n$ and there is no chain of greater length satisfying this condition. It is hard to draw a generic picture of a weak brush, but the following figures illustrate weak pins and how a weak brush could look like (where top squares correspond to final clusters) (Figs. 3 and 4).

It is well-known that the (D)-axiom corresponds to seriality and the (4)-axiom corresponds to transitivity of a Kripke frame, under the standard Kripke interpretation (see, e.g., [8, Chap. 4]). It is not very hard to see that the contraposition equivalent $\langle B\rangle\varphi \to \langle B\rangle B\langle B\rangle\varphi$ of our new axiom $B\langle B\rangle B\varphi \to B\varphi$ is a Sahlqvist formula and the first order property corresponding to this axiom is

$$\forall x \forall y (xRy \ \Rightarrow \exists z (xRz \wedge \forall w (zRw \Rightarrow wRy))). \tag{wE}$$

Therefore, a **wKD45** *frame* is a serial and transitive Kripke frame satisfying the above property (wE). We refer the reader to [8, Chap. 3.6] for a more detailed discussion on Sahlqvist formulas.

Let us recall that a point y in a reflexive and transitive Kripke frame (X, R) is called *quasi-maximal* if yRz for some $z \in X$ implies zRy.

Lemma 1. *A rooted Kripke frame* $\mathcal{F} = (X, R)$ *is a* **wKD45** *frame iff it is a cluster or it is a weak pin.*

Proof. The right-to-left direction is trivial. For the other direction, suppose $\mathcal{F} = (X, R)$ is a rooted **wKD45** frame that is not a cluster and assume $x \in X$ is the root. As $\mathcal{F}$ is serial, every quasi-maximal point of it is in a final cluster. Hence, for any $y \in X$, y is a quasi-maximal point iff there is a final cluster C of $\mathcal{F}$ such that $y \in C$, i.e. the set of quasi-maximal points of $\mathcal{F}$ is $\bigcup \mathfrak{C}_R$. Recall that a weak pin is a weakly cofinal frame with a singleton irreflexive $X \setminus \bigcup \mathfrak{C}_R$. We hence need to show that (1) x is an irreflexive point and (2) every successor of x is a quasi-maximal point. Since x is the root and $\mathcal{F} = (X, R)$ is not a cluster, there exists $y \in X$ such that xRy and $\neg(yRx)$.

For (1), suppose that we have xRx. Then, by (wE), there exists $z_0 \in X$ such that xRz_0 and for all $w \in X$ with z_0Rw, we have wRx. Since R is serial, z_0 has at least one successor w, therefore, it is guarateed that there is at least one element $w \in X$ such that wRx. Since $wRxRy$ and R is transitive, we obtain

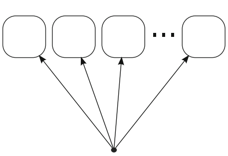


Fig. 3. Weak pin

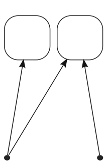


Fig. 4. An example of a weak brush

wRy implying, again by taransitivity, that z_0Ry. Therefore, by (wE), we have yRx, contradicting $\neg(yRx)$. So x is irreflexive.

For (2), suppose there exists $y_0 \in X$ such that xRy_0 and y_0 is not a quasi-maximal element. This means that there is $t_0 \in X$ such that y_0Rt_0 but $\neg(t_0Ry_0)$. By (wE), xRy_0 implies that there exists $z_0 \in X$ such that xRz_0 and for all $w \in X$ with z_0Rw, we have wRy_0. Similarly to the argument above, since it is guaranteed that z_0 has at least one successor w, R is transitive and $z_0RwRy_0Rt_0$ implies z_0Rt_0. Therefore, again by (wE), t_0Ry_0 contradicting our assumption. Thus, every successor of x is a quasi-maximal point. Finally, (1) and (2) together yield that (X, R) is a weak pin.

Lemma 2

1. *Each reflexive and transitive weakly cofinal frame is an* **S4**-*frame. Moreover,* **S4** *is sound and complete w.r.t. the class of finite rooted reflexive and transitive weakly cofinal frames.*
2. *Each weak brush is a* **wKD45**-*frame. Moreover,* **wKD45** *is sound and complete w.r.t. the class of finite weak brushes, indeed, w.r.t. the class of finite weak pins.*

Proof. (1) is well known, see e.g., [8,10]. For (2), we proved in Lemma 1 that the **wKD45**-frames are of finite depth. It is well known that every logic over **K4** that has finite depth is locally tabular and has the finite model property (e.g., [10, Theorem 12.21]). This implies that **wKD45** as well has the finite model property and thus it has the finite model property w.r.t. finite rooted **wKD45**-frames. Then by Lemma 1, we have that **wKD45** is in fact complete w.r.t. finite weak brushes and weak pins.

For any reflexive and transitive weakly cofinal frame (X, R) we define R_B on X by

$$xR_By \text{ if } y \in \bigcup \mathfrak{C}_{R(x)}$$

for each $x, y \in X$, where $\bigcup \mathfrak{C}_{R(x)} = R(x) \cap \bigcup \mathfrak{C}_R$. In other words, $R_B(x) = \bigcup \mathfrak{C}_{R(x)}$ for each $x \in X$. Moreover, we have the following equivalence.

Lemma 3. *For any reflexive and transitive weakly cofinal frame* (X, R) *we have*

$$\bigcup \mathfrak{C}_{R_B} = \bigcup \mathfrak{C}_R.$$

Proof. Let (X, R) be a reflexive and transitive weakly cofinal frame and $x \in X$.

($\subseteq$) Suppose $x \in \bigcup \mathfrak{C}_{R_B}$ and $x \notin \bigcup \mathfrak{C}_R$. Then $x \in \bigcup \mathfrak{C}_{R_B}$ means that $x \in C$ for some $C \in \mathfrak{C}_{R_B}$. As C is a final cluster, there is no $y \in X$ such that xR_By and $\neg(yR_Bx)$. On the other hand, since (X, R) is a weakly cofinal frame, there is a $C' \in \mathfrak{C}_R$ such that xRz for all $z \in C'$. Hence, $C' \subseteq \bigcup \mathfrak{C}_{R(x)}$. Thus, by the definition of R_B, we have $C' \subseteq R_B(x)$. However, as $x \notin \bigcup \mathfrak{C}_R$, we have that $\neg(zRx)$ and thus $\neg(zR_Bx)$ for any $z \in C'$ contradicting $x \in C$ for a final cluster C of (X, R_B). In fact, there is a unique $C \in \mathfrak{C}_{R_B}$ such that $R_B(x) = C$ since C is a final cluster.

($\supseteq$) Suppose $x \in \bigcup \mathfrak{C}_R$. Then, there is a (unique) $C \in \mathfrak{C}_R$ such that $x \in C$ and in fact $R(x) = C$. Also suppose that $x \notin \bigcup \mathfrak{C}_{R_B}$. Hence, there is a $y_0 \in X$ such that xR_By_0 and $\neg(y_0R_Bx)$. Then, $y_0 \in \bigcup \mathfrak{C}_{R(x)}$ but $x \notin \bigcup \mathfrak{C}_{R(y_0)}$ by definition of R_B. By definition of R_B, we have that xR_By_0 implies xRy_0. Hence, as $y_0 \in R(x)$, we also have $R(y_0) = R(x) = C$. Thus, $\bigcup \mathfrak{C}_{R(y_0)} = \bigcup \mathfrak{C}_{R(x)}$. As R is reflexive, $x \in \bigcup \mathfrak{C}_{R(x)}$ and hence $x \in \bigcup \mathfrak{C}_{R(y_0)}$ contradicting $\neg(y_0R_Bx)$.

Lemma 4. *For any reflexive and transitive weakly cofinal Kripke model $\mathcal{M} = (X, R, \nu)$, any $\varphi \in \mathcal{L}_K$ and any $x \in X$, we have*

$$\bigcup \mathfrak{C}_{R(x)} \subseteq \|\varphi\|^{\mathcal{M}} \textit{ iff } x \in \|K\langle K\rangle K\varphi\|^{\mathcal{M}}.$$

Proof. Let $\mathcal{M} = (X, R, \nu)$ be a reflexive and transitive weakly cofinal model, $\varphi \in \mathcal{L}_K$ and $x \in X$.

($\Rightarrow$) Suppose $\bigcup \mathfrak{C}_{R(x)} \subseteq \|\varphi\|^{\mathcal{M}}$. Let $y \in X$ be such that xRy. As R is transitive and xRy, we have $R(y) \subseteq R(x)$ implying that $\bigcup \mathfrak{C}_{R(y)} \subseteq \bigcup \mathfrak{C}_{R(x)}$. Hence, by our assumption, $\bigcup \mathfrak{C}_{R(y)} \subseteq \|\varphi\|^{\mathcal{M}}$. Thus, there is a $C \in \mathfrak{C}_R$ such that $C \subseteq R(y)$ and $C \subseteq \|\varphi\|^{\mathcal{M}}$. Since for all $z \in C$, we have $R(z) = C$ and $C \subseteq \|\varphi\|^{\mathcal{M}}$, we obtain $C \subseteq \|K\varphi\|^{\mathcal{M}}$. As $C \subseteq R(y)$, we have $y \in \|\langle K\rangle K\varphi\|^{\mathcal{M}}$. Therefore, as y has been chosen arbitrarily from $R(x)$ we obtain $x \in \|K\langle K\rangle K\varphi\|^{\mathcal{M}}$.

($\Leftarrow$) Suppose $\bigcup \mathfrak{C}_{R(x)} \not\subseteq \|\varphi\|^{\mathcal{M}}$. This implies that there exists a $y \in \bigcup \mathfrak{C}_{R(x)}$ such that $y \notin \|\varphi\|^{\mathcal{M}}$. Now $y \in \bigcup \mathfrak{C}_{R(x)}$ implies that there is a $C \in \mathfrak{C}_R$ such that $R(y) = C$ and $R(y) \subseteq R(x)$. As zRy for all $z \in C$ and $y \notin \|\varphi\|^{\mathcal{M}}$, we have $z \notin \|K\varphi\|^{\mathcal{M}}$ for all $z \in C$. Then, $R(y) = C$ yields $y \notin \|\langle K\rangle K\varphi\|^{\mathcal{M}}$. Finally, since xRy, we obtain $x \notin \|K\langle K\rangle K\varphi\|^{\mathcal{M}}$.

Lemma 5. *For any reflexive and transitive weakly cofinal frame (X, R),*

1. *(X, R_B) is a weak brush.*
2. *For any valuation ν on X and for each formula $\varphi \in \mathcal{L}_B$ we have $\|\varphi^{\circledast}\|^{\mathcal{M}} = \|\varphi\|^{\mathcal{M}_B}$, where $\mathcal{M} = (X, R, \nu)$ and $\mathcal{M}_B = (X, R_B, \nu)$.*

Proof. Let (X, R) be a reflexive and transitive weakly cofinal frame.

1. – *Transitivity:* Let $x, y, z \in X$ such that xR_By and yR_Bz. This means that $y \in \bigcup \mathfrak{C}_{R(x)}$ and $z \in \bigcup \mathfrak{C}_{R(y)}$. As R is transitive and xRy we have $\bigcup \mathfrak{C}_{R(y)} \subseteq \bigcup \mathfrak{C}_{R(y)}$. Hence, $z \in \bigcup \mathfrak{C}_{R(x)}$, i.e., xR_Bz.
   – *Seriality:* Let $x \in X$. Since (X, R) is weakly cofinal, there is a $y \in \bigcup \mathfrak{C}_{R(x)}$, i.e., xR_By.
   – *Irreflexive, antichain:* Suppose there is an $x \in X \setminus \bigcup \mathfrak{C}_{R_B}$ such that xR_Bx. This implies, $x \in \bigcup \mathfrak{C}_{R(x)}$, thus, $x \in \bigcup \mathfrak{C}_R$. By Lemma 3, $x \in \bigcup \mathfrak{C}_{R_B}$ which contradicts our assumption. Moreover, suppose that $X \setminus \bigcup \mathfrak{C}_{R_B}$ is not an antichain, i.e., there are $x, y \in X \setminus \bigcup \mathfrak{C}_{R_B}$ such that either xR_By or yR_Bx. W.l.o.g., assume xR_By. Hence, by definition of R_B, we have $y \in \bigcup \mathfrak{C}_{R(x)}$. Thus, $y \in \bigcup \mathfrak{C}_R$ and, by Lemma 3, $y \in \bigcup \mathfrak{C}_{R_B}$ contradicting $y \in X \setminus \bigcup \mathfrak{C}_{R_B}$.

2. We prove this item by induction on the complexity of φ. Let $\mathcal{M} = (X, R, \nu)$ be a model on (X, R). The cases for $\varphi = \bot$, $\varphi = p$, $\varphi = \neg\psi$, $\varphi = \psi \wedge \chi$ are straightforward. Now let $\varphi = B\psi$.
   ($\subseteq$) Let $x \in \|(B\psi)^{\circledast}\|^{\mathcal{M}} = \|K\langle K\rangle K\psi^{\circledast}\|^{\mathcal{M}}$. Then, by Lemma 4, $\bigcup \mathfrak{C}_{R(x)} \subseteq \|\psi^{\circledast}\|^{\mathcal{M}}$. By I.H, we obtain $\bigcup \mathfrak{C}_{R(x)} \subseteq \|\psi\|^{\mathcal{M}_B}$. Since $\bigcup \mathfrak{C}_{R(x)} = R_B(x)$, we have $R_B(x) \subseteq \|\psi\|^{\mathcal{M}_B}$ implying that $x \in \|B\psi\|^{\mathcal{M}_B}$.
   ($\supseteq$) Let $x \in \|B\psi\|^{\mathcal{M}_B}$. Then, by the standard Kripke semantics, we have $R_B(x) \subseteq \|\psi\|^{\mathcal{M}_B}$. By I.H, we obtain $R_B(x) \subseteq \|\psi^{\circledast}\|^{\mathcal{M}}$. Since $\bigcup \mathfrak{C}_{R(x)} = R_B(x)$, we have $\bigcup \mathfrak{C}_{R(x)} \subseteq \|\psi^{\circledast}\|^{\mathcal{M}}$. Thus, by Lemma 4, $x \in \|K\langle K\rangle K\psi^{\circledast}\|^{\mathcal{M}} = \|(B\psi)^{\circledast}\|^{\mathcal{M}}$.

Lemma 6. *For any weak brush* (X, R)*,*

1. (X, R^+) *is a reflexive and transitive weakly cofinal frame.*
2. *For any valuation* ν *on* X *and for each formula* $\varphi \in \mathcal{L}_B$ *we have* $\|\varphi\|^{\mathcal{M}} = \|\varphi^{\circledast}\|^{\mathcal{M}^+}$*, where* $\mathcal{M} = (X, R, \nu)$ *and* $\mathcal{M}^+ = (X, R^+, \nu)$*.*

Proof. Let (X, R) be a serial weak brush.

1. Since R is transitive, R^+ is also transitive and it is reflexive by definition. Moreover, (X, R^+) is weakly cofinal since (X, R) is a weak brush.
2. We prove (2) by induction on the complexity of φ. Let $\mathcal{M} = (X, \tau, \nu)$ be a model on (X, R). The cases for $\varphi = \bot$, $\varphi = p$, $\varphi = \neg\psi$, $\varphi = \psi \wedge \chi$ are straightforward. Let $\varphi = B\psi$.
   ($\subseteq$) Let $x \in \|B\psi\|^{\mathcal{M}}$. Then, by the standard Kripke semantics, we have $R(x) \subseteq \|\psi\|^{\mathcal{M}}$. Hence, by I.H., $R(x) \subseteq \|\psi^{\circledast}\|^{\mathcal{M}^+}$. Since (X, R) is a weak brush, $R(x) = \bigcup \mathfrak{C}_{R(x)} \subseteq \bigcup \mathfrak{C}_{R^+(x)}$. Hence, $x \in \bigcup \mathfrak{C}_{R^+(x)}$. Then, by Lemma 4, $x \in \|K\langle K\rangle K\psi^{\circledast}\|^{\mathcal{M}^+}$.
   ($\supseteq$) Let $x \in \|K\langle K\rangle K\psi^{\circledast}\|^{\mathcal{M}^+}$. Then, by Lemma 4, $\bigcup \mathfrak{C}_{R^+(x)} \subseteq \|\psi^{\circledast}\|^{\mathcal{M}^+}$. Thus, by I.H., $\bigcup \mathfrak{C}_{R^+(x)} \subseteq \|\psi\|^{\mathcal{M}}$. Then, as above, $R(x) \subseteq \|\psi\|^{\mathcal{M}}$ implying that $x \in \|B\psi\|^{\mathcal{M}}$.

Theorem 6. *For each formula* $\varphi \in \mathcal{L}_B$*,* $\mathbf{S4} \vdash \varphi^{\circledast}$ *iff* $\mathbf{wKD45} \vdash \varphi$*.*

Proof. Let $\varphi \in \mathcal{L}_B$.

($\Rightarrow$) Suppose $\mathbf{wKD45} \nvdash \varphi$. By Lemma 2(2), there exists a Kripke model $\mathcal{M} = (X, R, \nu)$, where (X, R) is a finite weak pin such that $\|\varphi\|^{\mathcal{M}} \neq X$. Then, by Lemma 6, $\mathcal{M}^+$ is a model based on the finite reflexive and transitive weakly cofinal frame (X, R^+) and $\|\varphi^{\circledast}\|^{\mathcal{M}^+} \neq X$. Hence, by Lemma 2(1), we have $\mathbf{S4} \nvdash \varphi^{\circledast}$.

($\Leftarrow$) Suppose $\mathbf{S4} \nvdash \varphi^{\circledast}$. By Lemma 2(1), there exists a Kripke model $\mathcal{M} = (X, R, \nu)$ where (X, R) is a finite reflexive and transitive weakly cofinal frame such that $\|\varphi^{\circledast}\|^{\mathcal{M}} \neq X$. Then, by Lemma 5, $\mathcal{M}_B$ is a model based on the (finite) weak brush (X, R_B) and $\|\varphi\|^{\mathcal{M}_B} \neq X$. Hence, by Lemma 2(2), we have $\mathbf{wKD45} \nvdash \varphi$.

Theorem 7. **wKD45** *is sound and complete w.r.t. the class of all topological spaces in the* **w***-topological belief semantics.*

Proof. As we noted in the beginning of this section, soundness can be proven directly. Another way of arguing for the topological soundness of **wKD45** is via Theorem 6: let $\varphi \in \mathcal{L}_B$ such that **wKD45** $\vdash \varphi$. Then, by Theorem 6, **S4** $\vdash \varphi^{\circledast}$. By the topological soundness of **S4** w.r.t. the class of all topological spaces in the interior semantics, we obtain that for any topological space (X, τ) we have $(X, \tau) \models \varphi^{\circledast}$. Then, by Proposition 3, we conclude that in the **w**-topological belief semantics $(X, \tau) \models \varphi$.

For completeness, let $\varphi \in \mathcal{L}_B$ be such that **wKD45** $\not\vdash \varphi$. By Theorem 6, **S4** $\not\vdash \varphi^{\circledast}$. Hence, by topological completeness of **S4** w.r.t. the class of all topological spaces in the interior semantics, there exists a topo-model $\mathcal{M} = (X, \tau, \nu)$ such that $[\varphi^{\circledast}]^{\mathcal{M}} \neq X$. Then, by Proposition 3, $[\![\varphi]\!]^{\mathcal{M}} \neq X$. Thus, we found a topological space (X, τ) which refutes φ in the **w**-topological belief semantics. Hence, **wKD45** is complete w.r.t. the class of all topological spaces in the **w**-topological belief semantics.

We point out that the above completeness proof crucially uses reasoning in Kripke frames rather than topology. However, as already mentioned earlier in the paper, topological (and geometrical) reading of our belief modality is key for its intuitive understanding as well as for viewing it as a Stalnaker-like belief operator.

5 The Topology of Static and Dynamic Belief Revision

5.1 Static Belief Revision: Conditional Beliefs

In this section, we explore the topological analogue of static conditioning by providing a topological semantics for conditional belief modalities based on arbitrary topological spaces[12]. We obtain the semantics for a conditional belief modality $B^{\varphi}\psi$ in a natural and standard way, as in [2], by relativizing the semantics for the simple belief modality to the extension of the learnt formula φ. Unlike model *restriction* in the case of updates, our conditional belief semantics does not lead to a change in the initial model. Conditional belief modalities intend to capture the *hypothetical* belief changes of an agent in case she would receive new information (see, e.g., [2] for a more detailed discussion on the topological interpretation of conditional beliefs).

Syntax and Semantics. We now consider the language $\mathcal{L}_{KCB}$ obtained by adding conditional belief modalities $B^{\varphi}\psi$ to $\mathcal{L}_{KB}$, where $B^{\varphi}\psi$ reads *if the agent would learn* φ*, then she would come to believe that* ψ *was the case before the learning* [4, p. 12].

[12] In [2], we propose topological semantics for conditional beliefs based on hereditarily extremally disconnected spaces.

For any subset P of a topological space (X, τ), we can generalize the belief modality B on the topo-models by relativizing the closure and the interior operators to the set P. More precisely, given a topological model $\mathcal{M} = (X, \tau, \nu)$, the additional semantic clause reads

$$[\![B^{\varphi}\psi]\!]^{\mathcal{M}} = \mathrm{Int}([\![\varphi]\!]^{\mathcal{M}} \to \mathrm{Cl}([\![\varphi]\!]^{\mathcal{M}} \cap \mathrm{Int}([\![\varphi]\!]^{\mathcal{M}} \to [\![\psi]\!]^{\mathcal{M}})))$$

where $[\![\varphi]\!]^{\mathcal{M}} \to [\![\psi]\!]^{\mathcal{M}} := (X \backslash [\![\varphi]\!]^{\mathcal{M}}) \cup [\![\psi]\!]^{\mathcal{M}}$.

One possible justification for the above semantics of conditional belief is that it validates an equivalence that generalizes the one for belief in a natural way:

Proposition 4. *The following equivalence is valid in all topological spaces wrt the refined topological semantics for conditional beliefs and knowledge*

$$B^{\varphi}\psi \;\leftrightarrow\; K(\varphi \to \langle K\rangle(\varphi \wedge K(\varphi \to \psi))).$$

This shows that, just like simple beliefs, conditional beliefs can be defined in terms of knowledge and this identity corresponds to the definition of the "conditional connective $\Rightarrow$" in [9]. Moreover, as a corollary of Proposition 4, we obtain that the equivalences

$$B^{\top}\psi \overset{(1)}{\leftrightarrow} K(\top \to \langle K\rangle(\top \wedge K(\top \to \psi)) \overset{(2)}{\leftrightarrow} K\langle K\rangle K\psi \overset{(3)}{\leftrightarrow} B\psi$$

valid in all topological spaces, and thus our semantics for conditional beliefs and simple beliefs (in terms of the interior of the closure of the interior operator) are perfectly compatible with each other. Last but not least, we obtain the complete logic **KCB** of knowledge and conditional beliefs w.r.t. all topological spaces in the following way.

Theorem 8. *The logic* **KCB** *of knowledge and conditional beliefs is axiomatized completely by the system* **S4** *for the knowledge modality K together with the following equivalences:*

1. $B^{\varphi}\psi \leftrightarrow K(\varphi \to \langle K\rangle(\varphi \wedge K(\varphi \to \psi)))$
2. $B\varphi \leftrightarrow B^{\top}\varphi$.

5.2 Dynamic Belief Revision: Updates on All Topological Spaces

In this section, we implement updates on arbitrary topological spaces and show that the problems occurred when we work with extremally disconnected spaces do not arise here: we in fact obtain a complete dynamic logic of knowledge and conditional beliefs with respect to the class of all topological spaces.

We now consider the language $\mathcal{L}_{!KCB}$ obtained by adding (existential) dynamic modalities $\langle !\varphi\rangle\psi$ to $\mathcal{L}_{KCB}$ and we model $\langle !\varphi\rangle\psi$ by means of subspaces exactly the same way as formalized in Sect. 3.2 , i.e., by using the restricted model $\mathcal{M}_{\varphi}$ with the semantic clause

$$[\![\langle !\varphi\rangle\psi]\!]^{\mathcal{M}} = [\![\psi]\!]^{\mathcal{M}_{\varphi}}.$$

In this setting, however, as the underlying static logic **KCB** is the logic of all topological spaces, we implement updates on arbitrary topological spaces. Since the resulting restricted model $\mathcal{M}_\varphi$ is always based on a topological (sub)space and no additional property of the initial topology needs to be inherited by the corresponding subspace (unlike the case for extremally disconnected spaces), we do not face the problem of loosing some validities of the corresponding static system: all the axioms of **KCB** (and, in particular, of **S4** and **wKD45**) will still be valid in the restricted space. Moreover, we obtain a complete axiomatization of the dynamic logic of knowledge and conditional beliefs:

Theorem 9. *The complete and sound dynamic logic* **!KCB** *of knowledge and conditional beliefs with respect to the class of all topological spaces is obtained by adding the following reduction axioms to any complete axiomatization of the logic* **KCB***:*

1. $\langle !\varphi\rangle p \leftrightarrow (\varphi \wedge p)$
2. $\langle !\varphi\rangle \neg\psi \leftrightarrow (\varphi \wedge \neg\langle !\varphi\rangle\psi)$
3. $\langle !\varphi\rangle(\psi \wedge \theta) \leftrightarrow (\langle !\varphi\rangle\psi \wedge \langle !\varphi\rangle\theta)$
4. $\langle !\varphi\rangle K\psi \leftrightarrow (\varphi \wedge K(\varphi \rightarrow \langle !\varphi\rangle\psi))$
5. $\langle !\varphi\rangle B^{\theta}\psi \leftrightarrow (\varphi \wedge B^{\langle !\varphi\rangle\theta}\langle !\varphi\rangle\psi)$
6. $\langle !\varphi\rangle\langle !\varphi\rangle\theta \leftrightarrow \langle !\langle !\varphi\rangle\psi\rangle\theta$

Proof. Proof of this theorem follows, in a standard way, by the soundness of the reduction axioms with respect to all topological spaces. For proof details, we refer to [2, Theorem 12].

6 Conclusion and Future Work

In this paper, we proposed a new topological semantics for belief in terms of the *interior of the closure of the interior* operator which coincides with the one introduced in [1,2,25] on extremally disconnected spaces and diverges from it on arbitrary topological spaces. This new topological semantics for belief comes with significant advantages especially concerning static and dynamic belief revision (in particular, concerning conditional belief and update semantics) and a few disadvantages compared to the setting in [1,2].

In [1,2], we worked with the knowledge system **S4.2** and the standard belief system **KD45**, however, on a restricted class of topological spaces, namely on extremally disconnected spaces. Although the framework of [1,2] provides a solid ground for the static systems of knowledge and belief and the relation between the two, the topological semantics based on extremally disconnected spaces falls short of dealing with updates as shown in Sect. 3.2. In particular, in order to deal with updates one needs to further restrict the class of extremally disconnected spaces to hereditarily extremely disconnected spaces.

In this paper, we did not only provide a semantics for belief based on *all topological spaces* but we also showed that its natural extension to conditional beliefs and updates gave us a 'well-behaved' semantics. In other words, while extending the class of topo-models we could work within the context of knowledge and belief, we also resolved the problem about updates present in the previous setting. The price we had to pay for these results, however, was a weakening of the underlying static knowledge and belief logics: we weakened the knowledge logic **S4.2** to **S4** and the belief logic **KD45** to a slightly weaker one **wKD45**.

This paper can be seen as a continuation of the research program that we have been pursuing on a topological semantics for belief: in [1] we proposed a topological belief semantics based on extremally disconnected spaces and in [2] we investigated a topological belief semantics on hereditarily extremally disconnected spaces and further extended this setting with conditional beliefs and updates. The current work takes a broader perspective and examines belief, conditional beliefs and updates on arbitrary topological spaces.

In on-going work, we investigate a more natural axiomatization of the logic of knowledge and conditional beliefs **KCB** and its dynamic counterpart with respect to arbitrary topological spaces. Moreover, we also investigate the topological semantics for evidence and evidence-based justification in connection with topological interpretations of knowledge and belief in [3] and, following [34], we further explore the dynamics of evidence in a topological setting in the extended version of [3].

Acknowledgments. We thank the anonymous referees for their valuable comments that help us improve the presentation of the paper significantly.

References

1. Baltag, A., Bezhanishvili, N., Özgün, A., Smets, S.: The topology of belief, belief revision and defeasible knowledge. In: Grossi, D., Roy, O., Huang, H. (eds.) LORI 2013. LNCS, vol. 8196, pp. 27–40. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40948-6_3
2. Baltag, A., Bezhanishvili, N., Özgün, A., Smets, S.: The topological theory of belief (2015). http://www.illc.uva.nl/Research/Publications/Reports/PP-2015-18.text.pdf
3. Baltag, A., Bezhanishvili, N., Özgün, A., Smets, S.: Justified belief and the topology of evidence. In: Väänänen, J., Hirvonen, Å., de Queiroz, R. (eds.) WoLLIC 2016. LNCS, vol. 9803, pp. 83–103. Springer, Heidelberg (2016). doi:10.1007/978-3-662-52921-8_6
4. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. Texts Logic Games **3**, 9–58 (2008)
5. Baskent, C.: Topics in Subset Space Logic. Master's thesis, ILLC. University of Amsterdam, The Netherlands (2007)
6. Baskent, C.: Geometric public announcement logics. In: Murray, R.C., McCarthy, P.M. (eds.) Proceedings of the 24th Florida Artificial Intelligence Research Society Conference (FLAIRS-24), pp. 87–88 (2011)
7. Bezhanishvili, G., Bezhanishvili, N., Lucero-Bryan, J., van Mill, J.: S4.3 and hereditarily extremally disconnected spaces. Georgian Mathemetical J. **22**(3), 469–475 (2015)
8. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science, vol. 53. Cambridge University Press, Cambridge (2001)
9. Boutilier, C.: Conditional logics of normality as modal systems. In: Shrobe, H.E., Dietterich, T.G., Swartout, W.R. (eds.) AAAI, pp. 594–599. AAAI Press/The MIT Press (1990)

10. Chagrov, A.V., Zakharyaschev, M.: Modal Logic. Oxford Logic Guides, vol. 35. Oxford University Press, Oxford (1997)
11. Chellas, B.F.: Modal Logic. Cambridge University Press, Cambridge (1980)
12. DeRose, K.: The Case for Contextualism, 1st edn. Oxford University Press, New York (2009)
13. Dugundji, J.: Topology. Allyn and Bacon Series in Advanced Mathematics. Prentice Hall, Saddle River (1965)
14. Engelking, R.: General Topology, vol. 6, 2nd edn. Heldermann Verlag, Berlin (1989)
15. Gettier, E.: Is justified true belief knowledge? Analysis **23**, 121–123 (1963)
16. Hendricks, V., Symons, J.: Epistemic logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, fall 2015 edn. Metaphysics Research Lab Stanford University (2015)
17. Hintikka, J.: Knowledge and Belief. Cornell University Press, Ithaca (1962)
18. Ichikawa, J.J., Steup, M.: The analysis of knowledge. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, fall 2013 edn. Metaphysics Research Lab, Stanford University (2013)
19. Klein, D., Gratzl, N., Roy, O.: Introspection, normality and agglomeration. In: Hoek, W., Holliday, W.H., Wang, W. (eds.) LORI 2015. LNCS, vol. 9394, pp. 195–206. Springer, Heidelberg (2015). doi:10.1007/978-3-662-48561-3_16
20. Klein, P.: A proposed definition of propositional knowledge. J. Philos. **68**, 471–482 (1971)
21. Lehrer, K., Paxson, T.J.: Knowledge: undefeated justified true belief. J. Philos. **66**, 225–237 (1969)
22. Lenzen, W.: Recent Work in Epistemic Logic, vol. 30. Acta Philosophica Fennica, North Holland (1978)
23. McKinsey, J.C.C., Tarski, A.: The algebra of topology. Ann. of Math. **45**(2), 141–191 (1944)
24. Nozick, R.: Philosophical Explanations. Harvard University Press, Cambridge (1981)
25. Özgün, A.: Topological Models for Belief and Belief Revision. Master's thesis, ILLC. University of Amsterdam, The Netherlands (2013)
26. Parikh, R., Moss, L.S., Steinsvold, C.: Topology and epistemic logic. In: Aiello, M., Pratt-Hartmann, I., van Benthem, J. (eds.) Handbook of Spatial Logics, pp. 299–341. Springer, Amsterdam (2007)
27. Rott, H.: Stability, strength and sensitivity: converting belief into knowledge. Erkenntnis **61**, 469–493 (2004)
28. Sikorski, R.: Boolean Algebras. Springer, Heidelberg (1964)
29. Sosa, E.: How to defeat opposition to moore. Nos **33**, 141–153 (1999)
30. Stalnaker, R.: On logics of knowledge and belief. Phil. Stud. **128**, 169–199 (2006)
31. Steinsvold, C.: Topological models of belief logics. Ph.D. thesis, New York, NY, USA (2007)
32. Steinsvold, C.: A grim semantics for logics of belief. J. Philos. Logic **37**, 45–56 (2008)
33. van Benthem, J., Bezhanishvili, G.: Modal logics of space. In: Aiello, M., van Benthem, J., Pratt-Hartman, I. (eds.) Handbook of Spatial Logics, pp. 217–298. Springer, Amsterdam (2007)
34. van Benthem, J., Pacuit, E.: Dynamic logics of evidence-based beliefs. Stud. Logica **99**(1), 61–92 (2011)
35. Voorbraak, F. As Far as I Know. Ph.D. thesis, Utrecht University (1993)
36. Williamson, T.: Knowledge and its Limits. Oxford University Press, Oxford (2000)
37. Wolter, F. Lattices of Modal Logics. Ph.D. thesis. Free University, Berlin (1993)
38. Zvesper, J.: Playing with Information. Ph.D. thesis, ILLC. University of Amsterdam (2010)

Universal Models for the Positive Fragment of Intuitionistic Logic

Nick Bezhanishvili[1], Dick de Jongh[1], Apostolos Tzimoulis[2], and Zhiguang Zhao[2](✉)

[1] ILLC, University of Amsterdam, Amsterdam, The Netherlands
{N.Bezhanishvili,D.H.J.deJongh}@uva.nl
[2] Delft University of Technology, Delft, The Netherlands
apostolos@tzimoulis.eu, zhaozhiguang23@gmail.com

Abstract. We describe the n-universal model $\mathcal{U}^\star(n)$ of the positive fragment of the intuitionistic propositional calculus IPC. We show that $\mathcal{U}^\star(n)$ is isomorphic to a generated submodel of $\mathcal{U}(n)$ – the n-universal model of IPC. Using $\mathcal{U}^\star(n)$, we give an alternative proof of Jankov's theorem stating that the intermediate logic KC, the logic of the weak law of excluded middle, is the greatest intermediate logic extending IPC that proves exactly the same positive formulas as IPC.

Keywords: Universal models · Positive morphism · Fragment of intuitionistic logic · Jankov's theorem

1 Introduction

In this paper, we use the tools of universal models to study the positive fragment of intuitionistic propositional calculus IPC, i.e., formulas containing only propositional variables, $\wedge$, $\vee$ and $\to$. Fragments of intuitionistic logic have been thoroughly investigated in the literature. For a detailed historic account we refer to [20]. Among these fragments, the locally finite ones, i.e., the fragments where for each $n \in \omega$ there are only finitely many non-equivalent formulas in n variables, attracted more attention. For example, [13] proved a classic result that the $[\wedge, \to]$-fragment of IPC is locally finite. The positive fragment is not locally finite, and as a result it has not received much attention in the literature. The major interest for the study of this fragment comes from minimal logic [19], the sublogic of intuitionistic logic obtained by dropping the axiom $\bot \to \varphi$.

Universal models of intuitionistic logic can be seen as duals to free Heyting algebras. The basic idea underlying the construction of universal models can be traced back to [9]. Universal models for the full IPC were described in [1,16,21,22]; for a detailed exposition see also [4, Sect. 3.2], [6, Sect. 8.7] and [11, Sect. 3]. We refer to [15, Sect. 3.2.1] for an overview of the history of universal models. The universal model for the $[\wedge, \to]$-fragment of IPC is characterized

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 229–250, 2017.
DOI: 10.1007/978-3-662-54332-0_13

in [5,14,23]. Universal models for other locally finite fragments of IPC are discussed[1] in [10,17]. In this paper we focus on the $[\wedge, \vee, \rightarrow]$-fragment of IPC.

The contribution of the present paper can be listed as follows:

- We describe the n-universal model $\mathcal{U}^\star(n)$ of the positive fragment of IPC and show that it is isomorphic to a generated submodel of the n-universal model $\mathcal{U}(n)$ of IPC and at the same time is a (positive morphism) quotient of $\mathcal{U}(n)$. We study the properties of $\mathcal{U}^\star(n)$ as well as its connection with the n-Henkin model $\mathcal{H}^\star(n)$ for the positive fragment of IPC.
- Using $\mathcal{U}^\star(n)$, we give an alternative proof of Jankov's theorem that the intermediate logic KC, the logic of the weak law of excluded middle, is the greatest intermediate logic extending IPC that proves exactly the same positive formulas as IPC.

The paper is organized as follows: In Sect. 2, we recall all the basic notions and results used consequently in the paper. We also discuss the top-model property and its relationship with the positive fragment of IPC. In Sect. 3, we define the universal models for the positive fragment of IPC. We also recall the definition of positive morphisms and show that every finite Kripke model can be mapped via a positive morphism into the universal model. We also define positive Jankov-de Jongh formulas and prove an analogue of the Jankov-de Jongh theorem for these formulas. In Sect. 4, we discuss the relationship between the n-Henkin models and the n-universal models of the positive fragment and in Sect. 5 we give an alternative proof of Jankov-de Jongh and Jankov's theorems. In Sect. 6 we summarize obtained results and discuss some future research directions.

2 Preliminaries

2.1 Basic Notations

In this section, we briefly recall the relational semantics for the intuitionistic propositional calculus IPC. For a detailed study of IPC we refer to [6].

Definition 1 (Kripke frames and models). *A* Kripke frame *is a pair* $\mathfrak{F} = (W, R)$ *where* W *is a set and* R *is a partial order on it. A* Kripke model *is a triple* $\mathfrak{M} = (W, R, V)$ *where* (W, R) *is a Kripke frame and* V *is a partial map* $V : \mathrm{Prop} \rightarrow \mathscr{P}(W)$ *(where* Prop *is the set of propositional variables and* $\mathscr{P}(W)$ *is the powerset of* W*) such that for any* $w, w' \in W$ *we have that* $w \in V(p)$ *and* wRw' *imply* $w' \in V(p)$.

The valuation can be extended to all formulas in a standard way. We call the upward closed subsets of W (with respect to R) *upsets*. The set of all upsets of W is denoted by $\mathrm{Up}(W)$. As usual $w \in V(\varphi)$ will be denoted as $w \models \varphi$.

[1] We note that [10,17] do not discuss universal (*exact* in their terminology) models of non-locally finite fragments of IPC.

Definition 2 (General frames).

1. *A* general frame *is a triple* $\mathfrak{F} = (W, R, \mathcal{P})$*, where* (W, R) *is a Kripke frame and* $\mathcal{P}$ *is a family of upsets containing* $\emptyset$ *and closed under* $\cap, \cup$ *and the following operation* $\Rightarrow$*: for every* $X, Y \subseteq W$*,*

$$X \Rightarrow Y = \{x \in W : \forall y \in W(xRy \wedge y \in X \to y \in Y)\}.^{2}$$

Elements of the set $\mathcal{P}$ *are called* admissible sets*.*

2. *A general frame* $\mathfrak{F} = (W, R, \mathcal{P})$ *is called* refined *if for any* $x, y \in W$*,*

$$\forall X \in \mathcal{P}(x \in X \to y \in X) \Rightarrow xRy.$$

3. $\mathfrak{F}$ *is called* compact*, if for any families* $\mathcal{X} \subseteq \mathcal{P}$ *and* $\mathcal{Y} \subseteq \{W \backslash X : X \in \mathcal{P}\}$*, for which* $\mathcal{X} \cup \mathcal{Y}$ *has the finite intersection property (i.e., finite intersections of the elements of* $\mathcal{X} \cup \mathcal{Y}$ *are non-empty), we have* $\bigcap(\mathcal{X} \cup \mathcal{Y}) \neq \emptyset$*.*
4. *A general frame* $\mathfrak{F}$ *is called a* descriptive frame *if it is refined and compact.*[3]

By an *n-formula* we mean a formula built from $p_1, \ldots, p_n$. An *n-model* is a model where Prop $= \{p_1, \ldots, p_n\}$. Next we recall some frame and model constructions that will be used consequently.

Definition 3 (Generated subframe and generated submodel).

1. *For any Kripke frame* $\mathfrak{F} = (W, R)$ *and* $X \subseteq W$*, the* subframe of $\mathfrak{F}$ generated by X *is* $\mathfrak{F}_X = (R(X), R')$*, where* $R(X) = \{w' \in W : wRw'$ *for some* $w \in X\}$ *and* R' *is the restriction of* R *to* $R(X)$*. If* $X = \{w\}$*, then we denote* $\mathfrak{F}_X$ *by* $\mathfrak{F}_w$ *and* $R(X)$ *by* $R(w)$*.*
2. *For any Kripke frame* $\mathfrak{F} = (W, R)$*, any valuation* V *on* $\mathfrak{F}$ *and* $X \subseteq W$*, the* submodel of $\mathfrak{M} = (\mathfrak{F}, V)$ generated by X *is* $\mathfrak{M}_X = (\mathfrak{F}_X, V')$*, where* V' *is the restriction of* V *to* $R(X)$*. If* X *is a singleton* $\{w\}$*, then we denote* $\mathfrak{M}_X$ *by* $\mathfrak{M}_w$*.*
3. *For any general frame* $\mathfrak{F} = (W, R, \mathcal{P})$ *and any* $X \subseteq W$*, the* (general) subframe of $\mathfrak{F}$ generated by X *is* $\mathfrak{F}_X = (R(X), R', \mathcal{Q})$*, where* $(R(X), R')$ *is the subframe of* (W, R) *generated by* X*, and* $\mathcal{Q} = \{U \cap R(X) : U \in \mathcal{P}\}$*.*

Let $\mathfrak{F} = (W, R, \mathcal{P})$ be a descriptive frame and let $W' \in \mathcal{P}$. Let $\mathfrak{G} = (W', R', \mathcal{Q})$ denote a general frame such that R' is the restriction of R to W' and $\mathcal{Q} = \{U \cap W' : U \in \mathcal{P}\}$. For a proof of the next lemma we refer to, e.g., [23]. In terms of Esakia spaces this lemma states that a restriction of the order and topology of an Esakia space to a clopen upset in it yields again an Esakia space.

Lemma 4. *Let* $\mathfrak{F} = (W, R, \mathcal{P})$ *be a descriptive frame and let* $W' \in \mathcal{P}$*. Then* $\mathfrak{G} = (W', R', \mathcal{Q})$ *is a descriptive frame.*

---

[2] In fact, $\Rightarrow$ is just the Heyting implication of the Heyting algebra of all upsets of W.

[3] Descriptive general frames are essentially the same as Esakia spaces (see e.g., [4, Sect. 2.3]). This topological perspective explains why compact general frames are called "compact" (the corresponding topology is compact). This also explains why $\mathcal{Q}$ in Definition 3(3) is defined this way.

Let $\mathfrak{F} = (W, R, \mathcal{P})$ be a descriptive frame. A *descriptive* (or an *admissible*) *valuation on* $\mathfrak{F}$ is a map V : Prop $\to \mathcal{P}$. A pair $(\mathfrak{F}, V)$ is a *descriptive model* is $\mathfrak{F}$ is a descriptive frame and V a descriptive valuation on $\mathfrak{F}$. The truth and validity of formulas in Kripke and descriptive frames and models are defined in a standard way. Next we recall the definition of p-morphisms.

Definition 5 (p-morphism).

1. *Let* $\mathfrak{F} = (W, R)$ *and* $\mathfrak{F}' = (W', R')$ *be Kripke frames. A map* $f : W \to W'$ *is called a* p-morphism *from* $\mathfrak{F}$ *to* $\mathfrak{F}'$ *if*
   - wRw' *implies* $f(w)R'f(w')$ *for any* $w, w' \in W$*;*
   - $f(w)R'v'$ *implies* $\exists v \in W(wRv \wedge f(v) = v')$*.*
2. *Let* $\mathfrak{F} = (W, R, \mathcal{P})$ *and* $\mathfrak{G} = (V, S, \mathcal{Q})$ *be general frames. We call a Kripke frame p-morphism* f *of* (W, R) *to* (V, S) *a* (general frame)p-morphism *of* $\mathfrak{F}$ *to* $\mathfrak{G}$*, if*
$$\forall X \in \mathcal{Q}, f^{-1}(X) \in \mathcal{P}.$$
3. *A* p-morphism f *from* $\mathfrak{M} = (W, R, V)$ *to* $\mathfrak{M}' = (W', R', V')$ *is a p-morphism from* (W, R) *to* (W', R') *such that* $w \in V(p) \Leftrightarrow f(w) \in V'(p)$ *for every* $p \in$ Prop*. For models based on general frames, we also require the condition for p-morphisms between general frames. For* n*-models, the definition is similar.*

The extra condition on p-morphisms in Definition 5.2 is again best explained by viewing descriptive frames as Esakia spaces. This condition is then just equivalent to continuity.

Next we recall the definition of n-Henkin model, which is the canonical model for the n-variable fragment of IPC.

Definition 6 (n-Henkin model).

1. *An* n-theory *is a set of* n*-formulas closed under deduction in* IPC*.*
2. *A set of formulas* Γ *has the* disjunction property*, if for all* n*-formulas* φ, ψ*, we have that* $\varphi \vee \psi \in \Gamma$ *implies* $\varphi \in \Gamma$ *or* $\psi \in \Gamma$*.*
3. *The* n-canonical model *or* n-Henkin model $\mathcal{H}(n) = (W_n, R_n, V_n)$ *is a model where* W_n *consists of all consistent* n*-theories with the disjunction property,* R_n *is the subset relation, and* $\Gamma \in V_n(p)$ *iff* $p \in \Gamma$*.*

2.2 The n-universal Model for the Full Language of IPC

In this section we recall the definition of the n-universal model for the full language of IPC, state its main properties, recall the definition of the de Jongh formulas and the statement of the Jankov-de Jongh theorem. Proofs of all the results stated here can be found in [4, Chap. 3], [6, Sect. 8.6] and [11, Sect. 3].

In the following, we use the terminology *color* to denote the valuation at a world in an n-model. In general, an n-color (n can be omitted if it is clear from the context) is a sequence $c_1 \dots c_n$ of 0's and 1's. The set of all n-colors is denoted by C^n. We define the order on colors as follows:

$$c_1 \dots c_n \leq c'_1 \dots c'_n \text{ iff } c_i \leq c'_i, \text{ for } 1 \leq i \leq n.$$

We write $c_1 \dots c_n < c'_1 \dots c'_n$ if $c_1 \dots c_n \leq c'_1 \dots c'_n$ but $c_1 \dots c_n \neq c'_1 \dots c'_n$.

A *coloring* on $\mathfrak{F} = (W, R)$ is a map $col : W \to C^n$ satisfying $uRv \Rightarrow col(u) \leq col(v)$. It is easy to see that colorings and valuations are in 1-1 correspondence. Given $\mathfrak{M} = (W, R, V)$, we can reconstruct the valuation by the coloring $col_V : W \to C^n$, where $col_V(w) = c_1 \dots c_n$, and for each $1 \leq i \leq n$ we have $c_i = 1$ if $w \in V(p_i)$, and 0 otherwise. We call $col_V(w)$ the *color of* w *under* V.

In any frame $\mathfrak{F} = (W, R)$, we say that $X \subseteq W$ *totally covers* w (notation: $w \prec X$), if X is the set of all immediate successors of w. When $X = \{v\}$, we write $w \prec v$. A set $X \subseteq W$ is called an *anti-chain* if $|X| > 1$ and for every $w, v \in X$, $w \neq v$ implies that $\neg(wRv)$ and $\neg(vRw)$. If uRv we say that u *is under* v.

We can now inductively define the n-universal model $\mathcal{U}(n)$ by cumulative layers $\mathcal{U}(n)^k$ for $k \in \omega$, where each layer contains all the points w such that the longest chain starting from w has length k, omitting n if it is clear from the context.

Definition 7 (n-universal model).

- *The first layer* $\mathcal{U}(n)^1$ *consists of* 2^n *nodes with the* 2^n *different* n*-colors under the discrete ordering.*
- *For* $k \geq 1$*, under each element* w *in* $\mathcal{U}(n)^k$*, for each color* $s < col(w)$*, we put a new node* v *in* $\mathcal{U}(n)^{k+1}$ *such that* $v \prec w$ *with* $col(v) = s$*, and we take the reflexive transitive closure of the ordering.*
- *For* $k \geq 1$*, under any finite anti-chain* X *with at least one element in* $\mathcal{U}(n)^k$ *and any color* s *with* $s \leq col(w)$ *for all* $w \in X$*, we put a new element* v *in* $\mathcal{U}(n)^{k+1}$ *such that* $col(v) = s$ *and* $v \prec X$ *and we take the reflexive transitive closure of the ordering.*

The whole model $\mathcal{U}(n)$ *is the union of its layers.*

It is easy to see from the construction that every $\mathcal{U}(n)^k$ is finite. As a consequence, the generated submodel $\mathcal{U}(n)_w$ is finite for any node w in $\mathcal{U}(n)$.

We now state some properties of the n-universal model. For a proof of the next lemma, we refer to, e.g., [6, Sect. 8.6], [23, Theorem 3.2.3] and [11, Lemma 11].

Lemma 8. *Let* $\mathfrak{M}$ *be a finite rooted Kripke* n*-model. Then there exist a unique* $w \in \mathcal{U}(n)$ *and a unique p-morphism* f *mapping* $\mathfrak{M}$ *onto* $\mathcal{U}(n)_w$*.*

The next theorem shows that $\mathcal{U}(n)$ is a counter-model to every n-formula not provable in IPC. This justifies the name "universal model" for $\mathcal{U}(n)$. For a proof, we refer to, e.g., [11, Theorem 13] and [23, Theorem 3.2.4].

Theorem 9.

1. *For any* n*-formula* φ *we have* $\mathcal{U}(n) \models \varphi$ *iff* $\vdash_{\mathsf{IPC}} \varphi$*.*
2. *For any* n*-formulas* φ *and* ψ*, and for all* $w \in \mathcal{U}(n)$ *we have*

$$(\mathcal{U}(n), w \models \varphi \Rightarrow \mathcal{U}(n), w \models \psi) \text{ iff } \varphi \vdash_{\mathsf{IPC}} \psi.$$

In the following, we recall the definition of de Jongh formulas for the full language of IPC and the fact that these formulas define point-generated submodels of universal models.

For any node w in an n-model $\mathfrak{M}$, if $w \prec \{w_1, \dots, w_m\}$, then we let

$$\begin{aligned} \mathrm{prop}(w) &= \{p_i | w \models p_i, 1 \leq i \leq n\}, \\ \mathrm{notprop}(w) &= \{q_i | w \nvDash q_i, 1 \leq i \leq n\}, \\ \mathrm{newprop}(w) &= \{r_j | w \nvDash r_j \text{ and } w_i \models r_j \text{ for each } 1 \leq i \leq m, \text{ for } 1 \leq j \leq n\} \end{aligned}$$

Here newprop(w) denotes the set of atoms which are "about to be true in w", i.e., the atoms that are false in w but are true in its all proper successors. For the definition of a depth of a point in a frame we refer to [4, Definition 3.1.9] or [6, p. 43]. Roughly speaking, a point w of a universal model has *depth* k if belongs to the k-th layer of $\mathcal{U}(n)$. The depth of a point w will be denoted by $d(w)$.

Definition 10. *Let w be a point in $\mathcal{U}(n)$. We inductively define the corresponding* de Jongh formulas *φ_w and ψ_w:*

If $d(w) = 1$, then let

$$\varphi_w = \bigwedge \mathrm{prop}(w) \wedge \bigwedge \{\neg p_k | p_k \in \mathrm{notprop}(w), 1 \leq k \leq n\},$$

and

$$\psi_w = \neg \varphi_w.$$

If $d(w) > 1$, and $\{w_1, \dots, w_m\}$ is the set of all immediate successors of w, then define

$$\varphi_w = \bigwedge \mathrm{prop}(w) \wedge (\bigvee \mathrm{newprop}(w) \vee \bigvee_{i=1}^{m} \psi_{w_i} \to \bigvee_{i=1}^{m} \varphi_{w_i}),$$

and

$$\psi_w = \varphi_w \to \bigvee_{i=1}^{m} \varphi_{w_i}.$$

The most important properties of the de Jongh formulas are recalled in the following proposition. For a proof, we refer to [4, Theorem 3.3.2].

Proposition 11. *For every $w \in \mathcal{U}(n)$, we have:*

- $V(\varphi_w) = R(w)$,
- $V(\psi_w) = \mathcal{U}(n) \setminus R^{-1}(w)$, *where* $R^{-1}(w) = \{w' \in \mathcal{U}(n) : w' R w\}$.

Now we state more properties of the universal model and de Jongh formulas. For a proof of the next proposition we refer to [11, Corollary 19]. We let

$$\begin{aligned} Cn_n(\varphi) &= \{\psi : \psi \text{ is an } n\text{-formula such that } \vdash_{\mathsf{IPC}} \varphi \to \psi\}, \\ Th_n(\mathfrak{M}, w) &= \{\varphi : \varphi \text{ is an } n\text{-formula such that } \mathfrak{M}, w \models \varphi\}, \end{aligned}$$

We will omit n if it is clear from the context.

Proposition 12. *For any point w in $\mathcal{U}(n)$, $Th_n(\mathcal{U}(n), w) = Cn_n(\varphi_w)$.*

The next lemma states that $\mathcal{U}(n)_w$ is isomorphic to the submodel of $\mathcal{H}(n)$ generated by the theory axiomatized by the de Jongh formula of w. For a proof, we refer to [11, Lemma 20].

Lemma 13. *For any $w \in \mathcal{U}(n)$, let φ_w be the de Jongh formula of w, then we have that $\mathcal{H}(n)_{Cn(\varphi_w)}$ is isomorphic to $\mathcal{U}(n)_w$.*

Let Upper($\mathfrak{M}$) denote the submodel $\mathfrak{M}_{\{w \in W | d(w) < \omega\}}$ generated by all the points of finite depth. Intuitively, Upper($\mathfrak{M}$) is the "upper" part of $\mathfrak{M}$. It can be shown that the n-universal model is isomorphic to the upper part of the n-Henkin model, i.e., to Upper($\mathcal{H}(n)$). For a proof, we refer to, e.g., [4, Theorem 3.2.9] and [11, Theorem 39].

Theorem 14. Upper($\mathcal{H}(n)$) *is isomorphic to $\mathcal{U}(n)$.*

The following result follows from Proposition 11 and Lemma 13. For a proof see [11, Corollary 21].

Proposition 15. *Let $\mathfrak{M}$ be any model and w be a point in $\mathcal{U}(n) = (W, R, V)$. For any point x in $\mathfrak{M}$, if $\mathfrak{M}, x \models \varphi_w$, then there exists a unique point v satisfying $\mathfrak{M}, x \models \varphi_v, \mathfrak{M}, x \not\models \varphi_{v_1}, \ldots, \mathfrak{M}, x \not\models \varphi_{v_m}$, where $v \prec \{v_1, \ldots, v_m\}$, and wRv.*

In the following we state the Jankov-de Jongh theorem for the full language of IPC. For a proof we refer to [4, Theorem 3.3.3] and [11, Theorem 26].

Theorem 16 (Jankov-de Jongh theorem for IPC). *Let $\mathfrak{G}$ be a descriptive frame and $w \in U(n)$ for some $n \in \omega$. Then $\mathfrak{G} \not\models \psi_w$ iff there is an n-valuation V on $\mathfrak{G}$ such that $\mathcal{U}(n)_w$ is a p-morphic image of a generated submodel of $(\mathfrak{G}, V)$.*

Notice that for each finite rooted frame $\mathfrak{F}$ there is a valuation V such that $(\mathfrak{F}, V)$ is isomorphic to $\mathcal{U}(n)_w$ for some $n \in \omega$ and $w \in U(n)$. (For this it is sufficient to introduce a new propositional variable p_s for each s in $\mathfrak{F}$ and let $V(p_s) = R(s)$.) So the above theorem applies to any finite rooted $\mathfrak{F}$.

2.3 The Top-Model Property

The *positive fragment* of IPC consists of the formulas constructed only by $\wedge, \vee, \rightarrow$. We denote this language by $\mathcal{L}_{\wedge,\vee,\rightarrow}$. Formulas in this fragment will be called *positive formulas*[4]. For the other fragments of IPC the notation is similar.

By replacing every occurrence of $\bot$ by $\neg(p \rightarrow p)$, every formula is IPC-equivalent to a $\bot$-free formula. For simplicity of discussion, we restrict our attention to $\bot$-free formulas (i.e. formulas in $\mathcal{L}_{\wedge,\vee,\rightarrow,\neg}$) only.

Definition 17 (Top model). *A Kripke model $\mathfrak{M} = (W, R, V)$ is called a* top model*, if it has a node $t \in W$ such that:*

[4] Notice that some authors, e.g., [6] call such formulas *negation-free*.

– *t is a successor of all nodes, i.e., we have wRt for each $w \in W$;*
– *all propositional variables are true in t.*

The node t is called the top point *of $\mathfrak{M}$.*

Definition 18 (Top-model property). *We say that a formula φ has the* top-model property, *if for all Kripke models $\mathfrak{M} = (W, R, V)$, all $w \in W$ we have $\mathfrak{M}, w \models \varphi$ iff $\mathfrak{M}^+, w \models \varphi$, where $\mathfrak{M}^+ = (W^+, R^+, V^+)$ is obtained by adding a top point t to $\mathfrak{M}$.*

The next proposition states that there is a procedure which with any intuitionistic formula φ associates a positive formula φ^* or $\perp$ equivalent to φ over top models. The algorithm describing how to compute φ^* given φ is provided in [12, Theorem 5].

Proposition 19. *There is an algorithm which transforms any formula φ in $\mathcal{L}_{\wedge,\vee,\to,\neg}$ into a formula φ^* in $\mathcal{L}_{\wedge,\vee,\to} \cup \{\perp\}$ such that for any top model $\mathfrak{M}$ and any node w in $\mathfrak{M}$, we have $\mathfrak{M}, w \models \varphi \leftrightarrow \varphi^*$. Furthermore if $\varphi^* \neq \perp$ and $\psi^* \neq \perp$ then $(\varphi \to \psi)^* = \varphi^* \to \psi^*$, $(\varphi \wedge \psi)^* = \varphi^* \wedge \psi^*$ and $(\varphi \vee \psi)^* = \varphi^* \vee \psi^*$.*

3 The Universal Models for the Positive Fragment of IPC

3.1 The Universal Model

We will now proceed by defining the n-universal model, $\mathcal{U}^\star(n)$, for the positive fragment of IPC. This model closely resembles the n-universal model for IPC: it is a generated submodel of it, and as we shall see below, is also a positive morphism quotient (see Definition 22 and Lemma 27). We now define $\mathcal{U}^\star(n) = (U^\star(n), R^\star, V^\star)$ inductively in a similar way as we defined $\mathcal{U}(n)$.

Definition 20.

– *The first layer $\mathcal{U}^\star(n)^1$ consists of $2^n - 1$ nodes with all the different n-colors —* excluding the color $1 \dots 1$ — *under the discrete ordering.*
– *For $k \geq 1$, under each element w in $\mathcal{U}^\star(n)^k$, for each color $s < col(w)$, we put a new node v in $\mathcal{U}^\star(n)^{k+1}$ such that $v \prec w$ with $col(v) = s$, and we take the reflexive transitive closure of the ordering.*
– *For $k \geq 1$, under any finite anti-chain X with at least one element in $\mathcal{U}^\star(n)^k$ and any color s with $s \leq col(w)$ for all $w \in X$, we put a new element v in $\mathcal{U}^\star(n)^{k+1}$ such that $col(v) = s$ and $v \prec X$, and we take the reflexive transitive closure of the ordering.*

The whole model $\mathcal{U}^\star(n)$ is the union of its layers.

Notice that $\mathcal{U}^\star(1)$ is very different from the Rieger-Nishimura ladder $\mathcal{U}(1)$. It is well known that $\mathcal{U}(1)$ is infinite while $\mathcal{U}^\star(1)$ consists of a single point that does not satisfy p. The only formulas satisfied at this point are the classical tautologies. For $n > 1$ we have that $\mathcal{U}^\star(n)$ is infinite. Below we present the first two layers of $\mathcal{U}^\star(2)$ (Fig. 1). The third layer consists of 72 points.

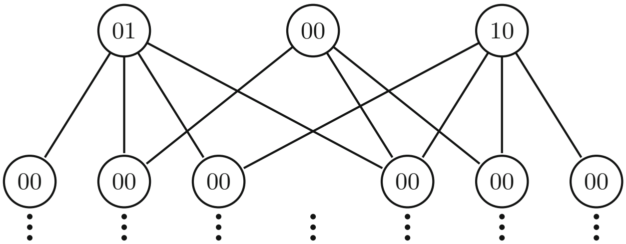


Fig. 1. The first two layers of $\mathcal{U}^{\star}(2)$

It is known (see e.g., [4, Theorem 3.2.19]) that $\mathcal{U}(n)$, where $n \geq 2$, has uncountably many upsets, and therefore, as our language is countable, not all of them are definable. A similar result holds for $\mathcal{U}^{\star}(n)$.

Lemma 21. *There are uncountably many upsets in* $\mathcal{U}^{\star}(n)$ *for* $n \geq 2$. *Thus, not all upsets of* $\mathcal{U}^{\star}(n)$ *are definable.*

Proof. We prove the result for $n = 2$. The case $n > 2$ follows since the underlying frame of $\mathcal{U}^{\star}(2)$ is a generated subframe of the underlying frame of $\mathcal{U}^{\star}(n)$ for $n \geq 2$. We show that $\mathcal{U}^{\star}(2)$ has a countable anti-chain. As every subset of this anti-chain generates a unique upset, the latter implies that there are $2^{\aleph_0}$ upsets of $\mathcal{U}^{\star}(2)$. This, in turn, means that not all upsets of $\mathcal{U}^{\star}(2)$ are definable.

The proof proceeds similarly to the proof that $\mathcal{U}(2)$ has a countable anti-chain (see e.g., [4, Theorem 3.2.19(2)]). The points of the first layer of $\mathcal{U}^{\star}(2)$ are of the colors 01, 00 and 10. It is easy to see that we can embed the Rieger-Nishimura ladder into the submodel of $\mathcal{U}^{\star}(2)$ that contains the points that are not below the 10-point. Hence there exists a countable chain of points below the 00-point, let us call them a_n (where $a_n R^{\star} a_m$ for $n \geq m$), such that the upsets they generate do not contain the 10-point. This means that the 10-point and a_n form an anti-chain for each $n \in \omega$. Therefore for each $n \in \omega$ there exists a point b_n totally covered by a_n and the 10-point. We will now argue that $\{b_n : n \in \omega\}$ is a countable anti-chain. Indeed, let $m \neq n$. We have that $\neg(a_n R^{\star} b_m)$, which implies that $\neg(b_n R^{\star} b_m)$ by the definition of b_n. By a symmetric argument, we also have $\neg(a_m R^{\star} b_n)$ and thus $\neg(b_m R^{\star} b_n)$. This concludes the proof of the lemma. □

3.2 Positive Morphisms

We will now recall the definition of positive morphisms between descriptive frames and models. In [6, Sect. 9.1] these morphisms are called *dense subreductions*. Positive morphisms are closely related to *strong partial Esakia morphisms* of [2]. However, strong Esakia morphisms satisfy additional conditions which guarantee a duality between these morphisms and $(\wedge, \vee, \to)$-homomorphisms between Heyting algebras.

Given two intuitionistic models (W, R, V) and (W', R', V') and a partial map $f : W \to W'$, for each $X \subseteq W'$ we let $f^*(X) = W \backslash R^{-1}(f^{-1}[W' \backslash X])$.

Definition 22. *Let* (W, R, V) *and* (W', R', V') *be models. A* positive morphism *is a partial map* $f : (W, R, V) \to (W', R', V')$ *such that:*

1. *If* $w, v \in \mathrm{dom}(f)$ *and* wRv *then* $f(w)R'f(v)$ (forth condition)*;*
2. *If* $w \in \mathrm{dom}(f)$ *and* $f(w)R'v$*, then there exists some* $u \in \mathrm{dom}(f)$ *such that* $f(u) = v$ *and* wRu (back condition)*;*
3. *If* $w \in \mathrm{dom}(f)$ *and* vRw*, then* $v \in \mathrm{dom}(f)$*;*
4. *For every* $p \in \mathrm{Prop}$ *we have* $V(p) = f^*(V'(p))$*.*

The last condition of Definition 22 guarantees that $f^*(V'(p))$ is an admissible upset. For descriptive frames $(W, R, \mathcal{P})$ and $(W', R', \mathcal{P}'')$, the corresponding condition is: $U \in \mathcal{P}'$ implies $f^*(U) \in \mathcal{P}$, which ensures that f^* is a well-defined map between $\mathcal{P}'$ and $\mathcal{P}$. In fact, conditions (1)-(2) ensure that it preserves $\wedge$ and $\to$, and condition (3) yields that it also preserves $\vee$, so f^* is a $(\wedge, \vee, \to)$-homomorphism.

Lemma 23. *Let* $f : W \to W'$ *be a positive morphism. If* $X \subseteq W'$ *is an upset of* W'*, then* $f^*(X) = f^{-1}[X] \cup (W \backslash \mathrm{dom}(f))$*.*

Proof. Let X be an upset of W'. Then $W' \backslash X$ is a downset of W'. By Definition 22(3), $w \in \mathrm{dom}(f)$ and $u \in R^{-1}(w)$ imply $u \in \mathrm{dom}(f)$ and Definition 22(1) yields $f(u)Rf(w)$. Since $W' \backslash X$ is a downset, $w \in f^{-1}[W' \backslash X]$ implies $u \in f^{-1}[W' \backslash X]$. Thus, $R^{-1}(f^{-1}[W' \backslash X]) = f^{-1}[W' \backslash X]$.

Therefore, we have that $f^*(X) = W \backslash R^{-1}(f^{-1}[W' \backslash X]) = W \setminus f^{-1}[W' \backslash X]$. But $W \backslash f^{-1}[W' \backslash X] = f^{-1}[X] \cup (W \backslash \mathrm{dom}(f))$, which finishes the proof of the lemma. $\square$

We will now give a more convenient characterization of positive morphisms.

Lemma 24. *A partial function* $f : (W, R, V) \to (W', R', V')$ *is a positive morphism iff the following conditions hold:*

1*. *If* $w, v \in \mathrm{dom}(f)$ *and* wRv *then* $f(w)R'f(v)$*;*
2*. *If* $w \in \mathrm{dom}(f)$ *and* $f(w)R'v$ *then there exists some* $u \in \mathrm{dom}(f)$ *such that* $f(u) = v$ *and* wRu*;*
3*. *If* $w \in \mathrm{dom}(f)$ *and* vRw*, then* $v \in \mathrm{dom}(f)$*;*
4*. *For every* $p \in \mathrm{Prop}$ *and* $w \in \mathrm{dom}(f)$ *we have* $w \in V(p) \iff f(w) \in V'(p)$*;*
5*. $\mathrm{dom}(f) \supseteq \{w \in W : \exists p \in \mathrm{Prop}\ w \notin V(p)\}$.

Proof. We need to prove that under the assumptions (1)–(3) of the definition of positive morphisms, (4) is equivalent to (4*) and (5*).

Let us assume (4*) and (5*). By Lemma 23 we have that $f^*(V'(p)) = f^{-1}[V'(p)] \cup W \backslash \mathrm{dom}(f)$. By (4*) we have that $f^{-1}[V'(p)] = V(p) \cap \mathrm{dom}(f)$. We also have that (5*) implies $W \backslash \mathrm{dom}(f) \subseteq V(p)$ since every element outside the domain of f satisfies all propositional variables. Therefore $V(p) = (V(p) \cap \mathrm{dom}(f)) \cup W \backslash \mathrm{dom}(f)$ and thus $f^*(V'(p)) = V(p)$.

For the other direction assume (4). Then $w \in \mathrm{dom}(f)$ and $f^{-1}[V'(p)] \cup W \backslash \mathrm{dom}(f) = V(p)$ yield that $w \in V(p)$ iff $f(w) \in V'(p)$. So (4*) holds. Also, for any $p \in \mathrm{Prop}$ we have that $W \backslash \mathrm{dom}(f) \subseteq V(p)$ and hence all elements outside the domain of f satisfy all propositional variables. So (5*) holds. $\square$

From now on, we will use this alternative characterization of positive morphisms. Obviously, every p-morphism is a positive morphism. Moreover, notice that if for all $w \in W$, there is some propositional variable p such that p is not satisfied in w, then the positive morphisms are p-morphisms. Finally, it is easy to check that the composition of two positive morphisms is a positive morphism.

The essential difference between p-morphisms and positive morphisms is that the latter are partial maps – domains of such maps may not contain worlds that satisfy all propositional variables. The reason why we can ignore these worlds when dealing with the positive fragment of IPC lies in a simple fact (which can be easily checked by induction) that in such worlds all positive formulas are true. Next we show that positive morphisms preserve positive formulas.

Proposition 25. *Let $f : (W, R, V) \to (W', R', V')$ be a positive morphism. Then for every positive formula φ and $w \in \mathrm{dom}(f)$ we have*

$$(W, R, V), w \models \varphi \quad \textit{iff} \quad (W', R', V'), f(w) \models \varphi.$$

Proof. We proceed by induction on the complexity of φ. The base case, i.e. when φ is a propositional variable, follows directly from Lemma 24(4). Now suppose that f preserves the positive formulas φ and ψ. That f also preserves $\varphi \vee \psi$ and $\varphi \wedge \psi$ trivially follows from the semantic definitions of the connectives and the induction hypothesis.

Let us now assume that $(W, R, V), w \models \varphi \to \psi$. Let $f(w)R'v$ and assume that $(W', R', V'), v \models \varphi$. Then by the definition of the positive morphisms, there is some $u \in \mathrm{dom}(f)$ such that $f(u) = v$ and wRu. By the induction hypothesis, we have $(W, R, V), u \models \varphi$. Hence $(W, R, V), u \models \psi$, which by the induction hypothesis gives us that $(W', R', V'), f(u) \models \psi$. So $(W', R', V'), f(w) \models \varphi \to \psi$.

For the converse direction, let us assume that $(W', R', V'), f(w) \models \varphi \to \psi$ and for some u such that wRu we have $(W, R, V), u \models \varphi$. If $u \in \mathrm{dom}(f)$, then the induction hypothesis readily implies that $(W, R, V), u \models \psi$. If $u \notin \mathrm{dom}(f)$, then by Lemma 24(5) for every propositional variable p we have that $u \in V(p)$, which implies that $(W, R, V), u \models \psi$, since all positive formulas are true in such worlds. □

The next corollary is a consequence of the proposition above.

Corollary 26. *Every formula in $\mathcal{L}_{\wedge,\vee,\to} \cup \{\bot\}$ has the top-model property.*

Proof. Let $\mathfrak{M} = (W, R, V)$ be an arbitrary Kripke model. We define a partial map $f : \mathfrak{M}^+ \to \mathfrak{M}$ such that it is the identity on all the elements of W and it is undefined in the top node. It is easy to see that f is a positive morphism. The result now follows directly from Proposition 25 for positive formulas. Finally, notice that for $\bot$ the result is trivially true. □

By the construction of the two universal models, we can see that $\mathcal{U}^\star(n)$ contains all the points in $\mathcal{U}(n)$ which are not below the node where all propositional variables are true. Therefore, it follows that $\mathcal{U}^\star(n)$ is isomorphic to $\mathcal{N} = (N, R, V)$ with $N = \{w \in \mathcal{U}(n) : w^\star \notin R(w)\}$, a generated submodel of $\mathcal{U}(n)$, where $w^\star$ is the greatest node of $\mathcal{U}(n)$ such that $\mathrm{col}(w^\star)_i = 1$ for every $i \leq n$. By Corollary 26, $(\mathcal{U}^\star(n))^+$ satisfies the same positive formulas as $\mathcal{U}^\star(n)$. Again, by the construction of the models, it follows that $(\mathcal{U}^\star(n))^+$ is (isomorphic to) a generated submodel of $\mathcal{U}(n)$, whose domain consists of the elements of $U(n)$ whose only successor of depth 1 satisfies all propositional variables. Let us call this submodel $\mathcal{M}(n)$, and let $G : (\mathcal{U}^\star(n))^+ \to \mathcal{M}(n)$ be this isomorphism.

The models $\mathcal{U}^\star(n)$ and $(\mathcal{U}^\star(n))^+$ can be viewed as two different ways of describing the universal models of the positive fragment of IPC. In the first approach, there are IPC-satisfiable positive formulas (for example $p_1 \wedge \ldots \wedge p_n$) that are satisfied nowhere in $\mathcal{U}^\star(n)$ and hence are indistinguishable from $\bot$ in this model. This is not the case in $\mathcal{U}(n)$, where every IPC-satisfiable formula is satisfied in some world. In $(\mathcal{U}^\star(n))^+$ all positive formulas are satisfied at the topmost point, and hence this model can distinguish positive formulas from $\bot$. As we will see below, every finite rooted model can be mapped onto a generated submodel of $\mathcal{U}^\star(n)$ via a positive morphism, which is not the case for $(\mathcal{U}^\star(n))^+$. On the other hand, for every finite rooted model $\mathfrak{M}$, the model $\mathfrak{M}^+$ can be mapped onto a generated submodel of $(\mathcal{U}^\star(n))^+$ via a p-morphism.

Lemma 27. *There exists a surjective positive morphism $F : \mathcal{U}(n) \to \mathcal{U}^\star(n)$ with* $\mathrm{dom}(F) = \{w \in \mathcal{U}(n) : \exists p \in \mathrm{Prop}(w \notin V(p))\}$*, and for every $w \in \mathrm{dom}(F)$ we have that the restriction of F to $\mathcal{U}(n)_w$ maps $\mathcal{U}(n)_w$ onto $\mathcal{U}^\star(n)_{F(w)}$.*

Proof. We will define F by induction on the depth of the elements of $\mathcal{U}(n)$ in such a way that the color of $F(w)$ is the same as the color of w. If $d(w) = 1$, then $F(w) = w'$, where $d(w') = 1$ and $\mathrm{col}(w) = \mathrm{col}(w')$. Let us now assume that F is defined for the elements of $\mathcal{U}(n)$ of depth m. Let $d(w) = m + 1$ and let us assume that $w \prec \{w_1, \ldots, w_k\}$. Let $A \subseteq F[\{w_1, \ldots, w_k\}]$ be the set of the R-minimal elements of $F[\{w_1, \ldots, w_k\}]$. Then A is finite as it is a subset of a finite set. If A is empty then let $F(w)$ be the element of $\mathcal{U}^\star(n)$ of depth 1 with the same color as w. If $A = \{u\}$ and u has the same color as w, then let $F(w) = u$. Otherwise, by the construction of $\mathcal{U}^\star(n)$, there is a unique $v \prec A$ (by the induction hypothesis for F) with the same color as w and we let $F(w) = v$.

It remains to show that F is a surjective positive morphism, and that for every $w \in \mathrm{dom}(F)$, the restriction of F to $\mathcal{U}(n)_w$ maps $\mathcal{U}(n)_w$ onto $\mathcal{U}^\star(n)_{F(w)}$.

That $w \in V(p)$ if an only if $F(w) \in V^\star(p)$ follows from the construction of F. It is also easy to see by the above construction that if uRw then $F(u)R^\star F(w)$. The surjectivity of F can be shown by viewing $\mathcal{U}^\star(n)$ as the generated submodel $\mathcal{N}$ of $\mathcal{U}(n)$ presented above. Then it is routine to check that F is the identity function on $\mathcal{N}$.

Next we show that the restriction of F to $\mathcal{U}(n)_w$ maps $\mathcal{U}(n)_w$ onto $\mathcal{U}^\star(n)_{F(w)}$. Since all elements of $\mathcal{U}^\star(n)$ have finite depth, it suffices to show that for all $u \in \mathrm{dom}(F)$, all the immediate successors of $F(u)$ are images of successors of u. Indeed, from the definition of F, the immediate successors of $F(w)$ form a subset of $F[\{w_1, \ldots, w_k\}]$, where $w_1, \ldots, w_k$ are the only immediate successors of w. Therefore, by an easy induction on immediate successors we can show that every element in $\mathcal{U}^\star(n)_{F(w)}$ is the image of some element in $\mathcal{U}(n)_w$.

Finally, the back clause that $F(w)R^\star v$ implies the existence of some u with wRu and such that $F(u) = v$ follows from the fact that the restriction of F to $\mathcal{U}(n)_w$ maps $\mathcal{U}(n)_w$ onto $\mathcal{U}^\star(n)_{F(w)}$. □

In the proof of Lemma 27 the map F is defined explicitly. An alternative proof of this lemma can be obtained by describing the same map F indirectly in the following way. Let us fix the injective partial map $i : (\mathcal{U}^\star(n))^+ \to \mathcal{U}^\star(n)$ between the two versions of the universal models to be the identity on $\mathcal{U}^\star(n)$ and undefined in the top node of $(\mathcal{U}^\star(n))^+$ (similarly to the positive morphism defined in the proof of Corollary 26). Moreover, by Proposition 25, for every $w \in U(n)$ we have that w satisfies the same positive formulas in $\mathcal{U}(n)$ and $(\mathcal{U}(n))^+$. Furthermore, by Lemma 8, there exists a unique p-morphism f_w from $((\mathcal{U}(n))^+)_w$ to $\mathcal{U}(n)$, and in particular to $\mathcal{M}(n)$ (see the paragraph after Corollary 26), since $\mathcal{M}(n)$ is a top model. By the uniqueness, we have that $f = \bigcup_{w\in U(n)} f_w$ is a p-morphism from $(\mathcal{U}(n))^+$ onto $\mathcal{M}(n)$. Then we can define $F = i \circ G^{-1} \circ f$ (where G is as in the paragraph after Corollary 26). This function is a positive morphism since it is a composition of positive morphisms. It is onto because f covers $\mathcal{M}(n)$ and G^{-1} and i are onto. Finally, since f was a union of maps from $((\mathcal{U}(n))^+)_w$ onto $\mathcal{U}(n)_{f(w)}$, it follows that the restriction of F to $\mathcal{U}(n)_w$ maps $\mathcal{U}(n)_w$ onto $\mathcal{U}^\star(n)_{F(w)}$.

Lemma 27 gives analogues of Lemma 8 and Theorem 9 for positive morphisms.

Lemma 28. *Let $\mathfrak{M} = (W, R, V)$ be a finite rooted intuitionistic n-model such that there exist $x \in W$ and $p \in \mathrm{Prop}$ with $x \notin V(p)$. Then there exist a unique $w \in U^\star(n)$ and a unique positive morphism f mapping $\mathfrak{M}$ onto $\mathcal{U}^\star(n)_w$.*

Proof. Given any finite rooted intuitionistic n-model $\mathfrak{M}$, Lemma 8 implies that there is a unique $w \in U(n)$ and a p-morphism f from $\mathfrak{M}$ onto $\mathcal{U}(n)_w$. By taking the F from Lemma 27, it follows that $F \circ f$ (with domain $\{x \in W : f(x) \in \mathrm{dom}(F)\}$) is a positive morphism (as a composition of positive morphisms) of $\mathfrak{M}$ onto $\mathcal{U}^\star(n)_{F(w)}$. Finally, since there exist $x \in W$ and $p \in \mathrm{Prop}$ such that $x \notin V(p)$ it follows that $\mathrm{dom}(F \circ f) \neq \varnothing$.

To show the uniqueness, we first observe that given two positive morphisms g_1, g_2 from $\mathfrak{M}$ to $\mathcal{U}^\star(n)$, we have

$$\mathrm{dom}(g_1) = \mathrm{dom}(g_2) = \{x \in W : \exists p \in \mathrm{Prop}(x \notin V(p))\},$$

because there do not exist points of $\mathcal{U}^\star(n)$ that satisfy all $p \in \mathrm{Prop}$. Notice that when restricted to $\mathrm{dom}(g_1)$ both g_1 and g_2 become p-morphisms. Thus, if $g_1 \neq g_2$, then there exist two different p-morphisms (g_1 and g_2) from $\mathrm{dom}(g_1)$ to $\mathcal{U}(n)$ (since $\mathcal{U}^\star(n)$ is a generated subframe of $\mathcal{U}(n)$), contradicting Lemma 8. □

The next theorem shows that $\mathcal{U}^\star(n)$ is indeed a "universal model" for all positive formulas.

Theorem 29. *For every positive n-formula φ, $\mathcal{U}^\star(n) \models \varphi$ iff $\vdash_{\mathsf{IPC}} \varphi$.*

Proof. The right to left direction is trivial. For the converse, let us assume that $\nvdash_{\mathsf{IPC}} \varphi$, i.e. there is a finite rooted model $\mathfrak{M}$ such that $\mathfrak{M}, x \nvDash \varphi$, where x is the root of $\mathfrak{M}$. Since φ is positive, we have that x does not satisfy all propositional variables. Then, by Lemma 28, there exists a unique $w \in U^\star(n)$ and a positive morphism f from $\mathfrak{M}$ onto $\mathcal{U}^\star(n)_w$. By Proposition 25, it follows that $\mathcal{U}^\star(n), f(x) \nvDash \varphi$. □

3.3 The Jankov-de Jongh Formulas

We will now define the de Jongh formulas for the positive fragment of IPC (for the description of the de Jongh formulas for the $[\wedge, \rightarrow]$-fragment of IPC, see [5]). These will be used in the next section for proving Jankov's theorem. We will present two ways of constructing the formulas: one that mirrors the construction of the standard de Jongh formulas, and one that derives the formulas through the procedure cited in Sect. 2.3. For $w \in U^\star(n)$ let $\mathrm{prop}(w), \mathrm{newprop}(w)$ and $\mathrm{notprop}(w)$ be defined as for the elements of $U(n)$.

Definition 30. *Let $w \in U^\star(n)$. We will define the formulas $\varphi_w^\star$ and $\psi_w^\star$ by induction on the depth of w:*

- *If $d(w) = 1$, then define*

$$\varphi_w^\star = \bigwedge \mathrm{prop}(w) \wedge (\bigvee \mathrm{notprop}(w) \rightarrow \bigwedge \mathrm{notprop}(w))$$

  and

$$\psi_w^\star = \varphi_w^\star \rightarrow \bigwedge_{i \in n} p_i.$$

- *If $d(w) > 1$, then let $w \prec \{w_1, \ldots, w_r\}$ and define*

$$\varphi_w^\star = \bigwedge \mathrm{prop}(w) \wedge (\bigvee \mathrm{newprop}(w) \vee \bigvee_{i \leq r} \psi_{w_i}^\star \rightarrow \bigvee_{i \leq r} \varphi_{w_i}^\star)$$

  and

$$\psi_w^\star = \varphi_w^\star \rightarrow \bigvee_{i \leq r} \varphi_{w_i}^\star.$$

The construction is motivated by the following observation: As we noted $(\mathcal{U}^\star(n))^+$ is a generated submodel of $\mathcal{U}^\star(n)$. Using the original de Jongh formula φ_w, for w the greatest element of $(\mathcal{U}^\star(n))^+$, we can define the de Jongh formulas from depth 2, using exactly the same construction as for the standard de Jongh formulas. Only now there is no need to take into consideration the ψ_w formula. This is because every positive formula is satisfied in a world that satisfies all propositional variables, and hence all positive formulas are true in w.

The above leads to the second way of constructing the de Jongh formulas for $\mathcal{U}^\star(n)$.

Definition 31. *For every $w \in \mathcal{U}^\star(n)$, we define $\varphi_w^\star$ and $\psi_w^\star$ as $[\varphi_{G(w)}]^*$ and $[\psi_{G(w)}]^*$ respectively, where $[\cdot]^*$ is the operation discussed in Proposition 19.*

The next proposition shows that the two definitions are in fact equivalent.

Proposition 32. *The formulas defined in Definitions 30 and 31 are equivalent.*

Proof. The proof is by induction on the depth of w. For $d(w) = 1$, we note that $[\varphi_{G(w)}]^*$ is $\bigwedge \text{prop}(w) \wedge (\bigvee \text{notprop}(w) \to \bigwedge \text{prop}(w) \wedge \bigwedge \text{notprop}(w))$, which is clearly equivalent to $\bigwedge \text{prop}(w) \wedge (\bigvee \text{notprop}(w) \to \bigwedge \text{notprop}(w))$. So $\varphi^\star_w$ is equivalent to $[\varphi_{G(w)}]^*$. Next we show that $\psi^\star_w$ is equivalent to $[\psi_{G(w)}]^*$. Since $G(w)$ is of depth 2 and its only successor is the node $w^\star$ where all propositional variables are true, by Proposition 19, $[\psi_{G(w)}]^* = [\varphi_{G(w)} \to \varphi_{w^\star}]^* = [\varphi_{G(w)}]^* \to [\varphi_{w^\star}]^* = [\varphi_{G(w)}]^* \to \bigwedge_{i\in n} p_i$, which is equivalent to $\psi^\star_w = \varphi^\star_w \to \bigwedge_{i\in n} p_i$.

For $d(w) = k + 1$, since φ and ψ formulas are inductively constructed in the same manner (in Definitions 10 and 30), by the induction hypothesis and the preservation of operations mentioned in Proposition 19, the equivalence follows immediately. ☐

We can now show that these formulas are indeed "positive analogues" of the standard de Jongh formulas (see Proposition 11).

Proposition 33. *For every $w \in \mathcal{U}^\star(n)$, we have:*

- $V^\star(\varphi_w) = R^\star(w)$;
- $V^\star(\psi_w) = \mathcal{U}^\star(n) \backslash (R^\star)^{-1}(w)$.

Proof. By Proposition 19 we have that any formula σ is equivalent to $[\sigma]^*$ in the top models. Hence $\varphi^\star_w$ is satisfied in the same worlds of $\mathcal{M}(n)$ (which is isomorphic to $(\mathcal{U}^\star(n))^+$, see the paragraph after Corollary 26) as φ_w (and likewise for ψ_w). But since $\varphi^\star_w$ are positive formulas, by Corollary 26, they will be satisfied in the same worlds in $(\mathcal{U}^\star(n))$. ☐

The proposition above implies that two distinct points of $\mathcal{U}^\star(n)$ can be distinguished via a positive formula. Indeed, if $w_1 \neq w_2$ are two worlds in $\mathcal{U}^\star(n)$, then either $\neg(w_1 R w_2)$ or $\neg(w_2 R w_1)$. In the first case $\mathcal{U}^\star(n), w_2 \not\models \varphi_{w_1}$, while in the second case $\mathcal{U}^\star(n), w_1 \not\models \varphi_{w_2}$.

4 *n*-Henkin Models

Let us denote the n-Henkin model for the positive fragment of IPC by $\mathcal{H}^\star(n)$. We write

$$\text{Cn}^\star_n(\varphi) = \{\psi \in \mathcal{L}_{\wedge,\vee,\to} : \psi \text{ is an } n\text{-formula and } \vdash_{\mathsf{IPC}} \varphi \to \psi\}$$

and

$$\text{Th}^\star_n(\mathfrak{M}, w) = \{\varphi \in \mathcal{L}_{\wedge,\vee,\to} : \varphi \text{ is an } n\text{-formula and } \mathfrak{M}, w \models \varphi\}.$$

The following proposition is analogous to Proposition 12.

Proposition 34. *For any point $w \in \mathcal{U}^\star(n)$ we have $\text{Th}^\star_n(\mathcal{U}^\star(n), w) = \text{Cn}^\star_n(\varphi^\star_w)$.*

Proof. It follows from Proposition 33 that the right hand side is a subset of the left hand side. For the other direction, assume $\mathcal{U}^\star(n), w \models \sigma$. Then if $\nvdash_{\mathsf{IPC}} \varphi^\star_w \to \sigma$, there is a finite model $\mathfrak{M}$ whose root, x, satisfies $\varphi^\star_w$ and does not satisfy σ. Then, since x does not satisfy all positive formulas, it does not satisfy all propositional variables. Hence there is a positive morphism f with non-empty domain from $\mathfrak{M}$ to $\mathcal{U}^\star(n)$. Since x satisfies φ_w, by Proposition 25 we have that $f(x)$ also satisfies φ_w. By Proposition 33, this implies that $f(x) \in R^\star(w)$. Finally, since $\mathcal{U}^\star(n), w \models \sigma$ we get $\mathcal{U}^\star(n), f(x) \models \sigma$, which contradicts Proposition 25 as $\mathfrak{M}, x \nvDash \sigma$. □

The next lemma will be used in the proof that the universal model is isomorphic to the upper part of the Henkin model (Theorem 37).

Lemma 35. *Let Γ be an n-theory of the positive fragment of* IPC*. If $\Gamma \supseteq \mathrm{Cn}^\star_n(\varphi^\star_w)$ for some $w \in U^\star(n)$, then either there exists some $v \in R^\star(w)$ such that $\Gamma = \mathrm{Cn}^\star_n(\varphi^\star_v)$, or Γ contains all positive formulas.*

Proof. Let $\Gamma \supseteq \mathrm{Cn}^\star_n(\varphi^\star_w)$ and let v be such that wRv and $\varphi^\star_v \in \Gamma$ while for all immediate successors of v (let $v_1, \ldots, v_k$ be all the immediate successors of v) we have that $\Gamma \cap \{\varphi^\star_{v_1}, \ldots, \varphi^\star_{v_k}\} = \varnothing$.

If this v is unique we can see that $\Gamma = \mathrm{Cn}^\star_n(\varphi^\star_v)$. The right to left inclusion is trivial. For the converse inclusion we observe that for every $\sigma \in \Gamma$ we have $\sigma \wedge \varphi^\star_v \nvdash \varphi^\star_{v_1} \vee \cdots \vee \varphi^\star_{v_k}$ which implies by Theorem 29 that there is a point of $\mathcal{U}^\star(n)$ that satisfies $\sigma \wedge \varphi^\star_v$ but not $\varphi^\star_{v_1} \vee \cdots \vee \varphi^\star_{v_k}$. By Proposition 33, there is only one such element, v. Hence $\sigma \in \mathrm{Th}^\star_n(\mathcal{U}^\star(n), v)$, which by Proposition 34 means that $\sigma \in \mathrm{Cn}^\star_n(\varphi^\star_v)$.

To complete the proof, we will show that the aforementioned v is unique or has depth 1. If $d(v) > 1$ and there is an element u ($v \neq u$) with the aforementioned property, then Proposition 33 implies that $\neg(vR^\star w)$ and $\neg(wR^\star v)$ and hence $\psi^\star_v \in \mathrm{Th}^\star_n(\mathcal{U}^\star(n), u)$, thus $\psi^\star_v \in \Gamma$. Therefore, since Γ has the disjunction property, there is some immediate successor v_i of v, such that $\varphi^\star_{v_i} \in \Gamma$. This is a contradiction. So if $d(v) > 1$, then v is unique.

Finally, if $\varphi^\star_v, \varphi^\star_u \in \Gamma$, where $v \neq u$ and $d(v) = d(u) = 1$, then we can assume without loss of generality that there is some propositional variable q true in v but not true in u. By the definition of $\varphi^\star_v$ we have that $q \in \Gamma$. By the definition of $\varphi^\star_u$ we have that $q \to \bigwedge_{i \leq n} p_i \in \Gamma$. Hence all propositional variables are in Γ, which implies that Γ contains all positive formulas. □

The next three statements are the positive-fragment analogues of Lemmas 13 and 14 and Proposition 15, respectively. Notice that the n-Henkin model here contains a top point where every positive formula is true.

Lemma 36. *For any $w \in \mathcal{U}^\star(n)$ we have that $\mathcal{H}^\star(n)_{\mathrm{Cn}^\star(\varphi^\star_w)}$ is isomorphic to $(\mathcal{U}^\star(n)_w)^+$.*

Proof. We will show that the function $g : (\mathcal{U}^\star(n)_w)^+ \to \mathcal{H}^\star(n)_{\mathrm{Cn}^\star(\varphi^\star_w)}$, where $g(v) = \mathrm{Cn}^\star_n(\varphi^\star_v)$ and the topmost element is mapped to the set of all positive formulas, is the required isomorphism. Proposition 33 implies that g is injective and Lemma 35 implies that g is surjective. The frame relations are preserved back and forth by the following chain of equivalences:

$$\begin{array}{lll} & uR^\star v & \\ \text{iff} & R^\star(v) \subseteq R^\star(u) & \\ \text{iff} & V^\star(\varphi^\star_v) \subseteq V^\star(\varphi^\star_u) & \text{(Proposition 33)} \\ \text{iff} & \vdash_{\mathsf{IPC}} \varphi^\star_v \to \varphi^\star_u & \text{(Theorem 29)} \\ \text{iff} & \varphi^\star_u \in \mathrm{Cn}^\star_n(\varphi^\star_v) & \\ \text{iff} & \mathrm{Cn}^\star_n(\varphi^\star_u) \subseteq \mathrm{Cn}^\star_n(\varphi^\star_v) & \\ \text{iff} & g(u) \subseteq g(v). & \end{array}$$

□

The next theorem shows that in the same way n-universal models for IPC are the "upper-parts" of the n-Henkin models, the n-universal models for positive IPC are the "upper-parts" of the n-Henkin models of positive IPC.

Theorem 37. $\mathrm{Upper}(\mathcal{H}^\star(n))$ *is isomorphic to* $(\mathcal{U}^\star(n))^+$.

Proof. As above, the isomorphism will be given by the function $g : (\mathcal{U}^\star(n))^+ \to \mathrm{Upper}(\mathcal{H}^\star(n))$, such that $g(v) = \mathrm{Cn}^\star_n(\varphi^\star_v)$ and the topmost element is mapped to the set of all positive formulas. That this map is injective follows from Proposition 34 and the fact that two distinct points of $(\mathcal{U}^\star(n))^+$ are separated by a positive formula (see the paragraph after Proposition 33). That the map preserves the relation follows from the fact that intuitionistic truth is upward preserving. What is left to show is that it is onto. Let $x \in \mathrm{Upper}(\mathcal{H}^\star(n))$, and x does not contain all positive formulas. Then, by Lemma 28, there is a positive morphism, f (which is non-empty by the assumptions for x) from $\mathrm{Upper}(\mathcal{H}^\star(n))_x$ onto some $\mathcal{U}^\star(n)_w$. Then we observe by Proposition 25 that $\mathrm{Th}^\star_n(\mathcal{U}^\star(n), w) = x$, i.e., by Proposition 34, $x = \mathrm{Cn}^\star_n(\varphi^\star_w)$. Therefore, g is surjective. □

Corollary 38. *Let* $\mathfrak{M} = (W, R, V)$ *be any* n*-model and let* $X \subseteq V(\varphi^\star_w)$ *be a non-empty set for some* $w \in U^\star(n)$*. Then there is a unique positive morphism* f *from* $\mathfrak{M}_X$ *to* $\mathcal{U}^\star(n)_w$*. Furthermore, if* $\mathfrak{M}_X$ *is rooted and does not satisfy all positive formulas, then there is a unique* $v \in U^\star(n)$ *with* $wR^\star v$ *and such that* f *maps* $\mathfrak{M}_X$ *onto* $\mathcal{U}^\star(n)_v$*.*

Proof. Since $X \subseteq V(\varphi^\star_w)$, for each $x \in X$ and $y \in W$ with xRy we have $\mathrm{Th}^\star_n(\mathfrak{M}, y) \supseteq \mathrm{Cn}^\star_n(\varphi^\star_w)$. By Lemma 35 such a theory is equal to some $\mathrm{Cn}^\star_n(\varphi^\star_v)$ or contains all positive formulas. We define a positive morphism f as follows:

$$f(y) = \begin{cases} u, & \text{if } \exists u \text{ such that } \mathrm{Th}^\star_n(\mathfrak{M}, y) = \mathrm{Cn}^\star_n(\varphi^\star_u); \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

If the domain of f is empty then it is vacuously a positive morphism. If the domain is non-empty, by the definition of f the only non-trivial step to show that f is a positive morphism is the back condition. For this we have: if $vR^\star u$ and $f(y) = v$, then by Proposition 33, it is the case that $\mathfrak{M}, y \nvDash \psi^\star_u$. Hence there is some $z \in W$ with yRz such that $\mathfrak{M}, z \models \varphi^\star_u$ and $\mathfrak{M}, z \nvDash \bigvee_{i \leq l} \varphi^\star_{u_i}$. This yields that $\mathrm{Th}^\star_n(\mathfrak{M}, z) = \mathrm{Cn}^\star_n(\varphi^\star_u)$, i.e. $f(z) = u$.

Finally, if $\mathfrak{M}_X$ is rooted and does not satisfy all positive formulas, then the root, x, is in the domain of f. Then we let $v = f(x)$. The back condition immediately yields that f is onto. □

Note that the underlying Kripke frame of $\mathcal{U}^\star(n)_w = (U^\star(n)_w, R^\star(n)_w, V^\star(n)_w)$ described in the previous lemma can be viewed as the general frame $(U^\star(n)_w, R^\star(n)_w, Up(\mathcal{U}^\star(n)_w))$, which is a descriptive frame since W is finite.

5 Jankov's Theorem for KC

In this section, we will first prove an analogue of the Jankov-de Jongh theorem (Theorem 16). This theorem will be used afterwards for an alternative proof of Jankov's theorem for KC.

Theorem 39 (Jankov-de Jongh theorem for positive fragment of IPC). *For every descriptive frame $\mathfrak{G}$ and $w \in U^\star(n)$ we have that $\mathfrak{G} \nvDash \psi^\star_w$ iff there is an n-valuation V on $\mathfrak{G}$ such that $\mathcal{U}^\star(n)_w$ is the image, through a positive morphism, of a generated submodel of $(\mathfrak{G}, V)$.*

Proof. Let $\mathcal{U}^\star(n)_w$ be the image, through a positive morphism f, of a generated submodel $\mathcal{K}$ of $(\mathfrak{G}, V)$. Proposition 33 implies that $\mathcal{U}^\star(n)_w, w \nvDash \psi^\star_w$. Since f is a positive morphism, Proposition 25 yields that $\mathcal{K}, x \nvDash \psi^\star_w$ for every $x \in f^{-1}[\{w\}]$. Now, because $\mathcal{K}$ is a generated submodel of $(\mathfrak{G}, V)$, we have that $(\mathfrak{G}, V), x \nvDash \psi^\star_w$, i.e. $\mathfrak{G} \nvDash \psi^\star_w$.

For the other direction, let us assume that there is some valuation and some x such that $(\mathfrak{G}, V), x \nvDash \psi^\star_w$. This implies that there is some y_0 such that xRy_0 and $(\mathfrak{G}, V), y_0 \models \varphi^\star_w$, while $(\mathfrak{G}, V), y_0 \nvDash \varphi^\star_{w_i}$, for all immediate successors w_i of w.

We take $(\mathfrak{G}, V)_{V(\varphi^\star_w)}$, the submodel of $(\mathfrak{G}, V)$ generated by $V(\varphi^\star_w)$. We note that by the above observation $V(\varphi^\star_w) \neq \varnothing$. Furthermore, we have that $(\mathfrak{G}, V)_{V(\varphi^\star_w)}$ does not satisfy all positive formulas since $y_0 \in V(\varphi^\star_w)$ and $(\mathfrak{G}, V), y_0 \nvDash \varphi^\star_{w_i}$, for all immediate successors w_i of w.

Therefore, by Corollary 38, we have that there is a positive morphism f from $(\mathfrak{G}, V)_{V(\varphi^\star_w)}$ to $\mathcal{U}^\star(n)_w$. It is onto because $\mathrm{Th}^\star((\mathfrak{G}, V), y_0) = \mathrm{Cn}^\star_n(\varphi^\star_w)$ and hence $f(y_0) = w$.

Finally, we have that $(\mathfrak{G}, V)_{V(\varphi_w)}$ is a descriptive model, by Lemma 4, since it is based on $V(\varphi_w)$. To show that the positive morphism is also descriptive, we only need to show that $f^{-1}[R^\star(v)] \cup (\mathfrak{G} \backslash \mathrm{dom}(f)) = V(\varphi^\star_v)$, for $v \in \mathcal{U}^\star(n)_w$. For the left to right inclusion we observe that anything outside the domain of f satisfies all positive formulas and f preserves positive formulas. For the right to left assume that $x \in V(\varphi^\star_v)$. Then $x \in V(\varphi^\star_w)$ and by Lemma 35 we get that $f(x) \in R^\star(w)$ or x satisfies all propositional variables and hence it is not in the domain of f. □

We recall that KC is complete with respect to the finite frames with a topmost node. Thus, by reflecting on Corollary 26, one can easily see that KC proves exactly the same positive formulas as IPC. In [18], Jankov proved that KC is maximal with that property (see also [6, Ex. 9.17] for a proof via Zakharyaschev's canonical formulas). In [11] an alternative proof based on the universal model for IPC is given.

Using the universal model for positive formulas we will provide yet another proof of this theorem, more perspicuous than the one in [11]. For this we will need the following auxiliary lemma.

Lemma 40. *Let $\mathfrak{F}$ be a descriptive frame with a topmost element, let $\mathfrak{G}$ be a descriptive frame, V and V' be admissible valuations and $f : (\mathfrak{G}, V) \to (\mathfrak{F}, V')$ a descriptive positive morphism between models. Then f can be extended to a descriptive frame p-morphism.*

Proof. First assume that the map f is total. Then it is a frame p-morphism. Now suppose f is not total. Then we extend f to f' such that for every $y \in \mathfrak{G} \backslash \text{dom}(f)$ we have $f'(y) = x_0$, where x_0 is the topmost element of $\mathfrak{F}$. We claim that f' is the desired frame p-morphism. That the forth condition holds is easy to see, since everything in $\mathfrak{F}$ is below x_0. For the back condition the only possible problem may arise if some $f'(y)Rx_0$. In that case, if $y \in \text{dom}(f)$ then $f(y)Rx_0$ and by the definition of positive morphisms a witness for the back condition exists. If $y \notin \text{dom}(f)$ then the witness is y. It remains to show that the f'-pre-image of an admissible set is admissible. Let Q be an admissible set in $\mathfrak{F}$. By the construction of f we have that $f'^{-1}[Q] = f^{-1}[Q] \cup (\mathfrak{G} \backslash \text{dom}(f))$, which is admissible since, by Lemma 23, it is equal to $f^*(Q)$ and f is a positive morphism between descriptive frames. □

Finally, we will give our alternative proof of Jankov's theorem stating that KC is the greatest intermediate logic that proves exactly the same positive formulas as IPC.

Theorem 41. (*Jankov*) *For every logic $\mathcal{L} \nsubseteq$ KC there exists some positive formula σ such that $\mathcal{L} \vdash \sigma$ while IPC $\nvdash \sigma$.*

Proof. Let us assume that $\mathcal{L} \nsubseteq$ KC. Then $\mathcal{L} \vdash \chi$ and KC $\nvdash \chi$ for some formula χ. As KC is complete with respect to finite rooted frames with a topmost element (see, e.g., [6, Proposition 2.37 and Theorem 5.33]), there is a finite rooted frame with a topmost element, $\mathfrak{F} = (W, R)$ with $\mathfrak{F} \nvDash \chi$. We define a valuation, V, on $\mathfrak{F}$ such that each of its elements has a different color and that there is a propositional variable, q, not satisfied at the topmost element. A way to do this is to introduce a propositional variable p_x for each $x \in W$ such that $V(p_x) = R(x)$ and $V(q) = \varnothing$. By Lemma 28, there is some $w \in U(n)$ and a positive morphism from $(\mathfrak{F}, V)$ onto $\mathcal{U}^\star(n)_w$. Since each element of $(\mathfrak{F}, V)$ has a different color, the positive morphism is 1-1 and since in every element of W at least one propositional variable is not satisfied, the positive morphism has W as its domain, hence $(\mathfrak{F}, V)$ is isomorphic to $\mathcal{U}^\star(n)_w$.

We claim that the required positive formula, σ is $\psi^\star_w$. For contradiction, let us assume that $\mathcal{L} \nvdash \psi^\star_w$. Then, as every logic is complete with respect to descriptive frames (e.g., [6, Theorem 8.36]), there exists a descriptive $\mathcal{L}$-frame, $\mathfrak{G}$ such that

$\mathfrak{G} \not\models \psi_w^\star$. By Theorem 39 there is a valuation V' on $\mathfrak{G}$, a generated submodel $\mathcal{K}$ of $(\mathfrak{G}, V')$, and a descriptive positive morphism f, from $\mathcal{K}$ onto $(\mathfrak{F}, V)$. By Lemma 40, f can be extended to a descriptive frame p-morphism f'. Since $\mathfrak{G}$ is an $\mathcal{L}$-frame and $\chi \in \mathcal{L}$, we have that $\mathfrak{G} \models \chi$. As f' is a descriptive frame p-morphism, $\mathfrak{G} \models \chi$ implies that $\mathfrak{F} \models \chi$, contradicting the assumption that $\mathfrak{F} \not\models \chi$. □

6 Conclusions and Future Directions

In this paper we described the universal models for the positive fragment of IPC, and using these models gave an alternative proof of Jankov's theorem which states that the logic KC of the weak law of excluded middle is the greatest logic that proves the same positive formulas as IPC. The main technical ingredients of our proofs are positive morphisms and Jankov-de Jongh formulas.

We also briefly underline some future research directions. In this paper we do not discuss algebraic aspects of universal models for positive IPC. It would be interesting to describe in all detail the algebraic counterparts of these universal models together with a full duality theory for the corresponding algebras. We refer to recent work [3] and [7] for topological dualities for similar algebraic structures. Here we only give a small hint towards algebraic analogues of the two different n-universal models $\mathcal{U}^\star(n)$ and $(\mathcal{U}^\star(n))^+$ for positive IPC discussed in Sect. 3.

From an algebraic point of view, the two universal models correspond to the Lindenbaum-Tarski algebras for the languages $\mathcal{L}_{\wedge,\vee,\to}$ and $\mathcal{L}_{\wedge,\vee,\to} \cup \{\bot\}$, respectively. In fact, one can show that the definable upsets of $\mathcal{U}^\star(n)$ form an algebra isomorphic to the Lindenbaum-Tarski algebra of the positive IPC. On the other hand, the definable upsets of $(\mathcal{U}^\star(n))^+$ form an algebra which is isomorphic to the Lindenbaum-Tarski algebra of the positive IPC with an additional bottom element $\bot$.

Finally, we point out a connection with minimal logic. Minimal logic can be seen as arising from positive intuitionistic logic by interpreting one propositional variable as the falsum without giving it any special properties and defining negation in the standard manner. The n-universal model for minimal logic is therefore directly available as the $n+1$-universal model of positive intuitionistic logic developed above. Recently, minimal logic with negation as a primitive and its sublogics have been studied in [8]. Colacito extended this work in [7] with proof-theoretic and algebraic results using top frames. We believe that the universal models for positive intuitionistic logic described in this paper will find fruitful application in this area as will the construction of the accompanying Jankov-de Jongh formulas.

Acknowledgement. We would like to thank the referees of this paper for careful reading and many useful comments.

References

1. Bellissima, F.: Finitely generated free Heyting algebras. J. Symbolic Logic **51**, 152–165 (1986)
2. Bezhanishvili, G., Bezhanishvili, N.: An algebraic approach to canonical formulas: intuitionistic case. Rev. Symbolic Logic **2**(3), 517–549 (2009)
3. Bezhanishvili, G., Moraschini, T., Raftery, J.: Epimorphisms in varieties of residuated structures (2016, submitted)
4. Bezhanishvili, N.: Lattices of intermediate and cylindric modal logics. Ph.D. thesis, University of Amsterdam (2006)
5. Bezhanishvili, N., Coumans, D., Gool, S.J., de Jongh, D.: Duality and universal models for the meet-implication fragment of IPC. In: Aher, M., Hole, D., Jeřábek, E., Kupke, C. (eds.) TbiLLC 2013. LNCS, vol. 8984, pp. 97–116. Springer, Heidelberg (2015). doi:10.1007/978-3-662-46906-4_7
6. Chagrov, A., Zakharyaschev, M.: Modal Logic. Clarendon Press, Oxford Logic Guides (1997)
7. Colacito, A.: Minimal and subminimal logic of negation. Master's Thesis, University of Amsterdam (2016)
8. Colacito, A., de Jongh, D., Vargas Sandoval, A.L.: Subminimal negation. In Proceedings of Coherence and Truth, Conference in Memoriam Franco Montagna (2015). To be published in Soft Computing
9. de Jongh, D.: Investigations on the intuitionistic propositional calculus. Ph.D. thesis, University of Wisconsin (1968)
10. de Jongh, D., Hendriks, L., de Lavalette, G.R.: Computations in fragments of intuitionistic propositional logic. J. Autom. Reasoning **7**(4), 537–561 (1991)
11. de Jongh, D., Yang, F.: Jankov's theorems for intermediate logics in the setting of universal models. In: Bezhanishvili, N., Löbner, S., Schwabe, K., Spada, L. (eds.) TbiLLC 2009. LNCS (LNAI), vol. 6618, pp. 53–76. Springer, Heidelberg (2011). doi:10.1007/978-3-642-22303-7_5
12. de Jongh, D., Zhao, Z.: Positive formulas in intuitionistic and minimal logic. In: Aher, M., Hole, D., Jeřábek, E., Kupke, C. (eds.) TbiLLC 2013. LNCS, vol. 8984, pp. 175–189. Springer, Heidelberg (2015). doi:10.1007/978-3-662-46906-4_11
13. Diego, A.: Sur les Algèbres de Hilbert, vol. 21. E. Nauwelaert, Gauthier-Villars, Louvain (1966)
14. Ghilardi, S.: Irreducible models and definable embeddings. In: Gabbay, D., Csirmaz, L., de Rijke, M. (eds.) Logic Colloquium 92 (Proceedings), Studies in Logic, Language and Information, pp. 95–213. CSLI Publications (1995)
15. Goudsmit, J.: Intuitionistic rules: admissible rules of intermediate logics. Ph.D. thesis, Utrecht University (2015)
16. Grigolia, R.: Free Algebras of Non-Classical Logics. Metsniereba, Tbilisi (1987). (Russian)
17. Hendriks, L.: Computations in propositional logic. Ph.D. thesis, University of Amsterdam (1996)
18. Jankov, V.A.: Calculus of the weak law of the excluded middle (in Russian). Izv. Akad. Nauk SSSR Ser. Mat. **32**(5), 1044–1051 (1968)
19. Johansson, I.: Der Minimalkalkül, Ein Reduzierter Intuitionistischer Formalismus. Compositio Mathematica **4**, 119–136 (1937)
20. Renardel de Lavalette, G., Hendriks, A., de Jongh, D.: Intuitionistic implication without disjunction. J. Logic Comput. **22**(3), 375–404 (2012)

21. Rybakov, V.: Rules of inference with parameters for intuitionistic logic. J. Symbolic Logic **57**, 33–52 (1992)
22. Shehtman, V.: Rieger-Nishimura lattices. Sov. Math. Dokl. **19**, 1014–1018 (1978)
23. Yang, F.: Intuitionistic subframe formulas, NNIL-formulas and n-universal models. Master's Thesis, University of Amsterdam (2008)

On Gödel Algebras of Concepts

Pietro Codara[1(✉)] and Diego Valota[2]

[1] Artificial Intelligence Research Institute (IIIA-CSIC),
Campus de la UAB, 08193 Bellaterra, Spain
codara@di.unimi.it
[2] Dipartimento di Informatica, Università degli Studi di Milano,
Via Comelico 39, 20135 Milano, Italy
diego@iiia.csic.es

Abstract. Beside algebraic and proof-theoretical studies, a number of different approaches have been pursued in order to provide a complete intuitive semantics for many-valued logics. Our intention is to use the powerful tools offered by *formal concept analysis* (FCA) to obtain further intuition about the intended semantics of a prominent many-valued logic, namely Gödel, or Gödel-Dummett, logic. In this work, we take a first step in this direction. Gödel logic seems particularly suited to the approach we aim to follow, thanks to the properties of its corresponding algebraic variety, the class of *Gödel algebras*. Furthermore, Gödel algebras are prelinear Heyting algebras. This makes Gödel logic an ideal contact-point between intuitionistic and many-valued logics.

In the literature one can find several studies on relations between FCA and fuzzy logics. These approaches often amount to equipping both intent and extent of concepts with connectives taken by some many-valued logic. Our approach is different. Since Gödel algebras are (residuated) lattices, we want to understand which type of concepts are expressed by these lattices. To this end, we investigate the concept lattice of the standard context obtained from the lattice reduct of a Gödel algebra. We provide a characterization of Gödel implication between concepts, and of the Gödel negation of a concept. Further, we characterize a Gödel algebra of concepts. Some concluding remarks will show how to associate (equivalence classes of) formulæ of Gödel logic with their corresponding formal concepts.

Keywords: Intended semantics · Concept lattice · Many-valued logic · FCA · Formal concept analysis · Fuzzy logic · Gödel Logic

1 Introduction

Gödel logic can be semantically defined as a many-valued logic, as follows. Consider the set FORM of well-formed formulæ over propositional variables $\{x_1, x_2, x_3, \dots\}$ in the language $(\wedge, \vee, \to, \bot, \top)$. An *assignment* is a function μ from FORM to $[0, 1] \subseteq \mathbb{R}$, such that, for any $\varphi, \psi \in$ FORM,

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 251–262, 2017.
DOI: 10.1007/978-3-662-54332-0_14

$$\mu(\bot) = 0\,, \qquad \mu(\top) = 1\,,$$
$$\mu(\varphi \wedge \psi) = \min\{\mu(\varphi), \mu(\psi)\}\,,$$
$$\mu(\varphi \vee \psi) = \max\{\mu(\varphi), \mu(\psi)\}\,,$$
$$\mu(\varphi \to \psi) = \begin{cases} 1 & \text{if } \mu(\varphi) \le \mu(\psi)\,, \\ \mu(\psi) & \text{otherwise}\,. \end{cases}$$

A formula φ such that $\mu(\varphi) = 1$ for every assignment μ is called a *tautology*. To indicate such a case we write $\vDash \varphi$.

Gödel logic can also be syntactically defined as a schematic extension of intuitionistic propositional calculus by the *prelinearity axiom*

$$(\varphi \to \psi) \vee (\psi \to \varphi). \tag{P}$$

We write $\vdash \varphi$ to mean that the formula φ is derivable from the axioms of Gödel logic using *modus ponens* as the only deduction rule. Gödel logic is complete with respect to the many-valued semantics defined above: in symbols, $\vdash \varphi$ if and only if $\vDash \varphi$. Details and proofs can be found in [22].

Even though Gödel logic is an axiomatic extension of intuitionistic logic, the constructive intended semantics[1] of the latter is not suitable for the former. Indeed, think of formulæ of FORM as problems for which we have an algorithmic solution. Then, (P) states that, for every choice of φ and ψ in FORM, the solution to φ can be reduced to the solution to ψ, or the solution to ψ can be reduced to the solution to φ. A rather strong assumption. This is a common problem of informal intended semantics. They are tailored over a specific logic. Applying them to some extension is not straightforward, or not even possible.

On the other hand, beside algebraic and proof-theoretical studies, a number of different approaches have been attempted to provide semantics for Gödel logics. To mention a few, we cite [5,18], where temporal-like and game-theoretic semantics, respectively, are investigated.

The possibility of connecting descriptions of real-world contexts with powerful formal instruments is what makes *formal concept analysis* (FCA) a promising framework, merging the intuitions of intended semantics with the advantages of formal semantics. In the present work, we study formal contexts associated with Gödel logic from the algebraic point of view. The algebraic semantics of Gödel logic is the subvariety of Heyting algebras satisfying prelinearity. A *Heyting algebra* is a structure $\mathbf{A} = (A, \wedge, \vee, \to, \top, \bot)$ of type $(2, 2, 2, 0, 0)$ such that $(A, \wedge, \vee, \top, \bot)$ is a distributive lattice and the couple $(\wedge, \to)$ forms a *residuated pair*. This means that the unique operation $\to$ that satisfies the residuation property, $x \wedge z \le y$ if and only if $z \le x \to y$, is the *residuum* of $\wedge$, defined as

$$x \to y = \max\{z \mid x \wedge z \le y\}. \tag{1}$$

[1] The intended semantics of a logical language consists of the collection of models that intuitively the language talks about. In this specific case the intended semantics' is the informal description of *truth as provability* given by Brouwer.

Hence, a *Gödel algebra* is a Heyting algebra satisfying the prelinearity equation $(x \to y) \vee (y \to x) = \top$, for $x, y \in A$. Horn [23] showed that the variety of Gödel algebras is *locally finite*. That is, the classes of finite, finitely generated and finitely presented algebras coincide.

For an integer $n \geq 1$, let FORM_n be the set of all formulæ whose propositional variables are contained in $\{x_1, \ldots, x_n\}$. Two formulæ $\varphi, \psi \in \mathrm{FORM}_n$ are called *logically equivalent* if both $\vdash \varphi \to \psi$ and $\vdash \psi \to \varphi$ hold. Logical equivalence is an equivalence relation, denoted by $\equiv$. We denote the equivalence class of a formula φ by $[\varphi]_\equiv$. It is straightforward to see that the quotient set $\mathrm{FORM}_n/\equiv$, endowed with the operations $\wedge, \vee, \top, \bot$ induced by the corresponding logical connectives, is a distributive lattice with top and bottom element $\top$ and $\bot$, respectively. If, in addition, $\mathrm{FORM}_n/\equiv$ is endowed with the operation $\to$ induced by the logical implication, then $\mathrm{FORM}_n/\equiv$ becomes a Gödel algebra. The specific Gödel algebra $\mathcal{G}_n = \mathrm{FORM}_n/\equiv$ is, by construction, the *Lindenbaum algebra* of Gödel logic over the language $\{x_1, \ldots, x_n\}$. Lindenbaum algebras are isomorphic to free algebras, thus $\mathcal{G}_n$ is the free n-generated Gödel algebra. Moreover, since the variety of Gödel algebras is locally finite, every finite Gödel algebra can be obtained as a quotient of a free n-generated Gödel algebra. For the rest of this paper, all Gödel algebras are assumed to be finite.

In the next section, we recall some basic notions on FCA. In Sect. 3 we deal with the concept lattice $\mathbf{C_A}$ of the standard context obtained from a Gödel algebra $\mathbf{A}$. We prove that endowing $\mathbf{C_A}$ with a suitable implication between concepts, we obtain an algebra of concepts isomorphic to $\mathbf{A}$. Further, we characterize the Gödel negation in terms of concepts. In Sect. 4 we characterize Gödel algebras of concepts. In Sect. 5 we show how to associate concepts belonging to a Gödel algebras of concepts with Gödel logic formulæ. Finally, in Sect. 6 we discuss the integration of this approach with the studies on many-valued (substructural) logics aimed to investigate their intended semantics.

2 Basic Notions on FCA

We recollect the basic definitions and facts about formal concept analysis needed in this work. For further details on this topics we refer the reader to [20].

Recall that an element j of a distributive lattice L is called a *join-irreducible* if j is not the bottom of L and if whenever $j = a \vee b$, then $j = a$ or $j = b$, for $a, b \in L$. Meet-irreducible elements are defined dually. Given a lattice $L = (L, \sqcap, \sqcup, 1)$, we denote by $\mathfrak{J}(L)$ the set of its join-irreducible elements, and by $\mathfrak{M}(L)$ the set of its meet-irreducible elements.

Let G and M be arbitrary sets of *objects* and *attributes*, respectively, and let $I \subseteq G \times M$ be an arbitrary binary relation. Then, the triple $\mathbb{K} = (G, M, I)$ is called a *formal context*. For $g \in G$ and $m \in M$, we interpret $(g, m) \in I$ as "the object g has attribute m". For $A \subseteq G$ and $B \subseteq M$, a Galois connection between the powersets of G and M is defined through the following operators:

$$A' = \{m \in M \mid \forall g \in A : gIm\} \qquad B' = \{g \in G \mid \forall m \in B : gIm\}$$

Every pair (A, B) such that $A' = B$ and $B' = A$ is called a *formal concept.* A and B are the *extent* and the *intent* of the concept, respectively. Given a context $\mathbb{K}$, the set $\mathfrak{B}(\mathbb{K})$ of all formal concepts of $\mathbb{K}$ is partially ordered by $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$ (or, equivalently, $B_2 \subseteq B_1$). The *basic theorem on concept lattices* [20, Theorem 3] states that the set of formal concepts of the context $\mathbb{K}$ is a complete lattice $(\mathfrak{B}(\mathbb{K}), \sqcap, \sqcup)$, called *concept lattice*, where meet and join are defined by:

$$\begin{aligned} \prod_{j\in J}(A_j, B_j) &= \left(\bigcap_{j\in J} A_j, \left(\bigcup_{j\in J} B_j\right)''\right), \\ \bigsqcup_{j\in J}(A_j, B_j) &= \left(\left(\bigcup_{j\in J} A_j\right)'', \bigcap_{j\in J} B_j\right), \end{aligned} \tag{2}$$

for a set J of indexes. The following proposition is fundamental for our treatise.

Proposition 1 **(**[20, Proposition 12]**).** *For every finite lattice L there is (up to isomorphisms) a unique context $\mathbb{K}_L$, with $L \cong \mathfrak{B}(\mathbb{K}_L)$:*

$$\mathbb{K}_L := (\mathfrak{J}(L), \mathfrak{M}(L), \leq).$$

The context $\mathbb{K}_L$ is called the *standard context* of the lattice L.

Since L is finite, $\mathfrak{J}(L)$ is finite. Hence, the concept $(\mathfrak{J}(L), \emptyset)$ is the top element of $\mathfrak{B}(\mathbb{K}_L)$. We denote it $\top_G$, emphasizing the fact that the join-irreducible elements of L are the objects of our context. Analogously, the concept $(\emptyset, \mathfrak{M}(L))$ is the bottom element of $\mathfrak{B}(\mathbb{K}_L)$, and we denote it by $\bot_M$.

Example 1. Let $L = (\{a, b, c, d, e, f\}, \leq)$ be the finite distributive lattice in Fig. 1(a). Then, $\mathfrak{J}(L) = \{b, c, e\}$, and $\mathfrak{M}(L) = \{b, d, e\}$. Let $G = \{g_1, g_2, g_3\}$, and $M = \{m_1, m_2, m_3\}$. We relabel $\mathfrak{J}(L)$, and $\mathfrak{M}(L)$ via the labeling functions $\lambda_J : \mathfrak{J}(L) \to G$, and $\lambda_M : \mathfrak{M}(L) \to M$ such that $\lambda_J(b) = g_1$, $\lambda_J(c) = g_2$, $\lambda_J(e) = g_3$, $\lambda_M(b) = m_1$, $\lambda_M(d) = m_2$, and $\lambda_M(e) = m_3$. The following tables show the standard context $\mathbb{K}_L$, and its relabeling in terms of G and M:

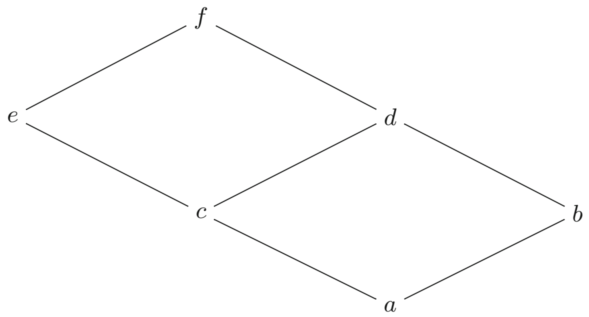
$\leq$	b	d	e
b	×	×	
c		×	×
e			×

$\leq$	m_1	m_2	m_3
g_1	×	×	
g_2		×	×
g_3			×

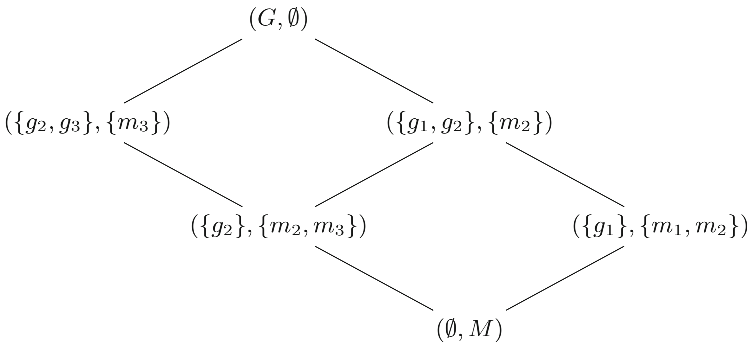
The concept lattice $\mathfrak{B}(\mathbb{K}_L)$ is depicted in Fig. 1(b).

3 Gödel Algebras of Concepts

Definition 1. *Let $\mathbb{K}$ be a finite context, and let $\mathfrak{B}(\mathbb{K})$ be its concept lattice. For every two concepts $C_1 = (G_1, M_1)$ and $C_2 = (G_2, M_2)$ in $\mathfrak{B}(\mathbb{K})$, we define the* p-implication $(\Rightarrow)$ *as:*



(a) L



(b) $\mathfrak{B}(\mathbb{K}_L)$

Fig. 1. A finite distributive lattice L, and its corresponding concept lattice $\mathfrak{B}(\mathbb{K}_L)$.

$$C_1 \Rightarrow C_2 = \bigsqcup \{(G_k, M_k) \in \mathfrak{B}(\mathbb{K}) \mid M_k \supseteq M_2 \backslash M_1\} . \qquad (\Rightarrow)$$

The following example better clarifies the previous definition.

Example 2. Consider the concept lattice depicted in Fig. 1(b). Then,

$$(\{g_1, g_2\}, \{m_2\}) \Rightarrow (\{g_2\}, \{m_2, m_3\}) = (\{g_2, g_3\}, \{m_3\}) ,$$
$$(\{g_2\}, \{m_2, m_3\}) \Rightarrow (\emptyset, M) = (\{g_1\}, \{m_1, m_2\}) .$$

The following proposition provides a way to build a concept lattice isomorphic to every Gödel algebra.

Proposition 2. *Let* $\mathbf{A} = (A, \wedge, \vee, \rightarrow, \top, \bot)$ *be a Gödel algebra, and let* $C_{\mathbf{A}} = \mathfrak{B}((\mathfrak{J}(\mathbf{A}), \mathfrak{M}(\mathbf{A}), \leq))$ *be the concept lattice of its standard context. Then, the algebra* $\mathbf{C_A} = (C_{\mathbf{A}}, \sqcap, \sqcup, \Rightarrow, \top_G, \bot_M)$*, where* $\Rightarrow$ *is the p-implication, is isomorphic to* $\mathbf{A}$.

Proof. Since each Gödel algebra is a finite lattice, it is isomorphic to the concept lattice of the associated standard context (c.f. Proposition 1). Let $f : A \rightarrow C_{\mathbf{A}}$

be such an isomorphism. We have to show that f extends to an isomorphism of Gödel algebras, that is

$$f(x \to y) = f(x) \Rightarrow f(y)\,, \tag{3}$$

for each $x, y \in A$. To this end, it suffices to prove the following claim.

Claim. The couple $(\sqcap, \Rightarrow)$ is a residuated pair.

We need to show that $(\sqcap, \Rightarrow)$ satisfies the residuum Eq. (1). That is

$$(C_1 \Rightarrow C_2) = \bigsqcup \{C_i \in C_{\mathbf{A}} \mid C_i \sqcap C_1 \leq C_2\}\,, \tag{4}$$

for every $C_1 = (G_1, M_1)$ and $C_2 = (G_2, M_2)$ in $\mathbf{C_A}$. We call $C_z = (G_z, M_z) = \bigsqcup \{C_i \in C_{\mathbf{A}} \mid C_i \sqcap C_1 \leq C_2\}$. By Definition 1, we have:

$$(C_1 \Rightarrow C_2) = \bigsqcup \{(G_i, M_i) \in C_{\mathbf{A}} \mid M_i \supseteq M_2 \backslash M_1\} = C_s = (G_s, M_s)\,. \tag{5}$$

We have to show that $M_s = M_z$ (equivalently, $G_s = G_z$). By (4), M_z is the smallest subset of M that belongs to a concept, and such that $M_z \cup M_1 \supseteq M_2$. In other words, M_z is precisely the smallest M_t such that $M_t \supseteq M_2 \backslash M_1$. Hence, by (5), M_z coincides with M_s. This settles the claim.

By the preceding claim, $\Rightarrow$ is precisely the unique (Gödel) residuum of $\sqcap$. Since the lattice isomorphisms f also preserves $\sqcap$, we have shown (3), and our statement is proved. □

We have derived the natural notion of implication between concepts in case the concept lattice is a Gödel algebra. Indeed, the p-implication satisfy the residuation law. It is now easy to provide a characterization of the Gödel negation of a concept.

Definition 2. *Let $\mathfrak{B}(\mathbb{K})$ be a concept lattice over a context $\mathbb{K}$, and let $(G_1, M_1) \in \mathfrak{B}(\mathbb{K})$. We call the* p-complement *of (G_1, M_1) the following operation*:

$$\sim (G_1, M_1) = \bigsqcup \{(G_k, M_k) \in \mathfrak{B}(\mathbb{K}) \mid M_k \supseteq M \backslash M_1\}\,.$$

Corollary 1. *The p-complement is the Gödel negation in a Gödel algebra of concepts.*

Proof. In Gödel logic the negation connective is derived from the implication: $\neg x := x \to \bot$. An easy computation shows that, if C is a concept of a Gödel algebra of concepts, then $\sim C = C \Rightarrow \bot$. □

Example 3. Consider the concept lattice depicted in Fig. 1(b). Then,

$$\sim (\{g_1, g_2\}, \{m_2\}) = (\emptyset, M)\,,$$
$$\sim (\{g_2\}, \{m_2, m_3\}) = (\{g_1\}, \{m_1, m_2\}).$$

Compare the second negation with Example 2.

4 Characterizing Gödel Algebras of Concepts

Let $\mathbb{K}$ be a finite context, and let $(\mathfrak{B}(\mathbb{K}), \sqcap, \sqcup)$ be its concept lattice. If, for each $C_1, C_2 \in \mathfrak{B}(\mathbb{K})$, there exists a greatest context $C \in \mathfrak{B}(\mathbb{K})$ such that $C_1 \sqcap C \leq C_2$, then $\mathfrak{B}(\mathbb{K})$ is a residuated lattice. The concept C is called the residuum, and it is denoted by $C_1 \Rightarrow C_2$. Since the residuum, if it exists, is unique, we have that $\Rightarrow$ must be exactly the p-implication defined in Definition 1. Indeed, in the proof of Proposition 2 it is shown that $(\sqcap, \Rightarrow)$ is a residuated pair. In general, a concept lattice need not be a distributive lattice. However, the existence of a residuum respect to the $\sqcap$ implies distributivity. Hence, in order to provide a characterization of Gödel algebras of concepts, we do not need to characterize distributivity. Nonetheless, the characterization of distributivity in concept lattices is an important topic in itself. An intrinsic characterization of distributivity in the finite case is provided in [26]. The infinite case has also been investigated, see [15].

The following proposition characterizes those concept lattices which are Gödel algebras.

Proposition 3. *Let $\mathbb{K}$ be a finite context, and let $(\mathfrak{B}(\mathbb{K}), \sqcap, \sqcup)$ be its concept lattice. Then,*

(i) $(\mathfrak{B}(\mathbb{K}), \sqcap, \sqcup, \Rightarrow, \top_G, \bot_M)$ is a Heyting algebra if and only if for each $C_1 = (G_1, M_1), C_2 = (G_2, M_2) \in \mathfrak{B}(\mathbb{K})$ there exists a greatest contest $C \in \mathfrak{B}(\mathbb{K})$ such that $C_1 \sqcap C \leq C_2$.

Moreover, let $C_l = (G_l, M_l) \in \mathfrak{B}(\mathbb{K})$ be such that M_l is the smallest set of attributes satisfying $M_l \supseteq M_2 \backslash M_1$. Analogously, let $C_r = (G_r, M_r) \in \mathfrak{B}(\mathbb{K})$ be such that M_r is the smallest set of attributes satisfying $M_r \supseteq M_1 \backslash M_2$.

(ii) The Heyting algebra $(\mathfrak{B}(\mathbb{K}), \sqcap, \sqcup, \Rightarrow, \top_G, \bot_M)$ is a Gödel algebra if and only if $M_l \cap M_r = \emptyset$.

Proof. The first part of the proposition is an immediate translation of the residuation property in terms of concepts. It has already been discussed in the beginning of the present section. We just need to prove (ii). Recall that Gödel algebras are Heyting algebras with a prelinear implication. We have to prove that the p-implication $\Rightarrow$ satisfies the prelinearity equation $(C_1 \Rightarrow C_2) \sqcup (C_2 \Rightarrow C_1) = \top_G$, for every $C_1, C_2 \in \mathfrak{B}(\mathbb{K})$, if, and only if, $M_l \cap M_r = \emptyset$.

Let

$$C_1 \Rightarrow C_2 = C_s = (G_s, M_s) = \bigsqcup \{(G_i, M_i) \in \mathfrak{B}(\mathbb{K}) \mid M_i \supseteq M_2 \backslash M_1\},$$

$$C_2 \Rightarrow C_1 = C_z = (G_z, M_z) = \bigsqcup \{(G_i, M_i) \in \mathfrak{B}(\mathbb{K}) \mid M_i \supseteq M_1 \backslash M_2\}.$$

Hence, prelinearity equation can be rewritten as:

$$C_s \sqcup C_z = (\mathfrak{I}(\mathfrak{B}(\mathbb{K})), \emptyset).$$

We observe that $M_l = M_s$, and $M_r = M_z$. Thus, $C_s \sqcup C_z = (\mathfrak{I}(\mathfrak{B}(\mathbb{K})), \emptyset)$ is equivalent to $M_l \cap M_r = \emptyset$, and (ii) is proved. □

5 Formal Concepts Described by Gödel Logic Sentences

In Sect. 3 we have associated formal concepts with elements of a finite Gödel algebra. Moreover, we have endowed the concept lattice with suitable operations, showing that every Gödel algebra is isomorphic to its associated concept lattice endowed with a p-implication. In this section, we advance some remarks on the logical counterpart of Gödel algebras, namely Gödel logic. Consider the free n-generated Gödel algebra $\mathcal{G}_n$. Since every finite Gödel algebra can be obtained as a quotient of a free n-generated Gödel algebra, we can effectively associate every Gödel logic formula with a corresponding concept. Knowing that $\mathcal{G}_n$ is a finite (distributive) lattice whose elements are formulæ in n variables (up to logical equivalence), and since for every finite lattice there is a unique reduced context $\mathbb{K}$, one can, indeed, relate (equivalence classes of) logical formulæ in $\mathcal{G}_n$ with the concepts in $\mathbb{K}$. That is precisely what we do in this section.

We start with a small example that can be dealt with via a trivial computation: the free 1-generated Gödel algebra $\mathcal{G}_1$. Comparing Figs. 1 and 2, one immediately notes that the lattice structure of $\mathcal{G}_1$ is isomorphic to $\mathfrak{B}(\mathbb{K}_L)$ in Fig. 1(b). Hence, by Proposition 1, there exists a lattice isomorphism $f : L(\mathcal{G}_1) \to \mathfrak{B}(\mathbb{K}_L)$ such that

$$f(\top) = (G, \emptyset)\,, \qquad f(\neg\neg x) = (\{g_2, g_3\}, \{m_3\})\,,$$
$$f(x \wedge \neg x) = (\{g_1, g_2\}, \{m_2\})\,, \qquad f(x) = (\{g_2\}, \{m_2, m_3\})\,,$$
$$f(\neg x) = (\{g_1\}, \{m_1, m_2\})\,, \qquad f(\bot) = (\emptyset, M)\,.$$

Moreover, by Proposition 2, $\mathfrak{B}(\mathbb{K}_L) = \mathbf{C}_{\mathcal{G}_1}$ and f is an isomorphism of algebras. Then,

$$\begin{aligned} &f([x \vee \neg x]_\equiv \to [x]_\equiv) = f([\neg\neg x]_\equiv) \\ &= (\{g_2, g_3\}, \{m_3\}) = (\{g_1, g_2\}, \{m_2\}) \Rightarrow (\{g_2\}, \{m_2, m_3\})\,, \end{aligned}$$

$$\begin{aligned} &f([x]_\equiv \to [\bot]_\equiv) = f([\neg x]_\equiv) \\ &= (\{g_1\}, \{m_1, m_2\}) = (\{g_2\}, \{m_2, m_3\}) \Rightarrow (\emptyset, M)\,. \end{aligned}$$

Compare with Example 2.

Let us consider a more complicated structure. Take the formula $\psi = \neg\neg x_1 \wedge \neg\neg x_2 \wedge (x_1 \vee x_2)$ over $\{x_1, x_2\}$, and let $\mathbf{A}$ be the Gödel algebra $\mathcal{G}_2/(\psi = \top)$ depicted in Fig. 3 (note that the equivalence classes displayed are the ones of $\mathcal{G}_2/(\psi = \top)$, not of $\mathcal{G}_2$).

Observe that $\mathfrak{J}(\mathbf{A}) = \{[x_1]_\equiv, [x_2]_\equiv, [x_1 \wedge x_2]_\equiv\}$, and $\mathfrak{M}(\mathbf{A}) = \{[x_1]_\equiv, [x_2]_\equiv\}$. Let $G = \{g_1, g_2, g_3\}$, and $M = \{m_1, m_2\}$, and define the labeling functions $\lambda_J : \mathfrak{J}(L) \to G$ and $\lambda_M : \mathfrak{M}(L) \to M$ by $\lambda_J([x_1 \wedge x_2]_\equiv) = g_1$, $\lambda_J([x_1]_\equiv) = g_2$, $\lambda_J([x_2]_\equiv) = g_3$, $\lambda_M([x_1]_\equiv) = m_1$, and $\lambda_M([x_1]_\equiv) = m_2$. The following two tables provide the standard context $C_{\mathbf{A}}$, and its relabeling in terms of G and M.

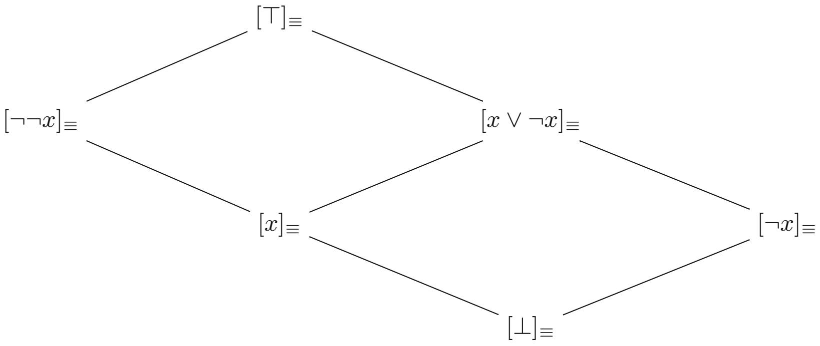


Fig. 2. The free 1-generated Gödel algebra $\mathcal{G}_1$.

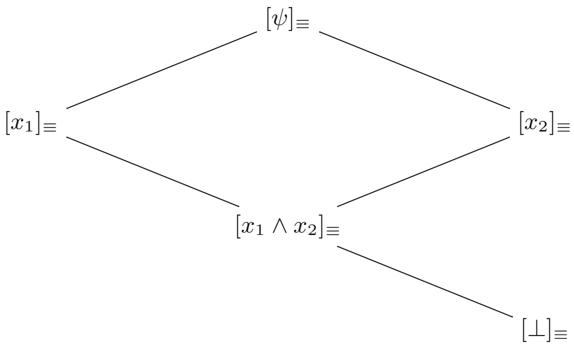


Fig. 3. A quotient of the free 2-generated Gödel algebra.

$\leq$	$[x_1]_\equiv$	$[x_2]_\equiv$
$[x_1 \wedge x_2]_\equiv$	×	×
$[x_1]_\equiv$	×	
$[x_2]_\equiv$		×

$\leq$	m_1	m_2
g_1	×	×
g_2	×	
g_3		×

Figure 4 shows the concept lattice associated with the Gödel algebra $\mathbf{A} = \mathcal{G}_2/(\psi = \top)$.

The characterization of free finitely generated Gödel algebras is a well-investigated topic that is beyond the scope of this paper. A functional representation is given in [21], while [1] is a state-of-the-art treatise on representations of many-valued logics. For our purposes it is sufficient to know that [2] contains a recursive description of $\mathcal{G}_n$, together with normal forms for Gödel logic, while in [14] the authors provide a combinatorial method to generate $\mathcal{G}_n$ and its quotients.

A general procedure to associate formal concepts with Gödel logic formulæ can be sketched out, based on the preceding examples. Let $\varphi_1, \ldots, \varphi_m, \psi$ be Gödel logic formulæ over $\{x_1, \ldots, x_n\}$, with $m \geq 0$, and $n \geq 1$. Generate $\mathcal{G}_n$ (see [2,14]) and apply Proposition 1, obtaining $C_{\mathcal{G}_n}$. Then, $\{\varphi_1, \ldots, \varphi_m\} \vdash \psi$

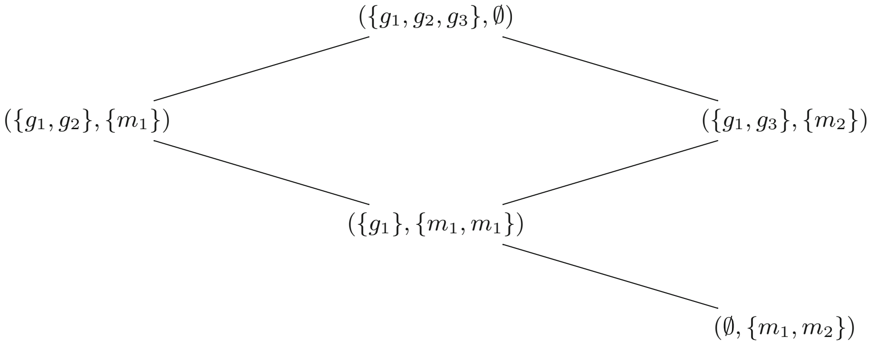


Fig. 4. The concept lattice associated with the Gödel algebra $\mathcal{G}_2/(\psi = \top)$

amounts to evaluating ψ over $\mathcal{G}_n/(\varphi_1 = \top, \ldots, \varphi_m = \top)$. Proposition 2 states that $\mathbf{C}_{\mathcal{G}_n}$ is isomorphic to $\mathcal{G}_n$. Hence, such evaluation provides also a concept in $\mathbf{C}_{\mathcal{G}_n}$, that is, precisely the concept associated with ψ. This allows us to express formal concepts associated with ψ, for every theory $\{\varphi_1, \ldots, \varphi_m\}$ in Gödel logic.

6 Concluding Remarks

In the basic setting of FCA (see Sect. 2) it is assumed that concepts are crisp. In the literature one can find several studies whose aim is the "fuzzification" of I, the relation between G and M. The first one being [10], while [7,8] are good overview of these investigations. A further generalization of this type of approach is given in [6], where the author considers both relation and order in FCA as defined over fuzzy sets (or residuated lattices in general). Our method diverges from those approaches. We exploit the classical notions of FCA to obtain new insight on algebraic semantics of many-valued logics. Indeed, in the above sections we have shown that it is possible to associate a formal concept with every formula of Gödel logic. Further, we have provided a characterization of concept lattices isomorphic to Gödel algebras in terms of formal contexts. In this way we could effectively find contexts over which Gödel logic can be used to reason about.

In other words, whenever a concept lattice satisfies Proposition 3, we are dealing with a Gödel algebra of concepts. Under such conditions, concepts can be combined via the lattice operators meet and join – see (2) –, but also via the operations of p-implication and p-complement introduced in Sect. 3. The latter operations correspond, respectively, with the Gödel logic implication and negation, as shown in Proposition 2 and Corollary 1. In this sense we can say that our new interpretation can be viewed as an alternative semantics for Gödel logic. In order to acquire a full understanding of this semantics, we aim to investigate, in future work, the effect of the p-implication and p-complement over concepts obtained from contexts describing real-world scenarios. The ultimate goal is to get more insight about the meaning of Gödel logic by running empirical experiments over real data. Through this work we believe that this can be done.

The approach used in this work is not limited to Gödel logic, but it can be generally applied to many non-classical logics. Broadly speaking, it is sufficient that the corresponding algebraic semantics has a complete lattice reduct. As a many-valued logic, Gödel logic is a schematic extension of the fundamental system BL introduced by Hájek in [22], which in turn is a schematic extension of the *Monoidal T-norm Logic* (MTL) [16]. Hence, we believe that extending our method to other logics in this hierarchy could be an interesting task. The first issue to deal with is the fact that these logics have a monoidal conjunction in addition to the lattice one. A good starting point would be investigate logics where representations of free algebras are already available, *e.g.*, Nilpotent Minimum logic [3,11], or Revised Drastic Product logic [27]. Further, many-valued logics are just particular substructural logics whose algebraic semantics is provided by the class of residuated lattices [19], giving thus space for further generalizations.

Additional research has to be done to compare our method with other investigations regarding alternative semantics and intended meaning of many-valued logics. For the former we can cite probabilistic [3,4], temporal [9] and game-theoretic [17] approaches, and [12,13,24,25] for the latter.

Acknowledgements. We thank Matteo Bianchi for useful discussions about the subject of the paper. We acknowledge the support of our Marie Curie INdAM-COFUND fellowships.

References

1. Aguzzoli, S., Bova, S., Gerla, B.: Free algebras and functional representation. In: Cintula, P., Hájek, P., Noguera, C. (eds.) Handbook of Mathematical Fuzzy Logic, Studies in Logic, vol. 38, chap. IX, pp. 713–791. College Publications, London (2011)
2. Aguzzoli, S., Gerla, B.: Normal forms and free algebras for some extensions of MTL. Fuzzy Sets Syst. **159**(10), 1131–1152 (2008)
3. Aguzzoli, S., Gerla, B.: Probability measures in the logic of nilpotent minimum. Stud. Logica **94**(2), 151–176 (2010)
4. Aguzzoli, S., Gerla, B., Marra, V.: De finetti's no-dutch-book criterion for Gödel logic. Stud. Logica **90**(1), 25–41 (2008)
5. Aguzzoli, S., Gerla, B., Marra, V.: Embedding Gödel propositional logic into Prior's tense logic. In: Magdalena, L., Ojeda-Aciego, M., Verdegay, J. (eds.) Proceedings of IPMU 2008, pp. 992–999. Torremolinos (Málaga), June 2008
6. Belohlávek, R.: Concept lattices and order in fuzzy logic. Ann. Pure Appl. Logic **128**(1–3), 277–298 (2004)
7. Belohlavek, R.: What is a fuzzy concept lattice? II. In: Kuznetsov, S.O., Ślzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS (LNAI), vol. 6743, pp. 19–26. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21881-1_4
8. Belohlávek, R., Vychodil, V.: What is a fuzzy concept lattice?. In: Belohlávek, R., Snášel, V. (eds.) Proceedings of CLA 2005, vol. 162, pp. 34–45. CEUR WS (2005)

9. Bianchi, M.: A temporal semantics for nilpotent minimum logic. Int. J. Approximate Reasoning **55**(1), 391–401 (2014)
10. Burusco Juandeaburre, A., Fuentes-González, R.: The study of the L-fuzzy concept lattice. Mathware Soft Comput. **1**(3), 209–218 (1994)
11. Busaniche, M.: Free nilpotent minimum algebras. Math. Logic Q. **52**(3), 219–236 (2006)
12. Codara, P., D'Antona, O., Marra, V.: An analysis of Ruspini partitions in Gödel logic. Int. J. Approximate Reasoning **50**(6), 825–836 (2009)
13. Codara, P., D'Antona, O., Marra, V.: Valuations in Gödel logic, and the euler characteristic. J. Multiple-Valued Logic Soft Comput. **19**(1–3), 71–84 (2012)
14. D'Antona, O., Marra, V.: Computing coproducts of finitely presented Gödel algebras. Ann. Pure Appl. Logic **142**(1–3), 202–211 (2006)
15. Erné, M.: Distributive laws for concept lattices. Algebra Univers. **30**(4), 538–580 (1993). http://dx.doi.org/10.1007/BF01195382
16. Esteva, F., Godo, L.: Monoidal t-norm based logic: towards a logic for left-continuous t-norms. Fuzzy Sets Syst. **124**(3), 271–288 (2001)
17. Fermüller, C.: Dialogue games for many-valued logics - an overview. Stud. Logica **90**(1), 43–68 (2008)
18. Fermüller, C.: On Giles style dialogues games and hypersequent systems. In: Hosni, H., Montagna, F. (eds.) Probability, Uncertainty and Rationality. Centro di Ricerca Matematica Ennio De Giorgi Series (No. 7), vol. 10, pp. 169–197. Edizioni della Normale, Pisa (2010)
19. Galatos, N., Jipsen, P., Kowalski, T., Ono, H.: Residuated Lattices: An Algebraic Glimpse at Substructural Logics. Elsevier, Amsterdam (2007)
20. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations, 1st edn. Springer-Verlag New York Inc., Secaucus (1997)
21. Gerla, B.: A note on functions associated with Gödel formulas. Soft Comput. **4**(4), 206–209 (2000)
22. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer, Dordrecht (1998)
23. Horn, A.: Free L-algebras. J. Symbol. Logic **34**, 475–480 (1969)
24. Marra, V.: Is there a probability theory of many-valued events? In: Hosni, H., Montagna, F. (eds.) Probability, Uncertainty and Rationality. Centro di Ricerca Matematica Ennio De Giorgi Series (No. 7), vol. 10, pp. 141–166. Edizioni della Normale, Pisa (2010)
25. Marra, V.: The problem of artificial precision in theories of vagueness: a note on the rôle of maximal consistency. Erkenntnis **79**(5), 1015–1026 (2013)
26. Monjardet, B., Wille, R.: On finite lattices generated by their doubly irreducible elements. Discrete Math. **73**(1–2), 163–164 (1989). https://dx.doi.org/10.1016/0012-365X(88)90144-6
27. Valota, D.: Poset representation for free RDP-Algebras. In: Hosni, H., Montagna, F. (eds.) Probability, Uncertainty and Rationality. Centro di Ricerca Matematica Ennio De Giorgi Series (No. 7), vol. 10. Edizioni della Normale, Pisa (2010)

A Representation Theorem for Stratified Complete Lattices

Zoltán Ésik(✉)

Department of Computer Science, University of Szeged, Szeged, Hungary
esikr@yahoo.com

Abstract. Stratified complete lattices are complete lattices equipped with a sequence of preorderings associated with the ordinals less than a given nonzero ordinal, typically a limit ordinal. They have been used to give semantics to recursive definitions involving nonmonotonic operations. We provide representation theorems for stratified complete lattices by inverse limits of complete lattices.

1 Introduction

A novel approach to the semantics of logic programs with negation, using an infinite supply of truth values, was introduced in [10]. The development of a fixed point theory underlying this approach has recently been undertaken in [4,6,7]. This fixed point theory has been applied to higher-order logic programs with negation [1] and to Boolean context-free languages [7].

The structures studied in this novel fixed point theory are stratified complete lattices, i.e., complete lattices $(L, \leq)$, equipped with a family of preorderings $\sqsubseteq_\alpha$, indexed by the ordinals α strictly less than a fixed nonzero ordinal κ, which without loss of generality can be taken to be a limit ordinal. In [6,7], several systems of axioms have been introduced. Some of the results, such as the 'Lattice Theorem' or the 'Fixed Point Theorem' of [6], were proved for a weaker class of stratified complete lattices, whereas some others, such as the 'Model Intersection Theorem' of [7], were established for stronger classes of stratified complete lattices. The Lattice Theorem asserts that every stratified complete lattice satisfying the axioms can be equipped with another complete lattice ordering $\sqsubseteq$ by defining $x \sqsubseteq y$ iff either $x = y$, or there is some $\alpha < \kappa$ with $x \sqsubset_\alpha y$ (i.e., $x \sqsubseteq_\alpha y$ but $y \not\sqsubseteq_\alpha x$). The Fixed Point Theorem states that certain nonmonotone functions $L \to L$ over an appropriate stratified complete lattice L have least fixed points w.r.t. the ordering $\sqsubseteq$.

In this paper, we mainly deal with two systems of axioms introduced in [6,7] that seem to be the most relevant for applications. In the stratified complete lattices satisfying these systems of axioms, called models and strong models, resp., the preorderings $\sqsubseteq_\alpha$, $\alpha < \kappa$, are completely determined by the complete

Partially supported by grant no. ANN 110883 from the National Foundation of Hungary for Scientific Research.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 263–289, 2017.
DOI: 10.1007/978-3-662-54332-0_15

lattice order $\leq$ and the equivalence relations $=_\alpha$ corresponding to the preorderings $\sqsubseteq_\alpha$.

The main results of the paper are:

1. Every model L is isomorphic to the stratified complete lattice determined by an inverse limit of complete lattices with locally completely additive projections, cf. Theorem 2.
2. Every strong model L is isomorphic to the stratified complete lattice determined by an inverse limit of complete lattices with completely additive projections, cf. Corollary 15.
3. A general result (Theorem 3) based on the above representation theorems implying the Lattice Theorem, the Fixed Point Theorem, and the fact that for every model L and weakly monotone function $f : L \to L$ w.r.t. $\sqsubseteq$, the fixed points of L form a complete lattice w.r.t. the ordering $\sqsubseteq$ (cf. Corollary 19).

2 Models and Examples

In this section, we introduce axioms for the structures we are going to discuss throughout the paper. We will also provide some examples and a construction. For unexplained notions regarding lattices we refer to [2].

Suppose that κ is a fixed limit ordinal. We will be considering structures of the sort $L = (L, \leq, (\sqsubseteq_\alpha)_{\alpha<\kappa})$, called *stratified complete lattices*, such that $(L, \leq)$ is a complete lattice with bottom and top elements $\bot$ and $\top$, resp., and for each $\alpha < \kappa$, $\sqsubseteq_\alpha$ is a preordering of L.

Our stratified complete lattices will satisfy the following axioms, where for each α, $=_\alpha$ denotes the equivalence relation determined by $\sqsubseteq_\alpha$.

A1. For all $\alpha < \beta < \kappa$, $\sqsubseteq_\beta$ is included in $=_\alpha$, so that if $x \sqsubseteq_\beta y$ then $x =_\alpha y$.
A2. The intersection of all the relations $=_\alpha$ for $\alpha < \kappa$ is the identity relation, so that if $x =_\alpha y$ for all $\alpha < \kappa$, then $x = y$.
A3. For all x and $\alpha < \kappa$ there exists y such that $x =_\alpha y$ and for all z, if $x \sqsubseteq_\alpha z$ then $y \leq z$.

It follows from the first two axioms that the intersection of all relations $\sqsubseteq_\alpha$, $\alpha < \kappa$, is also the identity relation. It is clear that the element y in A3 is uniquely determined by x and α and also satisfies $y \sqsubseteq_\alpha z$ whenever $x \sqsubseteq_\alpha z$. We will denote it by $x|_\alpha$.

A4. For all α with $\alpha < \kappa$ and x_i and y with $x_i =_\alpha y$, $i \in I$, where I is any nonempty index set, it holds that $\bigvee_{i\in I} x_i =_\alpha y$.
A5. For all x, y and $\alpha < \kappa$, if $x \leq y$ then $x|_\alpha \leq y|_\alpha$.
A6. For all x, y and $\alpha < \kappa$, if $x \leq y$ and $x =_\beta y$ for all $\beta < \alpha$, then $x \sqsubseteq_\alpha y$.

A stratified complete lattice satisfying the above axioms A1–A6 will be called a *model*, for short. Sometimes we will require a stronger variant of A4.

A4*. For all α with $\alpha < \kappa$ and x_i, y_i with $x_i =_\alpha y_i$, $i \in I$, where I is any nonempty index set, it holds that $\bigvee_{i\in I} x_i =_\alpha \bigvee_{i\in I} y_i$.

Models satisfying A4* will be called *strong*. We will discuss several consequences of the axioms in Sect. 5.

The following motivating example is from [6,10]. Consider the following linearly ordered set $V = V_\kappa$ of truth values:

$$F_0 < F1 < \cdots < F_\alpha < \cdots < 0 < \cdots < T_\alpha < \cdots < T_1 < T_0,$$

where α ranges over the ordinals strictly less than κ. Let Z denote a nonempty set of propositional variables and consider the set $L = V^Z$, equipped with the pointwise ordering. Thus, for all $f, g \in L$, $f \leq g$ iff $f(z) \leq g(z)$ for all $z \in Z$. Then $(L, \leq)$ is a complete lattice. For each $f, g \in L$ and $\alpha < \kappa$, define $f \sqsubseteq_\alpha g$ iff for all $z \in Z$,

(i) $\forall \beta < \alpha \ (f(z) = F_\beta \Leftrightarrow g(z) = F_\beta) \wedge (f(z) = T_\beta \Leftrightarrow g(z) = T_\beta)$, and
(ii) $(g(z) = F_\alpha \Rightarrow f(z) = F_\alpha) \wedge (f(z) = T_\alpha \Rightarrow g(z) = T_\alpha)$.

Then L is a strong model. When $f \in L$ and $\alpha < \kappa$, then for all $z \in Z$, $f|_\alpha(z) = f(z)$ if $f(z)$ is in the set $\{F_\beta, T_\beta : \beta < \alpha\}$, and $f|_\alpha(z) = F_{\alpha+1}$, otherwise. For κ being the least uncountable ordinal, this example was used in [10] to give semantics to possibly countably infinite propositional logic programs involving negation. The idea is to associate with a logic program P over Z a function $f_P : V_\Omega^Z \to V_\Omega^Z$, and to define the semantics of P as the unique least fixed point of f_P with respect to a new ordering $\sqsubseteq$, canonically defined for interpretations $I, J \in V_\Omega^Z$ by $I \sqsubseteq J$ iff $I = J$ or there is some $\alpha < \Omega$ with $f_P(I) \sqsubset_\alpha f_P(J)$ (i.e., $f_P(I) \sqsubseteq_\alpha f_P(J)$ but $f_P(J) \not\sqsubseteq_\alpha f_P(I)$). The function f_P is not necessarily monotone with respect to $\sqsubseteq$. It is argued in [10] that the semantics corresponds to the view of negation as failure. See Example 3 for more details. For an extension to higher order logic programs, see [1].

In particular, Z can be chosen to be a singleton set. It follows that V_κ is itself a strong model with the relations $\sqsubseteq_\alpha$, $\alpha < \kappa$, defined by $x \sqsubseteq_\alpha y$ iff $x = y$ or $x, y \in \{F_\gamma, T_\gamma : \gamma \geq \alpha\} \cup \{0\}$ such that if $x = T_\alpha$ then $y = T_\alpha$ and if $y = F_\alpha$ then $x = F_\alpha$.

The axioms A1–A6 are from [6,7]. Actually A3 is a weaker version of the corresponding axiom in [6] that we will denote A3*. (Axiom A3* will be recalled and established in all models in Proposition 5.)

3 Inverse Limits

In this section, we recall the notion of inverse systems and limits of inverse systems of complete lattices. Inverse limits will be used to construct further models of the axioms. We will make use of the following concept.

Suppose that $L = (L, \leq)$ and $L' = (L', \leq)$ are complete lattices. We say that $h : L' \to L$ *preserves all infima* if $h(\bigwedge Y) = \bigwedge h(Y)$ for all $Y \subseteq L$. Similarly, we say that $k : L \to L'$ *preserves all suprema*, or that k is *completely additive*, if $k(\bigvee X) = \bigvee k(X)$ for all $X \subseteq L$. It is clear that if $h : L' \to L$ preserves all infima, then it is monotone and preserves the greatest element. If h is additionally surjective, then it preserves the least element. Similar facts hold for functions preserving all suprema.

Suppose that L and L' are complete lattices and $h : L' \to L$ and $k : L \to L'$ are monotone functions. We say that (h, k) is a *(monotone) Galois connection* [2] (with h being the upper and k being the lower adjoint) if the identity function on L is less than or equal to $h \circ k : L \to L$ and $k \circ h : L' \to L'$ is less than or equal to the identity function on L' with respect to the pointwise ordering of functions. It is known, cf. [2], that for complete lattices L and L' and functions $h : L' \to L$ and $k : L \to L'$, (h, k) is a Galois connection iff h preserves all infima and k preserves all suprema. Moreover, we say that (h, k) is a *projection-embedding pair* [11] if $h \circ k : L \to L$ is the identity function on L and $k \circ h : L' \to L'$ is less than or equal to the identity function on L' with respect to the pointwise ordering of functions. Thus, a projection-embedding pair is a Galois connection.

Suppose that (h, k) is a Galois connection between complete lattices L, L' as above. If (h, k) is a projection embedding pair, then h is clearly surjective and k is injective. Conversely, if h is surjective or k is injective, then (h, k) is a projection-embedding pair (also called a Galois insertion). It is also clear that h uniquely determines k and vice versa. Indeed, for each $x \in L$, $k(x)$ is the least element y of L' with $x \leq h(y)$. And for each $y \in L'$, $h(y)$ is the greatest $x \in L$ with $k(x) \leq y$.

We call a monotone function $h : L' \to L$ a *projection* if there is a corresponding embedding $L \to L'$ (which is then uniquely determined), and call a monotone function $k : L \to L'$ an *embedding* if there is a corresponding projection $L' \to L$. A well-known useful fact is that any composition of projections is a projection and corresponds to the composition of the respective embeddings.

Suppose that for each $\alpha < \kappa$, $L_\alpha = (L_\alpha, \leq)$ is a complete lattice and a family of projections $h^\alpha_\beta : L_\alpha \to L_\beta$ for $\beta < \alpha < \kappa$ is specified such that $h^\beta_\gamma \circ h^\alpha_\beta = h^\alpha_\gamma$, for all $\gamma < \beta < \alpha$. Then we say that the complete lattices L_α, $\alpha < \kappa$, form an *inverse system*, c.f. [11],[1] with projections h^α_β, $\beta < \alpha < \kappa$.

For the rest of this section, suppose that we are given such an inverse system of complete lattices. We denote the embedding corresponding to each h^α_β by k^α_β. As noted above, it follows that $k^\alpha_\beta \circ k^\beta_\gamma = k^\alpha_\gamma$, for all $\gamma < \beta < \alpha < \kappa$. Also, for each $\beta < \alpha < \kappa$, h^α_β preserves all infima and k^α_β preserves all suprema. We will sometimes also suppose that the projections h^α_β are completely additive, or at least locally completely additive, see below. It will be convenient to define h^α_α and k^α_α for $\alpha < \kappa$ as the identity function $L_\alpha \to L_\alpha$.

Let L_∞ be the *inverse limit* determined by the above inverse system. Thus, $L_\infty \subseteq \prod_{\alpha<\kappa} L_\alpha$ is the collection of all κ-sequences $x = (x_\alpha)_{\alpha<\kappa}$ in $\prod_{\alpha<\kappa} L_\alpha$ with $h^\alpha_\beta(x_\alpha) = x_\beta$ for all $\beta < \alpha < \kappa$, ordered by the relation $\leq$ defined pointwise. A sequence in L_∞ will be referred to as a 'compatible sequence'. Since the functions h^α_β preserve all infima, L_∞ is indeed a complete lattice in which the

[1] The complete lattices and projections of an inverse system of [11] are continuous, and the ordinal κ is ω, the least infinite ordinal. Inverse systems of complete lattices over arbitrary directed partial orders are considered in [9], where following [11], the projections are usually assumed to be continuous as well.

infimum $\bigwedge X$ of any set $X \subseteq L_\infty$ is formed pointwise. This follows by noting that the pointwise infimum of any set of compatible sequences is compatible, since the functions h^α_β preserve all infima. The least element of L_∞ is the compatible sequence $(\bot_\alpha)_{\alpha<\kappa}$ composed of the least elements of the lattices L_α. The greatest element is the sequence $(\top_\alpha)_{\alpha<\kappa}$, where for each $\alpha < \kappa$, $\top_\alpha$ is the greatest element of L_α. If the functions h^α_β, $\beta < \alpha < \kappa$, are all completely additive, then the supremum $\bigvee X$ of any set X of sequences in L_∞ is also formed pointwise. To facilitate notation, we will denote the supremum and the infimum of a subset X of L_α by $\bigvee_\alpha X$ and $\bigwedge_\alpha X$, respectively.

For each $\alpha < \kappa$, let h^∞_α denote the function $L_\infty \to L_\alpha$ mapping each $x \in L_\infty$ to the α-component x_α of x. These functions form a *cone* over the inverse system $h^\alpha_\beta : L_\alpha \to L_\beta$, since $h^\alpha_\beta \circ h^\infty_\alpha = h^\infty_\beta$ for all $\beta < \alpha < \kappa$.

Lemma 1. *Suppose that the complete lattices L_α, $\alpha < \kappa$, form an inverse system with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$, and limit L_∞. Then each function $h^\infty_\alpha : L_\infty \to L_\alpha$ for $\alpha < \kappa$ is also a projection.*

Proof. For each $x \in L_\alpha$, where $\alpha < \kappa$, let $k^\infty_\alpha(x) = (y_\beta)_{\beta<\kappa}$ with $y_\beta = h^\alpha_\beta(x)$ if $\beta \le \alpha$, and $y_\beta = k^\beta_\alpha(x)$ if $\beta > \alpha$, where k^β_α is the embedding corresponding to h^β_α. Then $k^\infty_\alpha(x) \in L_\infty$ and clearly $h^\infty_\alpha(k^\infty_\alpha(x)) = x$. And if $z = (z_\beta)_{\beta<\kappa}$ is in L_∞, then $k^\infty_\alpha(h^\infty_\alpha(z)) \le z$, since if $\beta \le \alpha$ then the β-component of $k^\infty_\alpha(h^\infty_\alpha(z))$ is z_β, and if $\beta > \alpha$, then the β-component of $k^\infty_\alpha(h^\infty_\alpha(z))$ is $k^\beta_\alpha(z_\alpha) \le z_\beta$, since $z_\alpha = h^\beta_\alpha(z_\beta)$ and $(h^\alpha_\beta, k^\alpha_\beta)$ is a projection-embedding pair. Thus, $h^\infty_\alpha : L_\infty \to L_\alpha$ is a projection with corresponding embedding $k^\infty_\alpha : L_\alpha \to L_\infty$. □

It follows that the functions h^∞_α preserve all infima and the functions k^∞_α preserve all suprema.

The complete lattice L_∞ has the following property. Suppose that L is a complete lattice and the functions $g_\alpha : L \to L_\alpha$ form another cone, where $\alpha < \kappa$, so that $g_\beta = h^\alpha_\beta \circ g_\alpha$ for all $\beta < \alpha < \kappa$. Then there is a unique function $g : L \to L_\infty$ such that $h^\infty_\alpha \circ g = g_\alpha$, for all $\alpha < \kappa$. Indeed, for each $y \in L$, $g(y) = (g_\alpha(y))_{\alpha<\kappa}$. If the functions g_α, $\alpha < \kappa$, are monotone, then so is this *mediating* function g, and vice versa. We will call the functions h^∞_α, $\alpha < \kappa$, *limit functions*, or *limit projections.*

Lemma 2. *Suppose that the complete lattices L_α, $\alpha < \kappa$, form an inverse system with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$, and limit L_∞. Let L be a complete lattice with a cone of projections $g_\alpha : L \to L_\alpha$ and corresponding embeddings $f_\alpha : L_\alpha \to L$, for each $\alpha < \kappa$, and let g denote the mediating function $L \to L_\infty$, $y \mapsto (g_\alpha(y))_{\alpha<\kappa}$. Define $f : L_\infty \to L$ by $f(x) = \bigwedge\{y : y \in L,\ \forall\gamma < \kappa\ x_\gamma \le g_\gamma(y)\} = \bigwedge\{y : y \in L,\ x \le g(y)\}$ for all $x = (x_\gamma)_{\gamma<\kappa} \in L_\infty$. Then the pair of functions g and f forms a Galois connection between L_∞ and L.*

Proof. Indeed, we have already noted that g is monotone, and it is clear that f is also monotone. Let $x = (x_\gamma)_{\gamma<\kappa} \in L_\infty$. Then for all $\alpha < \kappa$, $g_\alpha(f(x)) = g_\alpha(\bigwedge\{y : y \in L,\ \forall\gamma < \kappa\ x_\gamma \le g_\gamma(y)\}) = \bigwedge_\alpha\{g_\alpha(y) : y \in L,\ \forall\gamma < \kappa\ x_\gamma \le g_\gamma(y)\}$, since g_α preserves arbitrary infima. It is clear that $x_\alpha \leq \bigwedge_\alpha \{g_\alpha(y) \in L : \forall \gamma < \kappa\ x_\gamma \leq g_\gamma(y)\}$, thus $x_\alpha \leq g_\alpha(f(x))$. Since this holds for all $\alpha < \kappa$, it follows that the identity function over L_∞ is less than or equal to $g \circ f$ with respect to the pointwise ordering. We still need to prove that $f \circ g$ is less than or equal to the identity function over L. But for all $y \in L$, $f(g(y)) = \bigwedge\{z : z \in L,\ g(y) \leq g(z)\} \leq y$, since $g(y) \leq g(y)$. □

Remark 1. For later use we note that if the mediating function g of Lemma 2 is surjective, or if for each $x = (x_\gamma)_{\gamma<\kappa}$ in L_∞ and $\alpha < \kappa$ there is some $y \in L$ with $x_\alpha = g_\alpha(y)$ and $x_\gamma \leq g_\gamma(y)$ for all $\gamma < \kappa$, then g is a projection. Indeed, if either of these assumptions applies, then $x_\alpha = g_\alpha(f(x))$ for all $\alpha < \kappa$ and $x \in L$, where f is defined as in Lemma 2.

If the projections h^α_β, $\beta < \alpha < \kappa$, satisfy a weak form of complete additivity, then we can prove that the mediating morphism g is in fact a projection. Call a monotone function $L' \to L$ *locally completely additive* if for all $Y \subseteq L'$ and $x \in L$ with $h(Y) = \{x\}$ (i.e., Y is nonempty and h maps each element of Y to x), it holds that $h(\bigvee Y) = x$. It is clear that when a function $h : L' \to L$ is completely additive, then it is locally completely additive.

There exist finite complete lattices L and L' and a projection $L' \to L$ which is not locally completely additive, cf. [5]. An example of finite lattices L and L' with a locally completely additive projection $L' \to L$ which is not completely additive is given in [5].

Lemma 3. *Let L and L' be complete lattices and let $g : L' \to L$ be monotone and surjective. Then g is locally completely additive iff $\bigvee g^{-1}(x) \in g^{-1}(x)$ for all $x \in L$.*

Proof. Suppose first that g is locally completely additive. Let $x \in L$ and $Y = g^{-1}(x)$. Then $g(Y) = \{x\}$, thus $g(\bigvee Y) = x$ and $\bigvee g^{-1}(x) = \bigvee Y \in g^{-1}(x)$, since g is locally completely additive.

Suppose now that $\bigvee g^{-1}(x) \in g^{-1}(x)$ for all $x \in L$. Let $x \in L$ and $Y \subseteq L'$ with $g(Y) = \{x\}$. Then Y is not empty, say $y_0 \in Y$. Since $y_0 \leq \bigvee Y \leq \bigvee g^{-1}(x)$ and g is monotone, it holds that $x = g(y_0) \leq g(\bigvee Y) \leq g(\bigvee g^{-1}(x)) = x$. Thus, $g(\bigvee Y) = x$. □

Lemma 4. *Let L_∞ be the limit of the inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Then the limit projections $h^\infty_\beta : L_\infty \to L_\beta$, $\beta < \kappa$, are also locally completely additive.*

Proof. Suppose that $x \in L_\beta$ and $Y = (h^\infty_\beta)^{-1}(x)$, where $\beta < \kappa$ is a fixed ordinal. We need to prove that $h^\infty_\beta(\bigvee Y) = x$.

For each α with $\beta < \alpha < \kappa$, let $Y_\alpha = (h^\alpha_\beta)^{-1}(x)$. If $\beta < \alpha < \alpha' < \kappa$, then $(h^{\alpha'}_\beta)^{-1}(x) = (h^{\alpha'}_\alpha)^{-1}((h^\alpha_\beta)^{-1}(x))$, hence $Y_{\alpha'} = (h^{\alpha'}_\alpha)^{-1}(Y_\alpha)$. Moreover, $h^{\alpha'}_\alpha(Y_{\alpha'}) = Y_\alpha$. Also, $Y = (h^\infty_\alpha)^{-1}(Y_\alpha)$ and $h^\infty_\alpha(Y) = Y_\alpha$ for all α with $\beta < \alpha < \kappa$.

For each α with $\beta < \alpha < \kappa$, define $y_\alpha = \bigvee Y_\alpha$. When $\alpha \leq \beta$, let $y_\alpha = h^\beta_\alpha(x)$. We intend to show that the sequence $(y_\alpha)_{\alpha<\kappa}$ is compatible, so that $y = (y_\alpha)_{\alpha<\kappa}$ is in L_∞.

We have $y_\alpha \in Y_\alpha$ for all α with $\beta < \alpha < \kappa$, since h^α_β is locally completely additive. Thus, if $\beta < \alpha < \alpha'$, then $h^{\alpha'}_\alpha(y_{\alpha'}) = y_\alpha$, since $h^{\alpha'}_\alpha(y_{\alpha'})$ is necessarily the greatest element of Y_α. When $\alpha < \alpha' < \kappa$ with $\alpha \leq \beta$, then $h^{\alpha'}_\alpha(y_{\alpha'}) = h^\beta_\alpha(x_\beta) = y_\alpha$. Thus, $y \in L_\infty$.

We claim that $y = \bigvee Y$ in L_∞. We have already shown that $y \in L_\infty$. We know that for each α with $\beta < \alpha < \kappa$, it holds that $y_\alpha = \bigvee Y_\alpha$. Thus, our claim holds if for all such α, Y_α is equal to the set of all α-components of the sequences in Y. But this is clear, since $Y_\alpha = h^\infty_\alpha(Y)$.

It follows now that h^∞_β is locally completely additive. □

Lemma 5. *Let L_∞ be the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that L is a complete lattice and the locally completely additive projections $g_\alpha : L \to L_\alpha$, $\alpha < \kappa$, form a cone. Then the unique mediating function $g : L \to L_\infty$ is a projection.*

Proof. We already know that g is a projection if it is surjective, cf. Lemma 2 and Remark 1. Below we prove that g is indeed surjective. We will also give a new description of the corresponding embedding.

For each $\alpha < \kappa$, let f_α denote the embedding corresponding to g_α. When $x = (x_\alpha)_{\alpha<\kappa}$ is in L_∞, define $f(x) = \bigvee_{\alpha<\kappa} f_\alpha(x_\alpha)$. We prove that $g(f(x)) = x$ for all $x \in L_\infty$ and that f is the embedding corresponding to g.

So let $x = (x_\alpha)_{\alpha<\kappa}$ in L_∞. If $\beta < \alpha < \kappa$, then $f_\beta(x_\beta) = \bigwedge\{y : y \in L,\ x_\beta \leq g_\beta(y)\} \leq \bigwedge\{y : y \in L,\ x_\alpha \leq g_\alpha(y)\} = f_\alpha(x_\alpha)$, since if $x_\alpha \leq g_\alpha(y)$ for some $y \in L$, then $x_\beta = h^\alpha_\beta(x_\alpha) \leq h^\alpha_\beta(g_\alpha(y)) = g_\beta(y)$. Hence the sequence $(f_\alpha(x_\alpha))_{\alpha<\kappa}$ is increasing. If $\gamma \leq \alpha < \kappa$, then $g_\gamma(f_\alpha(x_\alpha)) = h^\alpha_\gamma(g_\alpha(f_\alpha(x_\alpha))) = h^\alpha_\gamma(x_\alpha) = x_\gamma$. Thus, $g_\gamma(\bigvee_{\alpha<\kappa} f_\alpha(x_\alpha)) = g_\gamma(\bigvee_{\gamma\leq\alpha<\kappa} f_\alpha(x_\alpha)) = x_\gamma$, since g_γ is locally completely additive. Since this holds for all $\gamma < \kappa$, we conclude that $g(f(x)) = x$ for all $x \in L_\infty$.

Suppose now that $y \in L$. Then $f(g(y)) = f((g_\alpha(y))_{\alpha<\kappa}) = \bigvee_{\alpha<\kappa} f_\alpha(g_\alpha(y)) \leq y$, since $f_\alpha(g_\alpha(y)) \leq y$ for all $\alpha < \kappa$.

□

Corollary 1. *Under the assumptions of the previous lemma, for all $(x_\alpha)_{\alpha<\kappa} \in L_\infty$, $\bigwedge\{y : y \in L,\ \forall\alpha < \kappa\ x_\alpha \leq g_\alpha(y)\} = \bigvee_{\alpha<\kappa} f_\alpha(x_\alpha)$.*

Lemma 6. *Let L_∞ be the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that L is a complete lattice and the locally completely additive projections $g_\alpha : L \to L_\infty$, $\alpha < \kappa$, form a cone. Then the unique mediating function $g : L \to L_\infty$ is a locally completely additive projection.*

Proof. Let $h^\infty_\alpha : L_\infty \to L_\alpha$, $\alpha < \kappa$, be the limit functions defined above. We know that they are locally completely additive projections. Suppose that $Y \subseteq L$,

$x = (x_\alpha)_{\alpha<\kappa} \in L_\infty$ and $g(Y) = \{x\}$. Then $g_\alpha(Y) = h_\alpha^\infty(g(Y)) = x_\alpha$, hence $g_\alpha(\bigvee Y) = x_\alpha$ for all $\alpha < \kappa$, since g_α is locally completely additive. Since this holds for all α, we have $g(\bigvee Y) = x$. On the other hand, g is a projection by Lemma 5. □

We now consider inverse systems with completely additive projections.

Lemma 7. *Let L_∞ be the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with projections $h_\beta^\alpha : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that each h_β^α is completely additive. Then the limit projections $h_\alpha^\infty : L_\infty \to L_\alpha$, $\alpha < \kappa$, are also completely additive.*

Proof. Let $X \subseteq L_\infty$ and $\alpha < \kappa$. Let X_α denote the set of α-components of the sequences in X. Since the supremum of X in L_∞ is formed pointwise, $h_\alpha^\infty(\bigvee X) = \bigvee_\alpha X_\alpha = \bigvee_\alpha h_\alpha^\infty(X)$. □

Lemma 8. *Let L_∞ be the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with projections $h_\beta^\alpha : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that each h_β^α for $\beta < \alpha < \kappa$ is also completely additive. Let L be a complete lattice and suppose that the completely additive functions $g_\alpha : L \to L_\alpha$, $\alpha < \kappa$ form a cone. Then the mediating function $g : L \to L_\infty$ is also completely additive.*

Proof. Indeed, for all $X \subseteq L$, $g(\bigvee X) = (g_\alpha(\bigvee X))_{\alpha<\kappa} = (\bigvee_\alpha g_\alpha(X))_{\alpha<\kappa} = \bigvee\{(g_\alpha(x))_{\alpha<\kappa} : x \in X\} = \bigvee g(X)$. □

4 Inverse Limit Models

In this section, our aim is to prove that the limit of an inverse system of complete lattices with locally completely additive projections determines a model. Moreover, when the projections of the inverse system are completely additive, then the limit determines a strong model.

Suppose that L_α, $\alpha < \kappa$, is an inverse system of complete lattices with projections $h_\beta^\alpha : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Let L_∞ denote the limit of the inverse system with limit projections $h_\alpha^\infty : L_\infty \to L_\alpha$.

For each $\alpha < \kappa$, define the relation $\sqsubseteq_\alpha$ on L_α by $x \sqsubseteq_\alpha y$ iff $x \le y$ and $h_\beta^\alpha(x) = h_\beta^\alpha(y)$ for all $\beta < \alpha$. Clearly, $\sqsubseteq_\alpha$ is a partial ordering of L_α which is included in the complete lattice order $\le$ on L_α.

We also define preorderings $\sqsubseteq_\alpha$ on L_∞. For all $\alpha < \kappa$ and $x = (x_\gamma)_{\gamma<\kappa}$ and $y = (y_\gamma)_{\gamma<\kappa}$ in L_∞, let $x \sqsubseteq_\alpha y$ iff $x_\alpha \sqsubseteq_\alpha y_\alpha$ in L_α, i.e., when $x_\alpha \le y_\alpha$ and $x_\beta = y_\beta$ for all $\beta < \alpha$. Thus, for all $x, y \in L_\infty$ and $\alpha < \kappa$, if $x \sqsubseteq_\alpha y$ then $h_\alpha^\infty(x) \sqsubseteq_\alpha h_\alpha^\infty(y)$, hence $h_\alpha^\infty(x) \le h_\alpha^\infty(y)$ and $h_\beta^\infty(x) = h_\beta^\infty(y)$ for all $\beta < \alpha$.

By the above definition, each $\sqsubseteq_\alpha$ is a preorder, so that L_∞ is a stratified complete lattice. Moreover, the intersection of all equivalence relations $=_\alpha$, determined by the preorderings $\sqsubseteq_\alpha$, $\alpha < \kappa$, is the identity relation on L_∞. Thus, A1 and A2 hold. We show that A3 holds.

Lemma 9. *Let L_∞ be the stratified complete lattice determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Then for all $x \in L_\infty$ and $\alpha < \kappa$ there is some $y \in L_\infty$ with $x =_\alpha y$ and such that for all $z \in L_\infty$, if $x \sqsubseteq_\alpha z$ then $y \le z$.*

Proof. Suppose that $x = (x_\gamma)_{\gamma<\kappa}$ is in L_∞. Let $\alpha < \kappa$ and define $y = (y_\gamma)_{\gamma<\kappa}$ as follows. Let $y_\gamma = x_\gamma$ for all $\gamma \le \alpha$. And if $\alpha < \gamma$, define $y_\gamma = k^\gamma_\alpha(x_\alpha)$, where k^γ_α is the embedding determined by the projection h^γ_α. Note that $y \in L_\infty$ and $y =_\alpha x$, since $y_\alpha = x$. In fact, $y = k^\infty_\alpha(h^\infty_\alpha(x))$, where the limit projection h^∞_α and corresponding embedding k^∞_α were defined above.

Let $z = (z_\gamma)_{\gamma<\kappa}$ in L_∞. Suppose that $x \sqsubseteq_\alpha z$. Then $y_\alpha = x_\alpha \le z_\alpha$ and $y_\beta = x_\beta = z_\beta$ for all $\beta < \alpha$. Suppose now that $\alpha < \beta < \kappa$. Then $y_\beta = k^\beta_\alpha(y_\alpha) = k^\beta_\alpha(x_\alpha) \le k^\beta_\alpha(z_\alpha) \le z_\beta$, since $x_\alpha \le z_\alpha$ and k^β_α is monotone, and since $h^\beta_\alpha(z_\beta) = z_\alpha$. Thus, $y \le z$ and $y \sqsubseteq_\alpha z$. □

Under the assumptions of Lemma 9, we denote $x|_\alpha = k^\infty_\alpha(h^\infty_\alpha(x))$ for all $x \in L_\infty$ and $\alpha < \kappa$.

Lemma 10. *Let L_∞ be the stratified complete lattice determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Then for all $x \in L_\infty$, it holds that $x = \bigvee_{\alpha<\kappa} x|_\alpha$.*

Proof. For all $\alpha < \kappa$, $x|_\alpha \le x$ and $x =_\alpha x|_\alpha$, i.e., the α-component of x agrees with the α-component of $x|_\alpha$. Thus, $\bigvee_{\alpha<\kappa} x|_\alpha \le x$ and $x \le y$ whenever $x|_\alpha \le y$ for all $\alpha < \kappa$. □

It is also clear that A5 and A6 hold. We thus have:

Corollary 2. *Let L_∞ be the stratified complete lattice determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Then L_∞, equipped with the relations $\sqsubseteq_\alpha$, $\alpha < \kappa$, satisfies A1, A2, A3, A5, A6. Moreover, $x = \bigvee_{\alpha<\kappa} x|_\alpha$ for all $x \in L_\infty$.*

Lemma 11. *Suppose that L_∞ is the stratified complete lattice determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that X is a nonempty subset of L_∞, $y \in L_\infty$ and $\alpha < \kappa$ with $X =_\alpha y$, i.e., $x =_\alpha y$ for all $x \in X$. Then $\bigvee X =_\alpha y$.*

Proof. Since $X =_\alpha y$, it holds that $h^\infty_\alpha(X) = y$. Since by Lemma 4, h^∞_α is locally completely additive, we conclude that $h^\infty_\alpha(\bigvee X) = y$, i.e., $\bigvee X =_\alpha y$. □

Proposition 1. *Let L_α, $\alpha < \kappa$, be the stratified complete lattice determined by an inverse system of complete lattices with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Then the inverse limit L_∞ is a model satisfying the axioms A1–A6 iff each of the projections h^α_β for $\beta < \alpha < \kappa$ is locally completely additive. Moreover, in this case, the limit functions $h^\infty_\alpha : L_\infty \to L_\alpha$, $\alpha < \kappa$, are locally completely additive projections.*

Proof. Suppose first that the projections h^α_β are locally completely additive. Then L is a model by Corollary 2 and Lemma 11. Moreover, the limit functions h^∞_α are locally completely additive projections by Lemmas 1 and 4.

Suppose now that L_∞ is a model. We want to prove that each h^α_β is locally completely additive. First we show that each h^∞_α is. Suppose that $Y \subseteq L_\infty$ is not empty and $h^\infty_\alpha(Y) = x$. Then $Y =_\alpha k^\infty_\alpha(x)$, since the α-component of each sequence in Y is x as is the α-component of $k^\infty_\alpha(x)$. Since L_∞ is a model, it follows that $\bigvee Y =_\alpha k^\infty_\alpha(x)$. This means that the α-component of $\bigvee Y$ agrees with the α-component x of $k^\infty_\alpha(x)$, hence $h^\infty_\alpha(\bigvee_\alpha Y) = x$.

Suppose now that $\beta < \alpha < \kappa$ and $x \in L_\beta$. Let $Y = (h^\alpha_\beta)^{-1}(x)$ and $Z = (h^\infty_\beta)^{-1}(x) = (h^\infty_\alpha)^{-1}(Y)$. Since h^∞_β is locally completely additive, $\bigvee Z \in Z$ and thus $h^\infty_\alpha(\bigvee Z) \in Y$. But $Y = h^\infty_\alpha(Z) \leq h^\infty_\alpha(\bigvee Z)$, thus $\bigvee_\alpha Y = h^\infty_\alpha(\bigvee Z) \in Y$. □

If the projections h^α_β are completely additive, then the stronger version A4* of axiom A4 holds.

Lemma 12. *Suppose that L_∞ is the model determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Suppose that $\alpha < \kappa$ and $x_i \sqsubseteq_\alpha y_i$ in L_∞ for all $i \in I$. Then $\bigvee_{i \in I} x_i \sqsubseteq_\alpha \bigvee_{i\in I} y_i$.*

Proof. By our assumption, the β-component of x_i agrees with the β-component of y_i for all $i \in I$ and $\beta < \alpha$. Moreover, for all $i \in I$, the α-component of x_i is less than or equal to the α-component of y_i. Since the supremum is formed pointwise (cf. Lemma 7), it follows that for all $\beta < \alpha$, the β-component of $\bigvee_{i\in I} x_i$ agrees with the β-component of $\bigvee_{i \in I} y_i$, and the α-component of $\bigvee_{i\in I} x_i$ is less than or equal to the α-component of $\bigvee_{i\in I} y_i$. Thus $\bigvee_{i\in I} x_i \sqsubseteq_\alpha \bigvee_{i\in I} y_i$. □

Proposition 2. *Let L_α, $\alpha < \kappa$, be an inverse system of complete lattices with projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$, and denote by L_∞ the stratified complete lattice determined by limit of the system. If the projections h^α_β are completely additive, then the inverse limit L_∞ is a strong model, i.e., it satisfies A1, A2, A3, A4*, A5 and A6. Moreover, the limit projections $h^\infty_\alpha : L_\infty \to L_\alpha$, $\alpha < \kappa$, are completely additive.*

Conversely, if L_∞ is a strong model, then the projections h^α_β, $\beta < \alpha < \kappa$, are completely additive.

Proof. Suppose that the projections h^α_β, $\beta < \alpha < \kappa$, are completely additive. Then they are locally completely additive, hence L_∞ is a model by Proposition 1. Thus, by Lemma 12, L_∞ is a strong model.

Suppose now that L_∞ is a strong model. Let $X \subseteq L_\infty$ and $\alpha < \kappa$. Since $x =_\alpha k^\infty_\alpha(h^\infty_\alpha(x))$ for all $x \in X$ and L_∞ is a strong model, we have $\bigvee X =_\alpha \bigvee k^\infty_\alpha(h^\infty_\alpha(X)) = k^\infty_\alpha(\bigvee_\alpha(h^\infty_\alpha(X))$, where the last equality holds since k^∞_α preserves all suprema. Applying h^∞_α this gives $h^\infty_\alpha(\bigvee X) = h^\infty_\alpha(k^\infty_\alpha(\bigvee_\alpha h^\infty_\alpha(X))) = \bigvee_\alpha h^\infty_\alpha(X)$. Thus each h^∞_α is completely additive. It follows that for each $\beta < \alpha < \kappa$, $h^\alpha_\beta = h^\infty_\beta \circ k^\infty_\alpha$, h^α_β is also completely additive. □

Example 1. Let $\kappa = \Omega$ be the least uncountable ordinal, and for each $\alpha < \Omega$, let L_α be the linearly ordered lattice $F_0 < \cdots < F_\alpha < 0 < T_\alpha < \cdots < T_0$. For all $\beta < \alpha < \Omega$, define $h^\alpha_\beta : L_\alpha \to L_\beta$ by $h^\alpha_\beta(F_\gamma) = F_\gamma$ and $h^\alpha_\beta(T_\gamma) = T_\gamma$, for all $\gamma \leq \beta$, and let $h^\alpha_\beta(x) = 0$, otherwise. Then all of the assumptions of Proposition 2 are satisfied so that L_∞ is a strong model. In fact, L_∞ is isomorphic to V_Ω. An isomorphism $L_\infty \to V_\Omega$ is given by the assignment that maps the sequence $(0, 0, \ldots, F_\alpha, F_\alpha, \ldots)$ to F_α, the sequence $(0, 0, \ldots, T_\alpha, T_\alpha, \ldots)$ to T_α, where $\alpha < \Omega$ and the first F_α or T_α occurs in position α, and the 0-sequence $(0, 0, \ldots)$ to 0.

We will prove in Sect. 7 that every model satisfying the axioms A1–A6 is isomorphic to a model determined by the limit of an inverse system of complete lattices with locally completely additive projections. Moreover, we will prove that every strong model is isomorphic to a model determined by the limit of an inverse system of complete lattices with completely additive projections.

5 Some Properties of Models

In this section, we establish several consequences of the axioms. These results will be used in our proof of the fact that every model is isomorphic to an inverse limit model. Suppose that L satisfies the axioms A1–A6. For each $x \in L$ and $\alpha < \kappa$, let $[x]_\alpha = \{y \in L : x =_\alpha y\}$. Moreover, for each $\alpha < \kappa$, let $L|_\alpha = \{x|_\alpha : x \in L\}$.

Lemma 13. *For each $x \in L$ and $\alpha < \kappa$, it holds that $x =_\alpha x|_\alpha$, $x|_\alpha \leq x$, and $x|_\alpha$ is the $\leq$-least element of $[x]_\alpha$.*

Proof. The first claim is clear, since by A3, $x =_\alpha x|_\alpha$. Suppose that $y \in [x]_\alpha$. Then $x =_\alpha y$ and so $x \sqsubseteq_\alpha y$. Thus, $x|_\alpha \leq y$, by A3 and the definition of $x|_\alpha$. In particular, since $x \in [x]_\alpha$, it holds that $x|_\alpha \leq x$. □

Corollary 3. *For all $x \in L$ and $\alpha < \kappa$, it holds that $x|_\alpha = \bigwedge [x]_\alpha = \bigwedge \{y : x \sqsubseteq_\alpha y\}$.*

Corollary 4. *For all $x \in L$, $\bigvee_{\alpha<\kappa} x|_\alpha \leq x$.*

Corollary 5. *For all $x, y \in L$ and $\alpha < \kappa$, it holds that $x \sqsubseteq_\alpha y$ iff $x|_\alpha \sqsubseteq_\alpha y$ iff $x \sqsubseteq_\alpha y|_\alpha$ iff $x|_\alpha \sqsubseteq_\alpha y|_\alpha$.*

Proof. This follows from the fact $x =_\alpha x|_\alpha$ and $y =_\alpha y|_\alpha$, proved in Lemma 13. □

Corollary 6. *For all $x, y \in L$ and $\alpha < \kappa$, it holds that $x =_\alpha y$ iff $x|_\alpha =_\alpha y$ iff $x|_\alpha =_\alpha y|_\alpha$. Moreover, $x =_\alpha y$ iff $x|_\alpha = y|_\alpha$.*

Proof. This follows from Corollary 5 and Lemma 13, by noting that if $x =_\alpha y$, then $[x]_\alpha = [y]_\alpha$, so $x|_\alpha$ and $y|_\alpha$ are $\leq$-least elements of the same set. □

Lemma 14. *Suppose that $x \in L$ and $\alpha < \beta < \kappa$. Then $x|_\alpha =_\alpha x|_\beta$ and $x|_\alpha \leq x|_\beta$.*

Proof. By Lemma 13, it holds that $x|_\alpha =_\alpha x =_\beta x|_\beta$. Since by A1 the relation $=_\beta$ is included in the relation $=_\alpha$, we conclude that $x|_\alpha =_\alpha x|_\beta$. Since $[x]_\beta \subseteq [x]_\alpha$, the $\leq$-least element of $[x]_\alpha$ is less than or equal to the $\leq$-least element of $[x]_\beta$. Thus, by Lemma 13, $x|_\alpha \leq x|_\beta$. □

Lemma 15. *Suppose that $x \in L$ and $\alpha, \beta < \kappa$. If $\alpha \leq \beta$ then $(x|_\alpha)|_\beta = x|_\alpha$. If $\beta < \alpha$ then $(x|_\alpha)|_\beta = x|_\beta$.*

Proof. By Lemma 13, it holds that $(x|_\alpha)|_\beta \leq x|_\alpha$. If $\alpha \leq \beta$ then, since $x|_\alpha =_\alpha x$, by A1 we have $[x|_\alpha]_\beta \subseteq [x]_\alpha$, hence the $\leq$-least element of $[x]_\alpha$ is less than or equal to the $\leq$-least element of $[x|_\alpha]_\beta$. Thus, by Lemma 13, $x|_\alpha \leq (x|_\alpha)|_\beta$. We conclude that $(x|_\alpha)|_\beta = x|_\alpha$.

Suppose now that $\beta < \alpha$. Then by $x|_\alpha =_\alpha x$, which holds by Lemma 13, and by the fact that the relation $=_\alpha$ is included in $=_\beta$, which holds by A1, we have $[x|_\alpha]_\beta = [x]_\beta$. Thus, $(x|_\alpha)|_\beta = x|_\beta$ by Lemma 13. □

Corollary 7. *For all $x \in L$ and $\alpha < \kappa$, $x \in L|_\alpha$ iff $x = x|_\alpha$.*

Proof. Recall that $L|_\alpha = \{y|_\alpha : y \in L\}$. Thus, if $x = y|_\alpha$ is in $L|_\alpha$, then $x|_\alpha = (y|_\alpha)|_\alpha = y|_\alpha = x$. If $x = x|_\alpha$, then clearly $x \in L|_\alpha$. □

Corollary 8. *For all $x, y \in L|_\alpha$, $x =_\alpha y$ iff $x = y$.*

Proof. Suppose that $x, y \in L|_\alpha$. Then $x = x|_\alpha$ and $y = y|_\alpha$. We conclude by Corollary 6. □

Lemma 16. *For all $x, y \in L$ and $\alpha < \kappa$, if $x \sqsubseteq_\alpha y$ then $x|_\alpha \leq y|_\alpha$.*

Proof. If $x \sqsubseteq_\alpha y$ then by $y =_\alpha y|_\alpha$, also $x \sqsubseteq_\alpha y|_\alpha$, hence $x|_\alpha \leq y|_\alpha$ by A3. □

Corollary 9. *For all $\alpha < \kappa$ and $x, y \in L|_\alpha$, if $x \sqsubseteq_\alpha y$ then $x \leq y$.*

The above facts were all consequences of the first and the third axiom. We will now make use of A2 and A4 in order to prove a strengthened version of Corollary 4.

Lemma 17. *For all $x \in L$ and $\alpha < \kappa$, $x = \bigvee_{\alpha<\kappa} x|_\alpha$.*

Proof. Let $\gamma < \kappa$ be any ordinal. By Lemma 14, the sequence $(x|_\alpha)_{\alpha<\kappa}$ is an increasing chain in L. Thus $\bigvee_{\alpha<\kappa} x|_\alpha = \bigvee_{\gamma\leq\alpha<\kappa} x|_\alpha$. But for all α with $\gamma \leq \alpha$, $x|_\alpha =_\gamma x|_\gamma =_\gamma x$ by Lemmas 13 and 14. Hence, by A4, $\bigvee_{\gamma\leq\alpha<\kappa} x|_\alpha =_\gamma x$ and thus $\bigvee_{\alpha<\kappa} x|_\alpha =_\gamma x$. Since this holds for all $\gamma < \kappa$, we conclude by A2 that $x = \bigvee_{\alpha<\kappa} x|_\alpha$. □

Lemma 18. *For all $\alpha < \kappa$, nonempty families $x_i \in L$, $i \in I$, and $y \in L$, if $x_i|_\alpha = y$ for all $i \in I$, then $(\bigvee_{i\in I} x_i)|_\alpha = y$.*

Proof. This is clear from A4 and Corollary 8, since our assumption implies that $y \in L|_\alpha$. □

Remark 2. A certain converse of Lemma 18 also holds. If A1, A2, A3 and the condition formulated in Lemma 18 hold, and if $y \in L$ and $x_i \in L$ with $x_i =_\alpha y$ for all $i \in I$, where I is a nonempty set, then by Corollary 8, $x_i|_\alpha = y|_\alpha$ for all $i \in I$, hence $(\bigvee_{i \in I} x_i)|_\alpha = y|_\alpha$. By Corollary 8 this means that $\bigvee_{i \in I} x_i =_\alpha y$, i.e., A4 holds.

The next facts also use A5.

Corollary 10. *For all $x, y \in L$, $x \leq y$ iff $x|_\alpha \leq y|_\alpha$ for all $\alpha < \kappa$.*

Proof. Suppose that $x|_\alpha \leq y|_\alpha$ for all $\alpha < \kappa$. Then by Lemma 17, $x = \bigvee_{\alpha<\kappa} x|_\alpha \leq \bigvee_{\alpha<\kappa} y|_\alpha = y$. The reverse direction holds by A5. □

Corollary 11. *For all $x, y \in L$ and $\alpha < \kappa$, $x|_\alpha \leq y$ iff $x|_\alpha \leq y|_\alpha$.*

Proof. This follows from Corollary 10 using the fact that $(x|_\alpha)|_\alpha = x|_\alpha$, proved in Lemma 15. □

The next facts depend on A6.

Lemma 19. *The following conditions are equivalent for all $x, y \in L$ and $\alpha < \kappa$: (i) $x \sqsubseteq_\alpha y$, (ii) $x|_\alpha \leq y|_\alpha$ and $x =_\beta y$ for all $\beta < \alpha$, (iii) $x|_\alpha \leq y$ and $x =_\beta y$ for all $\beta < \alpha$.*

Proof. Suppose that $x \sqsubseteq_\alpha y$. Then $x =_\beta y$ for all $\beta < \alpha$ by A1, and $x|_\alpha \leq y|_\alpha$ by Lemma 16. But if $x|_\alpha \leq y|_\alpha$, then also $x|_\alpha \leq y$, since by Lemma 13, $y|_\alpha \leq y$.

Suppose that $x|_\alpha \leq y$ and $x =_\beta y$ for all $\beta < \alpha$. Then $x|_\alpha =_\beta y$ for all $\beta < \alpha$, hence $x|_\alpha \sqsubseteq_\alpha y$ by A6. Thus, by Lemma 13, $x \sqsubseteq_\alpha y$. □

Corollary 12. *For all $x, y \in L$ and $\alpha < \kappa$, $x|_\alpha \sqsubseteq_\alpha y|_\alpha$ iff $x|_\alpha \leq y|_\alpha$ and $x|_\beta =_\beta y|_\beta$ for all $\beta < \alpha$.*

Proof. Immediate from Lemma 19 and Corollaries 7 and 8. □

Corollary 13. *For all $x, y \in L|_\alpha$, $x \sqsubseteq_\alpha y$ iff $x|_\alpha \leq y|_\alpha$ and $x|_\beta = y|_\beta$ for all $\beta < \alpha$.*

Proof. This is immediate from Corollaries 9 and 12. □

For each set $X \subseteq L$ and ordinal $\alpha < \kappa$, let us define $X|_\alpha = \{x|_\alpha : x \in X\}$. Note that this notation is consistent with the notation $L|_\alpha$ introduced earlier.

Suppose now that L is a strong model satisfying A4*.

Lemma 20. *For all $X \subseteq L$ and $\alpha < \kappa$, $\bigvee X|_\alpha = (\bigvee X)|_\alpha$.*

Proof. Let $X \subseteq L$ and $\alpha < \kappa$. Since by Lemma 13 $x =_\alpha x|_\alpha$ for all $x \in X$, it holds by A4* that $\bigvee X =_\alpha \bigvee X|_\alpha$. Thus, $(\bigvee X)|_\alpha \leq \bigvee X|_\alpha$, again by Lemma 13.

Since $x \leq \bigvee X$ for all $x \in X$, by A5 we have $x|_\alpha \leq (\bigvee X)|_\alpha$ for all $x \in X$. It follows that $\bigvee X|_\alpha \leq (\bigvee X)|_\alpha$. □

Remark 3. Suppose that A1, A2 and A3 hold. Moreover, suppose that the property described in Lemma 20 holds. Then we can show that A4* and A5 hold. Thus, in the definition of strong models, these two axioms may be replaced by the property in Lemma 20.

Indeed, if $x \leq y$ then for all $\alpha < \kappa$, $y|_\alpha = (x \vee y)|_\alpha = x|_\alpha \vee y|_\alpha$, hence $x|_\alpha \leq y|_\alpha$. And if $x_i =_\alpha y_i$ for all $i \in I$, where $\alpha < \kappa$, then by Corollary 6, $x_i|_\alpha = y_i|_\alpha$ for all $i \in I$, thus $(\bigvee_{i \in I} x_i)|_\alpha = \bigvee_{i \in I} x_i|_\alpha = \bigvee_{i \in I} y_i|_\alpha = (\bigvee_{i \in I} y_i)|_\alpha$. We conclude that $\bigvee_{i \in I} x_i =_\alpha \bigvee_{i \in I} y_i$.

6 An Alternative Axiomatization

We used axiom A3 to equip a model L with an operation $|_\alpha : L \to L$ for each $\alpha < \kappa$, mapping $x \in L$ to $x|_\alpha$ in $L|_\alpha \subseteq L$. In this section we give an alternative axiomatization using these operations $|_\alpha$ instead of the preorderings $\sqsubseteq_\alpha$.

Theorem 1. *Suppose that L is a model satisfying the axioms A1–A6. For each $\alpha < \kappa$ and $x \in L$, let $x|_\alpha$ be defined by the following property (cf. A3):*

C. $x|_\alpha =_\alpha x$ and for all $y \in L$, if $x \sqsubseteq_\alpha y$ then $x|_\alpha \leq y$.

Then, equipped with the operations $|_\alpha : L \to L$ for $\alpha < \kappa$, the following hold:

B1. For all $x \in L$ and $\beta \leq \alpha < \kappa$, $(x|_\alpha)|_\beta = x|_\beta$.
B2. For all $x, y \in L$ and $\alpha < \kappa$, if $x \leq y$ then $x|_\alpha \leq y|_\alpha$.
B3. For all $x \in L$, $x = \bigvee_{\alpha < \kappa} x|_\alpha$.
B4. For all $\alpha < \kappa$ and y and $x_i \in L$, $i \in I$, where I is a nonempty index set, if $x_i|_\alpha = y$ then $(\bigvee_{i \in I} x_i)|_\alpha = y$.

Moreover, the following holds:

D. For each $\alpha < \kappa$ and $x, y \in L$, it holds that $x \sqsubseteq_\alpha y$ iff $x|_\alpha \leq y|_\alpha$ and $x|_\beta = y|_\beta$ for all $\beta < \alpha$.

Suppose that $(L, \leq)$ is a complete lattice equipped with a family of functions $|_\alpha : L \to L$, $\alpha < \kappa$, satisfying the axioms B1–B4. For each $\alpha < \kappa$, define the relation $\sqsubseteq_\alpha$ on L by the condition D. Then, equipped with these relations $\sqsubseteq_\alpha$, L is a model satisfying the axioms A1–A6. Moreover, C holds.

Proof. We have already proved that when L is a model satisfying the axioms A1–A6, then equipped with the operations $|_\alpha : L \to L$, $\alpha < \kappa$, uniquely defined by C, L satisfies B1–B4. In fact, B2 is the same as A5. Moreover, D holds. (See Lemmas 15, 17, 18 and Corollary 13.)

Suppose now that L is a complete lattice equipped with a family of functions $|_\alpha : L \to L$, $\alpha < \kappa$, satisfying B1–B4. Define the relations $\sqsubseteq_\alpha$, $\alpha < \kappa$, by D. Then each of the relations $\sqsubseteq_\alpha$, $\alpha < \kappa$, is clearly a preordering, and if $\beta < \alpha$, then $\sqsubseteq_\alpha$ is contained in $=_\beta$. Thus A1 holds.

In order to prove that A2 holds, note first that if $x \leq y$ then $x|_\alpha \leq y|_\alpha$ for all $\alpha < \kappa$, by B2, and if $x|_\alpha \leq y|_\alpha$ for all $\alpha < \kappa$, then $x \leq y$, by B3. Thus, $x \leq y$

iff $x|_\alpha \leq y|_\alpha$ for all $\alpha < \kappa$, and $x = y$ iff $x|_\alpha = y|_\alpha$ for all $\alpha < \kappa$ iff $x =_\alpha y$ for all $\alpha < \kappa$, proving A2.

Now we prove A3. First note that for all $\alpha < \kappa$ and $x \in L$, $x =_\alpha x|_\alpha$, since by B1, $(x|_\alpha)|_\beta = x|_\beta$ for all $\beta \leq \alpha$. Moreover, if $x \sqsubseteq_\alpha y$, then by D and B3, $x|_\alpha \leq y|_\alpha \leq y$.

Axiom A4 holds by B4 and Remark 2. Axiom A5 holds since it is the same as B2. Finally, axiom A6 holds, since if $x \leq y$ in L and $x|_\beta = y|_\beta$ for all $\beta < \alpha$, where $\alpha < \kappa$, then, by B2, also $x|_\alpha \leq y|_\alpha$ and thus $x \sqsubseteq_\alpha y$ by D. □

Corollary 14. *Suppose that L is a strong model satisfying the axioms A1, A2, A3, A4*, A5 and A6. For each $\alpha < \kappa$ and $x \in L$, let $x|_\alpha$ be defined by the property C above. Then, equipped with the operations $|_\alpha : L \to L$ for $\alpha < \kappa$, B1, B3 and the following hold:*

B2*. For all $X \subseteq L$ and $\alpha < \kappa$, $(\bigvee X)|_\alpha = \bigvee X|_\alpha$.

Moreover, D holds.

Suppose that L is a complete lattice equipped with a family of functions $|_\alpha : L \to L$, $\alpha < \kappa$, satisfying the axioms B1, B2* and B3. For each $\alpha < \kappa$, define the relation $\sqsubseteq_\alpha$ on L by the condition D. Then, equipped with these relations $\sqsubseteq_\alpha$, L is a strong model. Moreover, C holds.

Proof. One uses Lemma 20 and Remark 3. □

Remark 4. The proof of Theorem 1 entails also the following result. Suppose that L is a stratified complete lattice satisfying the axioms A1, A2, A3, A5, A6 and B3, where for each $\alpha < \kappa$ and $x \in L$, $x|_\alpha$ is defined by the property C. Then, equipped with the operations $|_\alpha : L \to L$ for $\alpha < \kappa$, B1, B2 and D hold.

Suppose that $(L, \leq)$ is a complete lattice equipped with a family of functions $|_\alpha : L \to L$, $\alpha < \kappa$, satisfying the axioms B1, B2, B3. For each $\alpha < \kappa$, define the relation $\sqsubseteq_\alpha$ on L by the condition D. Then, equipped with these relations $\sqsubseteq_\alpha$, L satisfies A1, A2, A3, A5 and A6. Moreover, C holds.

7 The Representation Theorem

In this section, we prove that every model satisfying the axioms A1–A6 introduced in Sect. 2 is isomorphic to an inverse limit model. In our argument, we will make use of the properties of models established in the previous sections.

Proposition 3. *Suppose that L is a model satisfying A1–A6. Then for each $\alpha < \kappa$, $L|_\alpha$, equipped with the ordering inherited from L, is a complete lattice. Moreover, for all $X \subseteq L|_\alpha$, the infimum $\bigwedge_\alpha X$ of X in $L|_\alpha$ is $(\bigwedge X)|_\alpha$, where $\bigwedge X$ is the infimum of X in L. Similarly, the supremum $\bigvee_\alpha X$ of X in $L|_\alpha$ is $(\bigvee X)|_\alpha$, where $\bigvee X$ is the supremum of X in L.*

Proof. Suppose that L is a model. Let $\alpha < \kappa$ and $X \subseteq L|_\alpha$.

Since by Lemma 13 (or B3), $(\bigwedge X)|_\alpha \leq \bigwedge X$, we have $(\bigwedge X)|_\alpha \leq X$. Suppose that $z \in L|_\alpha$ with $z \leq X$. Then $z \leq \bigwedge X$, hence $z \leq (\bigwedge X)|_\alpha$ by Corollaries 7 and 11, or B1 and B2. We have completed the proof of the fact that $(\bigwedge X)|_\alpha$ is the infimum of X in $L|_\alpha$, i.e., $\bigwedge_\alpha X = (\bigwedge X)|_\alpha$.

The proof of $\bigvee_\alpha X = (\bigvee X)|_\alpha$ is similar. First, $X \leq \bigvee X$, hence $X \leq (\bigvee X)|_\alpha$ by Corollary 11, or B1 and B2. And if $z \in L|_\alpha$ with $X \leq z$, then $\bigvee X \leq z$, hence $(\bigvee X)|_\alpha \leq z$ by Lemma 13 (or B3). □

An example of a five-element model L such that there exist $x, y \in L|_0$ with $x \wedge y \neq x \wedge_0 y$ is given in [5].

Proposition 4. *Suppose that L is a model satisfying A1–A6. For any ordinals α, β with $\beta \leq \alpha < \kappa$, define $h^\alpha_\beta : L|_\alpha \to L|_\beta$ by $h^\alpha_\beta(x) = x|_\beta$ for all $x \in L|_\alpha$. Then each of the functions h^α_β for $\beta \leq \alpha < \kappa$ is surjective. For all $\alpha < \kappa$, h^α_α is the identity function $L|_\alpha \to L|_\alpha$, and for all $\gamma < \beta < \alpha < \kappa$, $h^\beta_\gamma \circ h^\alpha_\beta = h^\alpha_\gamma$. Moreover, the following hold:*

(i) For all $\beta < \alpha < \kappa$, $h^\alpha_\beta : L|_\alpha \to L|_\beta$ is a projection.
(ii) For all $\beta < \alpha < \kappa$, h^α_β is locally completely additive.
(iii) For all $\alpha < \kappa$ and $x, y \in L|_\alpha$, $x \sqsubseteq_\alpha y$ iff $x \leq y$ and $h^\alpha_\beta(x) = h^\alpha_\beta(y)$ for all $\beta < \alpha$.

Proof. Suppose that $\beta \leq \alpha < \kappa$. For all $x \in L$, it holds by Lemma 15 (or B1) that $(x|_\alpha)|_\beta = x|_\beta$. Thus, h^α_β is surjective.

By Lemma 15 (or B1), h^α_α is the identity function $L|_\alpha \to L|_\alpha$ for all $\alpha < \kappa$. The fact that $h^\beta_\gamma \circ h^\alpha_\beta = h^\alpha_\gamma$ for all $\gamma < \beta < \alpha$ also follows from Lemma 15 (or B1), since for all $x \in L$, $((x|_\alpha)|_\beta)|_\gamma = x|_\gamma = (x|_\alpha)|_\gamma$.

Suppose that $\beta < \alpha < \kappa$. If $x \leq y$ in $L|_\alpha$, then $x|_\beta \leq y|_\beta$ by A5 or B2. Thus, h^α_β is monotone. It follows from Lemma 15 that for all $\beta < \alpha < \kappa$, $L|_\beta \subseteq L|_\alpha$. Let $x \in L|_\beta$ and $y \in L|_\alpha$ with $x \leq y|_\beta$. Since $x \in L|_\beta$, it holds that $x = x|_\beta$, by Lemma 15 or B1. But again by Lemma 13 (or B3), $x|_\beta = x \leq y|_\beta \leq y$, so $x \leq y$. Also, if $x \leq y$, then $x \leq y|_\beta$. Thus, h^α_β is a projection with corresponding embedding $k^\alpha_\beta : L|_\beta \to L|_\alpha$ being the inclusion function.

Next we prove that each function h^α_β for $\beta < \alpha < \kappa$ is locally completely additive. To this end, suppose that $Y \subseteq L|_\alpha$ and $x \in L|_\beta$ with $h^\alpha_\beta(Y) = \{x\}$, so that Y is not empty and $y|_\beta = x$ for all $y \in Y$. Then, by Corollary 7 and Corollary 8, or B1 and D, $y =_\beta x$ for all $y \in Y$, i.e., $Y =_\beta x$. We conclude by A4 that $\bigvee Y =_\beta x$ and thus $(\bigvee Y)|_\beta = x$, again by Corollaries 7 and 8, or B1 and D. Thus, $h^\alpha_\beta(\bigvee_\alpha Y) = (\bigvee_\alpha Y)|_\beta = ((\bigvee Y)|_\alpha)|_\beta = (\bigvee Y)|_\beta = x$, by Proposition 3 and either Lemma 15 or B1.

The last claim holds by Corollary 13 or D. □

We are now ready to prove the Representation Theorem. By Proposition 4, for every model L satisfying the axioms A1–A6, the complete lattices $L|_\alpha$ equipped with the locally completely additive projections $h^\alpha_\beta : L|_\alpha \to L|_\beta$ defined by $h^\alpha_\beta(x) = x|_\beta$ for all $x \in L|_\alpha$ and $\beta < \alpha < \kappa$ form an inverse system. We can thus form the limit model L_∞ as in Sect. 4. We know that L_∞ is a model satisfying the axioms A1–A6. But actually L_∞ is isomorphic to L.

Theorem 2. *Every model L satisfying the axioms A1–A6 is isomorphic to the model determined by the limit of the inverse system of the complete lattices $L|_\alpha$, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L|_\alpha \to L|_\beta$, defined by $h^\alpha_\beta(x) = x|_\beta$ for all $x \in L|_\alpha$, where $\beta < \alpha < \kappa$.*

Proof. Let L_∞ denote the inverse limit. We intend to show that L is isomorphic to L_∞. Recall that for each $\alpha < \kappa$, the limit projection $h^\infty_\alpha : L_\infty \to L|_\alpha$ maps a sequence $x \in L_\infty$ to its α-component x_α. We know from Proposition 1 that these functions are locally completely additive projections and constitute a cone over the inverse system $h^\alpha_\beta : L|_\alpha \to L|_\beta$.

We define another cone. For each $\alpha < \kappa$, let $f_\alpha : L \to L|_\alpha$ be defined by $f_\alpha(x) = x|_\alpha$. Note that each f_α is monotone and locally completely additive (Lemma 18) and a projection (Corollaries 7 and 11). Moreover, by Lemma 15 (or B1), $h^\alpha_\beta(f_\alpha(x)) = (x|_\alpha)|_\beta = x|_\beta = f_\beta(x)$ for all $\beta < \alpha$ and $x \in L$. Thus, there is a unique function $f : L \to L_\infty$ with $h^\infty_\alpha \circ f = f_\alpha$ for all $\alpha < \kappa$. We know that the function f, given by $f(x) = (x|_\alpha)_{\alpha<\kappa}$, is a locally completely additive projection (Lemmas 5 and 6). By Corollary 10, f is an isomorphism.

To complete the proof, we still need to show that f creates an isomorphism between $(L, \sqsubseteq_\alpha)$ and $(L_\infty, \sqsubseteq_\alpha)$ for each α. But this is clear, since for all $x, y \in L$, $x \sqsubseteq_\alpha y$ iff $x|_\alpha \sqsubseteq_\alpha y|_\alpha$, as shown above (Corollary 5). □

Example 2. Consider the model $L = V_\Omega$ defined above and recall Example 1. Then for each $\alpha < \Omega$, $L|_\alpha$ is isomorphic to L_α and the functions $h^\alpha_\beta : L|_\alpha \to L|_\beta$ given by $x \mapsto x|_\beta$ correspond to the functions $h^\alpha_\beta : L_\alpha \to L_\beta$ described in Example 1.

Corollary 15. *Every strong model L is isomorphic to the model determined by the limit of the inverse system of the complete lattices $L|_\alpha$, $\alpha < \kappa$, with completely additive projections $h^\alpha_\beta : L|_\alpha \to L|_\beta$, defined by $h^\alpha_\beta(x) = x|_\beta$ for all $x \in L|_\alpha$, where $\beta < \alpha < \kappa$.*

Proof. By Theorem 2 and Proposition 2. □

Corollary 16. *Let L be a stratified complete lattice equipped with a preordering $\sqsubseteq_\alpha$ for each $\alpha < \kappa$. Then L is a model satisfying the axioms A1–A6 iff L is isomorphic to the model determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$.*

Corollary 17. *Let L be a stratified complete lattice equipped with a preordering $\sqsubseteq_\alpha$ for each $\alpha < \kappa$. Then L is a strong model iff L is isomorphic to the model determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$.*

8 Some Further Properties of Models

In this section, we establish several further properties of models. Some of these properties have been axioms in [6,7], see Propositions 5, 6. Some others, such as the ones formulated in Corollaries 20 and 21, were proved in [6] for a larger class of models. Our aim here is to use the Representation Theorem to provide alternative proofs of these results. In Corollary 20, we will prove that if L is a model, then it may naturally be equipped with another complete partial order $\sqsubseteq$. Then, in Corollary 21, we will show that certain weakly monotone functions over L have least pre-fixed points with respect to the ordering $\sqsubseteq$, and that these least pre-fixed points are in fact fixed points. Actually we will derive these facts from a new technical result formulated in Theorem 3, which also implies that the collection of all fixed points is in fact a complete lattice in itself w.r.t. the ordering $\sqsubseteq$, cf. Corollary 19.

In this section, we will without loss of generality suppose that a model L is given as the model determined by the limit L_∞ of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$ and corresponding embeddings $k^\alpha_\beta : L_\beta \to L_\alpha$, $\beta < \alpha < \kappa$. As before, we will denote the limit projection $L \to L_\alpha$ for $\alpha < \kappa$ by h^∞_α. As noted above, the embeddings k^α_β, as well as the embeddings $k^\infty_\alpha : L_\alpha \to L$, corresponding to the projections h^∞_α, are locally completely additive. Recall that an element of an inverse limit model L_∞ is a sequence $x = (x_\alpha)_{\alpha<\kappa}$, which is compatible in the sense that $h^\alpha_\beta(x_\alpha) = x_\beta$ for all $\beta < \alpha < \kappa$. As opposed to previous sections, instead of $\bigvee_\alpha X$ and $\bigwedge_\alpha X$, we will simply denote the supremum and infimum of a set $X \subseteq L_\alpha$, $\alpha < \kappa$, by $\bigvee X$ and $\bigwedge X$, respectively.

The properties established in all models by Propositions 5 and 6 below have been axioms in [6]. We include these propositions in order to connect this paper with [6].

Proposition 5. *Suppose that L is model satisfying A1–A6. Let $x \in L$, $\alpha < \kappa$ and $X \subseteq (x]_\alpha = \{z : \forall\beta < \alpha\ x =_\beta z\}$. Then there exists some $y \in (x]_\alpha$ with the following properties:*

(i) $X \sqsubseteq_\alpha y$ (i.e., $x \sqsubseteq_\alpha y$ for all $x \in X$),
(ii) for all $z \in (x]_\alpha$, if $X \sqsubseteq_\alpha z$ then $y \le z$ and $y \sqsubseteq_\alpha z$.

Proof. Before giving the proof, let us remark that for the notion of model as used in this paper, Proposition 5 greatly simplifies. Using the above assumption and notation, since $X \subseteq (x]_\alpha$ and $y, z \in (x]_\alpha$, $X \sqsubseteq_\alpha y$ holds iff $X|_\alpha \le y$, and similarly for $X \sqsubseteq_\alpha z$. Moreover, $y \sqsubseteq_\alpha z$ iff $y|_\alpha \le z$. See Lemma 19. But since $y|_\alpha \le y$ (cf. Lemma 13), we have $y \sqsubseteq_\alpha z$ and $y \le z$ iff $y \le z$. Thus, the above property amounts to the following assertion: for each $X \subseteq (x]_\alpha$ in a model L satisfying A1–A6, there is some $y \in (x]_\alpha$ with $X|_\alpha \le y$ and such that for all $z \in L$, if $X|_\alpha \le z$ then $y \le z$.

In our proof, we make use of Theorem 2. So without loss of generality suppose that $L = L_\infty$ is the model determined by the limit of an appropriate inverse system as described above. Then $x = (x_\beta)_{\beta<\kappa}$ is a compatible sequence, and $(x]_\alpha = \{(z_\beta)_{\beta<\kappa} \in L : \forall\beta < \alpha\ x_\beta = z_\beta\}$.

If X is empty, let $y = \bigvee_{\gamma<\alpha} k^\infty_\gamma(x_\gamma)$, which is the least element of $(x]_\alpha$. Indeed, for any $\beta < \alpha$, $h^\infty_\beta(y) = h^\infty_\beta(\bigvee_{\gamma<\alpha} k^\infty_\gamma(x_\gamma)) = h^\infty_\beta(\bigvee_{\beta\leq\gamma<\alpha} k^\infty_\gamma(x_\gamma))$, since the sequence $(k^\infty_\gamma(x_\gamma))_{\gamma<\alpha}$ is increasing. But for all γ with $\beta \leq \gamma < \alpha$, $h^\infty_\beta(k^\infty_\gamma(x_\gamma)) = h^\gamma_\beta(x_\gamma) = x_\beta$. Thus, since h^∞_β is locally completely additive, we have $h^\infty_\beta(\bigvee_{\gamma<\alpha} k^\infty_\gamma(x_\gamma)) = \bigvee_{\beta\leq\gamma<\alpha} h^\infty_\beta(k^\infty_\gamma(x_\gamma)) = \bigvee_{\beta\leq\gamma<\alpha} x_\beta = x_\beta$. And if $z = (z_\beta)_{\beta<\kappa} \in (x]_\alpha$, then $x_\beta = z_\beta = h^\infty_\beta(z)$ for all $\beta < \alpha$, hence $k^\infty_\beta(x_\beta) \leq z$ for all $\beta < \alpha$, so that $y = \bigvee_{\beta<\alpha} k^\infty_\beta(x_\beta) \leq z$.

If X is not empty, then define $y = k^\infty_\alpha(\bigvee X_\alpha) = \bigvee k^\infty_\alpha(X_\alpha)$, where X_α is the set of all α-components of the elements of X. Since $(h^\infty_\alpha, k^\infty_\alpha)$ is a projection-embedding pair, y is the least element of L with $X_\alpha \leq h^\infty_\alpha(y)$, or equivalently, $\bigvee X_\alpha \leq h^\infty_\alpha(y)$. To complete the proof, we still need to show that $y \in (x]_\alpha$. But for all $\beta < \alpha$, $h^\infty_\beta(y) = h^\infty_\beta(k^\infty_\alpha(\bigvee X_\alpha)) = h^\alpha_\beta(\bigvee X_\alpha) = x_\beta$, since $h^\alpha_\beta(X_\alpha) = x_\beta$ and h^α_β is locally completely additive. □

We will denote the element y constructed above by $\bigsqcup_\alpha X$. Note that when X is empty, $\bigsqcup_\alpha X$ depends on x, but if X is not empty, then $\bigsqcup_\alpha X$ is independent of x. In particular, we may use the notation $\bigsqcup_\alpha X$ without specifying the element x whenever X is not empty and $z =_\beta z'$ holds for all $z, z' \in X$ and $\beta < \alpha$.

We note that a short description of $\bigsqcup_\alpha X$ is $\bigvee(X|_\alpha \cup \{\bar{x}\})$, where $\bar{x}$ is the least element of $(x]_\alpha$.

Proposition 6. *Suppose that L is a strong model. Let I be an arbitrary non-empty index set and $x_{i,n} \in L$ for all $i \in I$ and $n \geq 0$. Suppose that $\alpha < \kappa$ and $x_{i,n} \sqsubseteq_\alpha x_{i,n+1}$ for all $i \in I$ and $n \geq 0$. Then $\bigvee_{i\in I} \bigsqcup_\alpha \{x_{i,n} : n \geq 0\} =_\alpha \bigsqcup_\alpha \{\bigvee_{i\in I} x_{i,n} : n \geq 0\}$.*

Proof. First note that $\bigsqcup_\alpha \{\bigvee_{i\in I} x_{i,n} : n \geq 0\}$ exists, since by Lemma 12 we have $\bigvee_{i\in I} x_{i,n} \sqsubseteq_\alpha \bigvee_{i\in I} x_{i,n+1}$ for all $n \geq 0$, hence $\bigvee_{i\in I} x_{i,n} =_\beta \bigvee_{i\in I} x_{i,n+1}$ for all $n \geq 0$ and $\beta < \alpha$.

Again, we assume that L is an inverse limit model. A routine calculation shows that both sides of the required equality are equal to $\bigvee_{i\in I, n\geq 0}(x_{i,n})_\alpha$, where for each $i \in I$ and $n \geq 0$, $(x_{i,n})_\alpha$ is the α-component of $x_{i,n}$. □

Actually the above fact extends to all nonempty chains.

Suppose that L is model satisfying A1–A6. Following [6], we define the relation $\sqsubseteq$ on L by $x \sqsubseteq y$ iff $x = y$, or there is some $\alpha < \kappa$ with $x \sqsubset_\alpha y$, i.e., $x \sqsubseteq_\alpha y$ but $y \not\sqsubseteq_\alpha x$. When L is an inverse limit model and $x = (x_\alpha)_{\alpha<\kappa}$, $y = (y_\alpha)_{\alpha<\kappa}$, this gives $x \sqsubseteq y$ iff either $x = y$, i.e., $x_\alpha = y_\alpha$ for all $\alpha < \kappa$, or there is some $\alpha < \kappa$ with $x_\alpha < y_\alpha$ and $x_\beta = y_\beta$ for all $\beta < \alpha$.

Lemma 21. *For every model L satisfying A1–A6, the relation $\sqsubseteq$ is a partial order. Moreover, for every $x, y \in L$, if $x \leq y$ then $x \sqsubseteq y$.*

Proof. Let L be the model determined by the limit of an inverse system L_α, $\alpha < \kappa$, of complete lattices with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Let $x = (x_\alpha)_{\alpha<\kappa}$ and $y = (y_\alpha)_{\alpha<\kappa}$ in L. If $x = y$ then clearly $x \sqsubseteq y$. Suppose that $x < y$. Then there is some α with $x_\alpha < y_\alpha$ and $x_\beta = y_\beta$ for all $\beta < \alpha$. Thus, $x \sqsubset_\alpha y$ and $x \sqsubset y$.

It is clear $\sqsubseteq$ is reflexive and transitive. To prove that it is anti-symmetric, let x, y in L. Suppose that $x \sqsubseteq y$ and $y \sqsubseteq x$. If $x \neq y$ then there exist $\alpha, \beta < \kappa$ such that $x \sqsubset_\alpha y$ and $y \sqsubset_\beta x$. Then $x =_\gamma y$ for all $\gamma < \max\{\alpha, \beta\}$, which implies that $\alpha = \beta$ and hence $x_\alpha < y_\alpha$ and $y_\alpha < x_\alpha$, a contradiction. Thus $x = y$. We note that when each L_α is linearly ordered, then $\sqsubseteq$ is a linear ordering of L. □

Note that on inverse limit models, $\sqsubseteq$ is the lexicographic order. In [5], it is shown that it is necessary that the projections be locally completely additive in order to have the result of Lemma 21.

Below we will often make use of the following observation. Let L be the model determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, where $\beta < \alpha < \kappa$. Suppose that $\alpha < \kappa$ and $(x_\beta)_{\beta<\alpha}$ is a (partial) compatible sequence, so that $h^\beta_\gamma(x_\beta) = x_\gamma$ for all $\gamma < \beta < \alpha$. Then there is a least element x_α of L_α such that the sequence $(x_\beta)_{\beta\leq\alpha}$ is still compatible, namely $x_\alpha = \bigvee_{\beta<\alpha} k^\alpha_\beta(x_\beta)$. Moreover, the set of all elements x_α with this property is a complete sublattice of L_α which is a closed interval. Indeed, if Y is a nonempty set of such elements of L_α, then so is $\bigvee Y$, since $h^\alpha_\beta(Y) = \{x_\beta\}$ and thus $h^\alpha_\beta(\bigvee Y) = \bigvee h^\alpha_\beta(Y) = x_\beta$ for all $\beta < \alpha$. Finally, if x_α and x'_α in L_α satisfy $h^\alpha_\beta(x_\alpha) = h^\alpha_\beta(x'_\alpha) = x_\beta$ for all $\beta < \alpha$, and if $x_\alpha \leq y \leq x'_\alpha$, then by $h^\alpha_\beta(x_\alpha) \leq h^\alpha_\beta(y) \leq h^\alpha_\beta(x'_\alpha)$ we must have $h^\alpha_\beta(y) = x_\beta$ for all $\beta < \alpha$.

Suppose that $f : L \to L$, where L is a model. Following [6], we say that f is α-*monotone* for some $\alpha < \kappa$ if $x \sqsubseteq_\alpha y$ implies $f(x) \sqsubseteq_\alpha f(y)$ for all $x, y \in L$. When L is an inverse limit model as above, this means that if $x, y \in L$ are such that for each $\beta < \alpha$, the β-component of x agrees with the corresponding component of y and the α-component of x is less than or equal to the corresponding component of y, then the same hold for $f(x)$ and $f(y)$. Call a function $g : L_\alpha \to L_\alpha$ *conditionally monotone* if for all $x, y \in L_\alpha$, if $h^\alpha_\beta(x) = h^\alpha_\beta(y)$ for all $\beta < \alpha$ and $x \leq y$, then $g(x) \leq g(y)$.

Lemma 22. *Suppose that L is a model determined by an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Let $f : L \to L$. Then f is α-monotone for all $\alpha < \kappa$ iff there exist conditionally monotone functions $f_\alpha : L_\alpha \to L_\alpha$, $\alpha < \kappa$, such that $f((x_\alpha)_{\alpha<\kappa}) = (f_\alpha(x_\alpha))_{\alpha<\kappa}$ for all $(x_\alpha)_{\alpha<\kappa}$ in L.*

Proof. In order to prove the sufficiency part of the lemma, suppose that $f : L \to L$ and f_γ, $\gamma < \kappa$, is a family of conditionally monotone functions such that $f(x) = (f_\gamma(x_\gamma))_{\gamma<\kappa}$ for all $x = (x_\gamma)_{\gamma<\kappa} \in L$. Let $\alpha < \kappa$ and $x, y \in L$ with $x \sqsubseteq_\alpha y$. Suppose that $x = (x_\gamma)_{\gamma<\kappa}$ and $y = (y_\gamma)_{\gamma<\kappa}$. We want to prove that $f(x) = x' \sqsubseteq_\alpha y' = f(y)$. But for all $\beta < \alpha$, the β-component x'_β of x' agrees with the β-component y'_β of y', since by $x_\beta = y_\beta$ we have $x'_\beta = f_\beta(x_\beta) = f_\beta(y_\beta) = y'_\beta$.

Also, since $x_\alpha \leq y_\alpha$ and f_α is conditionally monotone, for the α-components we have $x'_\alpha = f_\alpha(x_\alpha) \leq f_\alpha(y_\alpha) = y'_\alpha$.

In order to prove the necessity part of the lemma, suppose that f is α-monotone for all $\alpha < \kappa$. For each $\alpha < \kappa$, define $f_\alpha : L_\alpha \to L_\alpha$ as the function $h_\alpha^\infty \circ f \circ k_\alpha^\infty$. If $x \leq y$ in L_α with $h_\beta^\alpha(x) = h_\beta^\alpha(y)$ for all $\beta < \alpha$, then for all $\beta < \alpha$, the β-component of $k_\alpha^\infty(x)$ agrees with the β-component of $k_\alpha^\infty(y)$, while the α-component of $k_\alpha^\infty(x)$ is x and the α-component of $k_\alpha^\infty(y)$ is y, so that the α-component of $k_\alpha^\infty(x)$ is less than or equal to the α-component of $k_\alpha^\infty(y)$. Since f is α-monotone, the same holds for $f(k_\alpha^\infty(x))$ and $f(k_\alpha^\infty(y))$. In particular, the α-component of $f(k_\alpha^\infty(x))$ is less than or equal to the α-component of $f(k_\alpha^\infty(y))$, i.e., $f_\alpha(x) = h_\alpha^\infty(f(k_\alpha^\infty(x))) \leq h_\alpha^\infty(f(k_\alpha^\infty(y))) = f_\alpha(y)$.

We still need to prove that $f(x) = (f_\alpha(x_\alpha))_{\alpha<\kappa}$ for all $x = (x_\alpha)_{\alpha<\kappa}$ in L. Let $\alpha < \kappa$ be a fixed ordinal. Since f is α-monotone and $x =_\alpha k_\alpha^\infty(x_\alpha)$, also $f(x) =_\alpha f(k_\alpha^\infty(x))$, hence the α-component of $f(x)$ agrees with the α-component of $f(k_\alpha^\infty(x_\alpha))$, which is in turn equal to $f_\alpha(x_\alpha)$. Since α was an arbitrary ordinal less than κ, this proves the required equality. □

In particular, when f is α-monotone for all $\alpha < \kappa$, then f_0 is a monotone function over L_0.

A function $L \to L$ which is α-monotone for all $\alpha < \kappa$ need not be monotone w.r.t. the partial order $\sqsubseteq$, cf. [6].

Remark 5. Thus, if L is an inverse limit model as above and $f : L \to L$ is α-monotone for all $\alpha < \kappa$, then f determines and is determined by a necessarily unique family of conditionally monotone functions $f_\alpha : L_\alpha \to L_\alpha$, $\alpha < \kappa$. Moreover, this family of functions is compatible in the sense that $h_\beta^\alpha \circ f_\alpha = f_\beta \circ h_\beta^\alpha$ for all $\beta < \alpha < \kappa$.

Conversely, if f_α, $\alpha < \kappa$, is a compatible sequence of conditionally monotone functions, then for each compatible sequence $x = (x_\alpha)_{\alpha<\kappa}$, the sequence $(f_\alpha(x_\alpha))_{\alpha<\kappa}$ is also compatible, and the function $f : L \to L$ defined by $f(x) = (f_\alpha(x_\alpha))_{\alpha<\kappa}$ for all $x = (x_\alpha)_{\alpha<\kappa}$ in L is α-monotone for all $\alpha < \kappa$.

We will also use the following fact. Suppose that L is an inverse limit model as above and $f : L \to L$ is α-monotone for all $\alpha < \kappa$. Suppose that $(x_\beta)_{\beta<\alpha}$ is a compatible sequence, so that $h_\gamma^\beta(x_\beta) = x_\gamma$ for all $\gamma < \beta < \alpha$. Consider the sublattice Z_α of L_α of those elements x_α such that the sequence $(x_\beta)_{\beta\leq\alpha}$ is still compatible. If for each $\beta < \alpha$, x_β is a fixed point of f_β, see below, then f_α maps Z_α into itself and is monotone on Z_α.

Recall that a *pre-fixed point* (resp. *post-fixed point*) of a function f over a partially ordered set P is an element $x \in P$ with $f(x) \leq x$ (resp. $x \leq f(x)$). Moreover, x is a *fixed point* of f if $f(x) = x$, i.e., when x is both a pre-fixed point and a post-fixed point. By the well-known Knaster-Tarski fixed point theorem [2], every monotone endofunction over a complete lattice has a least fixed point which is also the least pre-fixed point. Dually, every monotone endofunction over a complete lattice has a greatest fixed point, which is also the greatest post-fixed point. And if L is a complete lattice and $f : L \to L$ is monotone, then the fixed points of f form a complete lattice. This immediately follows from the existence of the least fixed point using the fact that if x is a post-fixed point, then there is a least pre-fixed point over x which is a fixed point. More generally, if X is a set of post-fixed points, then there is a least pre-fixed point over X which is a fixed point. Of course, the dual statement also holds.

In order to prove the above claim, suppose that L is a complete lattice, $f : L \to L$ is monotone, and X is a set of post-fixed points of f. Let $Z = \{z \in L : X \leq z,\ f(z) \leq z\}$ and $y = \bigwedge Z$. We need to prove that y is a fixed point of f.

We have $X \leq y$ and thus $f(X) \leq f(y)$, hence $X \leq f(y)$ since X is a set of post fixed points. And if $z \in Z$ then $y \leq z$, hence $f(y) \leq f(z) \leq z$. Since this holds for all $z \in Z$ and $y = \bigwedge Z$, we conclude that $f(y) \leq y$. But then $f(y) \in Z$ and thus $y \leq f(y)$, proving $f(y) = y$.

Theorem 3. *Let L be a model satisfying the axioms A1–A6 and $f : L \to L$ be α-monotone for all $\alpha < \kappa$. Suppose that $X \subseteq L$ is a set of post-fixed points of f with respect to the ordering $\leq$. Then there is a (necessarily unique) $y \in L$ with the following properties:*

(i) $X \sqsubseteq y$ and $f(y) = y$,
(ii) for all $z \in L$, if $X \sqsubseteq z$ and $f(z) \sqsubseteq z$, then $y \sqsubseteq z$.

Proof. Without loss of generality we may assume that L is the model determined by the limit of an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$, $\beta < \alpha < \kappa$. Since f is α-monotone for all $\alpha < \kappa$, it is determined by a family of conditionally monotone functions $f_\alpha : L_\alpha \to L_\alpha$, $\alpha < \kappa$.

For each $\alpha < \kappa$, let X_α denote the set of all α-components x_α of the elements x of X. Define $Y_\alpha = \{z \in X_\alpha : \forall \beta < \alpha\ h^\alpha_\beta(z) = y_\beta\}$, and let y_α be the least (pre-)fixed point of f_α over Y_α in Z_α, where Z_α is the set of all elements z of L_α with $h^\alpha_\beta(z) = y_\beta$ for all $\beta < \alpha$. In particular, $Y_0 = X_0$ and y_0 is the least (pre-)fixed point of f_0 in $Z_0 = L_0$.

It is clear that the sequence $y = (y_\alpha)_{\alpha<\kappa}$ is in L. Moreover, $f(y) = y$, as each y_α is a fixed point of f_α. The fact that $X \sqsubseteq y$ follows from the following:

Claim. For all $x \in X$ and $\alpha < \kappa$, either $x_\beta = y_\beta$ for all $\beta < \alpha$, or there is some $\beta \leq \alpha$ with $x_\beta < y_\beta$.

Indeed, if $x_\alpha \in Y_\alpha$ for all $\alpha < \kappa$, then $x_\alpha = y_\alpha$ for all $\alpha < \kappa$. In the opposite case there is a least α with $x_\alpha \notin Y_\alpha$. Then $\alpha > 0$, and $x_\beta \in Y_\beta$ for all $\beta < \alpha$. Hence, if $\beta < \alpha$, then $x_\gamma = y_\gamma$ for all $\gamma < \beta$, showing that α is not a limit ordinal. Thus, α is successor ordinal, say $\alpha = \beta + 1$. Moreover, $x_\beta \in Y_\beta$ and $x_\alpha \notin Y_\alpha$. This implies that $x_\beta < y_\beta$ and $x_\gamma = y_\gamma$ for all $\gamma < \beta$, so that $x \sqsubset_\beta y$.

Claim. Let $z = (z_\alpha)_{\alpha<\kappa} \in L$ with $X \sqsubseteq z$ and $f(z) \sqsubseteq z$. Then for all $\alpha < \kappa$, either $y_\beta = z_\beta$ for all $\beta < \alpha$, or there is some $\beta < \alpha$ with $y_\beta < z_\beta$.

Indeed, suppose that $\alpha < \kappa$ and the claim holds for all ordinals less than α. If $y_\beta < z_\beta$ for some $\beta < \alpha$ then we are done. Suppose now that $y_\beta = z_\beta$ for all

$\beta < \alpha$. Then $f_\beta(z_\beta) = f_\beta(y_\beta) = y_\beta = z_\beta$ for all $\beta < \alpha$. Thus, if Y_α is empty, then y_α is the least (pre-)fixed point of f_α in Z_α, whereas z_α is another pre-fixed point of f_α in Z_α. Hence $y_\alpha \le z_\alpha$. Suppose now that Y_α is not empty. Then y_α is the least pre-fixed point of f_α in Z_α above Y_α, while z_α is another such pre-fixed point, since by $f(z) \sqsubseteq z$, $X \sqsubseteq z$ and $f_\beta(z_\beta) = z_\beta$ and $y_\beta = z_\beta$ for all $\beta < \alpha$ we have $f_\alpha(z_\alpha) \le z_\alpha$ and $Y_\alpha \le z_\alpha$. We conclude that $y_\alpha \le z_\alpha$. It follows now that $y \sqsubseteq z$ whenever $X \sqsubseteq z$ and $f(z) \sqsubseteq z$. □

Corollary 18. *Let L be a model satisfying the axioms A1–A6 and $f : L \to L$ be α-monotone for all $\alpha < \kappa$. Suppose that $X \subseteq L$ is a set of pre-fixed points of f with respect to the ordering $\le$. Then there is a (necessarily unique) $y \in L$ with the following properties:*

(i) $y \sqsubseteq X$ and $f(y) = y$,
(ii) for all $z \in L$, if $z \sqsubseteq X$ and $z \sqsubseteq f(z)$, then $z \sqsubseteq y$.

Proof. Again, we may assume that L is a limit model. Using the notation introduced in the previous proof, for each $\alpha < \kappa$ define $Y_\alpha = \{x \in X_\alpha : \forall \beta < \alpha\ h^\alpha_\beta(x) = y_\beta\}$ and let y_α be the greatest (post-)fixed point of f_α below Y_α in Z_α, where Z_α is the set of all elements z of L_α with $h^\alpha_\beta(z) = y_\beta$ for all $\beta < \alpha$. Then $y = (y_\alpha)_{\alpha<\kappa}$ is the required element of L. □

Corollary 19. *Suppose that L is a model and $f : L \to L$ is α-monotone for all $\alpha < \kappa$. Then the fixed points of f form a complete lattice with respect to the ordering $\sqsubseteq$.*

Corollary 20. *For every model L satisfying the axioms A1–A6, $(L, \sqsubseteq)$ is a complete lattice.*

Proof. Let f be the identity function in Corollary 19. In particular, we obtain that if $X \subseteq L$, then the supremum $\bigsqcup X$ of X w.r.t. the ordering $\sqsubseteq$ can be constructed as follows. For each $\alpha < \kappa$, define $Y_\alpha = \{x \in X_\alpha : \forall \beta < \alpha\ h^\alpha_\beta(x) = y_\beta\}$ and let y_α be the supremum of Y_α and the least element of Z_α in the complete lattice L_α (or in Z_α). Then $\bigsqcup X = (y_\alpha)_{\alpha<\kappa}$. Note that if Y_α is empty, then $y_\alpha = \bigvee_{\alpha<\kappa} k^\alpha_\beta(y_\beta)$.

The infimum $\bigsqcap X$ can be constructed dually. □

Corollary 21. *Let L be a model satisfying the axioms A1–A6 and suppose that $f : L \to L$ is α-monotone for all $\alpha < \kappa$. Then f has a least pre-fixed point w.r.t. the ordering $\sqsubseteq$ which is a fixed point. Hence, if x is the least fixed point of f and $f(y) \sqsubseteq y$, then $x \sqsubseteq y$.*

Remark 6. Suppose that L is a model and $f : L \to L$ is α-monotone for all $\alpha < \kappa$. Let x denote the least (pre-)fixed point of f w.r.t. $\sqsubseteq$. If $f(z) \le z$ for some $z \in L$, then also $f(z) \sqsubseteq z$, hence $x \sqsubseteq z$.

Example 3. Suppose that Z is a denumerable set of propositional variables and P is an at most countably infinite propositional logic program over Z, possibly involving negation. Thus P is a countable set of instructions of the form $z \leftarrow \ell_1 \wedge \cdots \wedge \ell_k$, where $z \in Z$ and ℓ_i is a literal for each i. Consider the model $L = V_\Omega^Z$, defined in Sect. 2, where Ω is the least uncountable ordinal. Then P induces a function $f_P : L \to L$ which maps an interpretation $I \in L$ to the interpretation $J = f_P(I)$ such that $J(z) = \bigvee_{z \leftarrow \ell_1 \wedge \cdots \wedge \ell_k \in P}(I(\ell_1) \wedge \cdots \wedge I(\ell_k))$, where for a negative literal $\ell = \neg y$, $I(\ell) = T_{\alpha+1}$ if $I(y) = F_\alpha$, $I(\ell) = F_{\alpha+1}$ if $I(y) = T_\alpha$, and $I(\ell) = 0$ if $I(y) = 0$. Then f_P is α-monotone for all $\alpha < \Omega$. The semantics of P is defined in [10] as the least fixed point of f_P w.r.t. $\sqsubseteq$.

We end this section by mentioning a result from [7]. A new proof of it, based on the inverse limit representation, can be found in [5].

Theorem 4. *Suppose that L is a model satisfying A1–A6 and $f : L \to L$ is α-monotone for each $\alpha < \kappa$. Let $X \subseteq L$ be a set of post-fixed points of f w.r.t. the ordering $\leq$. Then $y = \bigsqcup X$ is also a post-fixed point of f w.r.t. $\leq$.*

9 Symmetric Models

The first two axioms A1 and A2 and the axiom A6 introduced in Sect. 2 are self dual, but the others are not. The dual forms of A3, A4 and A6 are given below. (For missing proofs, see [5].)

A3d. For all x and $\alpha < \kappa$ there exists y such that $x =_\alpha y$ and for all z, if $z \sqsubseteq_\alpha x$ then $z \leq y$. (It is clear that y is uniquely determined by x and α and we will denote it by $x|^\alpha$.)
A4d. For all $\alpha < \kappa$ and x_i, y, $i \in I$, where I is a nonempty index set, if $x_i =_\alpha y$ for all $i \in I$, then $\bigwedge_{i \in I} x_i =_\alpha y$.
A5d. For all x, y and $\alpha < \kappa$, if $x \leq y$ then $x|^\alpha \leq y|^\alpha$.

We also define the dual of A4*.

A4*d. For all $\alpha < \kappa$ and x_i, y_i with $x_i =_\alpha y_i$, $i \in I$, where I is any index set, it holds that $\bigwedge_{i \in I} x_i =_\alpha \bigwedge_{i \in I} y_i$.

Lemma 23. *There is a model not satisfying A3d.*

Lemma 24. *Every model satisfying the axioms A1–A6 satisfies A4* d.*

Lemma 25. *Every strong model satisfies A3d and A5d.*

Suppose that L is a stratified complete lattice. We say that L is a *dual model* if it satisfies A1, A2, A3d, A4d, A5d and A6. Moreover we call L a *strong dual model* if it satisfies A1, A2, A3d, A4*d, A5d and A6. Alternatively, L is a (strong) dual model iff its dual L^{op}, obtained by reversing the relation $\leq$ and each relation $\sqsubseteq_\alpha$, is a (strong) model.

Of course, if a property holds in all models, then the dual property holds in all dual models, and similarly for strong models. In particular, every (strong) dual model can be constructed as an inverse limit model. However, one uses dual projection-embedding pairs and locally infimum preserving or infimum preserving functions $h^\alpha_\beta : L_\alpha \to L_\beta$ of complete lattices. Here, when L and L' are complete lattices, we say that $g : L' \to L$ is a dual projection with corresponding dual embedding $f : L \to L'$ if f and g are monotone, $g \circ f : L \to L$ is the identity function on L, and $f \circ g : L' \to L'$ is greater than or equal to the identity function on L'. Alternatively, this means that g is a projection $(L')^{\mathrm{op}} \to L^{\mathrm{op}}$ and f is the corresponding embedding $L^{\mathrm{op}} \to (L')^{\mathrm{op}}$. And a function $h : L' \to L$ is locally infimum preserving if for all $Y \subseteq L'$ and $x \in L$ with $h(Y) = x$, it holds that $h(\bigwedge Y) = x$. This clearly means that h is locally completely additive as a mapping of $(L')^{\mathrm{op}}$ into L^{op}.

Every dual model is isomorphic to a model determined by the limit of an inverse system $h^\alpha_\beta : L_\alpha \to L_\beta$ of locally infimum preserving dual projections. Moreover, every strong dual model is determined by such an inverse system where each h^α_β is a dual projection preserving all infima. Dual models share several properties of models, e.g., each dual model L gives rise to a complete lattice $(L, \sqsubseteq)$, and if $f : L \to L$ is α-monotone for all $\alpha < \kappa$, where L is a dual model, then the set of all fixed points of f, ordered by $\sqsubseteq$, is a complete lattice.

We also define *symmetric models* which are both models and dual models. Similarly, a *strong symmetric model* is a strong model that is a strong dual model. As an immediate consequence of Lemma 24 we have:

Corollary 22. *A model is symmetric iff it satisfies A3d and A5d.*

The standard model V^Z discussed in Sect. 2 is a strong symmetric model. But a model may not be symmetric. See Lemma 23. However, we have:

Theorem 5. *The following conditions are equivalent for a model L satisfying the axioms A1–A6: (i) L is a strong model, (ii) L is a strong symmetric model, (iii) L is a symmetric model.*

Proof. By Corollary 22, Lemmas 24 and 25. □

Corollary 23. *Let L be a model determined by an inverse system of complete lattices L_α, $\alpha < \kappa$, with locally completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$. Then L is a (strong) symmetric model iff the functions h^α_β, $\beta < \alpha < \kappa$ are completely additive.*

Thus, in this case, the functions h^α_β preserve arbitrary infima and suprema.

Corollary 24. *A model is a (strong) symmetric model iff it is isomorphic to the model determined by an inverse system of complete lattices L_α, $\alpha < \kappa$, with completely additive projections $h^\alpha_\beta : L_\alpha \to L_\beta$.*

10 Conclusion

An axiomatic framework as an abstraction of the treatment of the semantics of logic programs with negation in [10] has recently been introduced in [6,7]. Here, we dealt with the models of two of the axiom systems of [6,7], and established representation theorems for them. We proved that every model can be constructed from an inverse system of complete lattices with locally completely additive projections, and that every strong model can be constructed from an inverse system of complete lattices with completely additive projections. Using the inverse limit representation, we proved that the fixed points of a weakly monotone function over a model form a complete lattice with respect to a new ordering, cf. Corollary 19. In particular, there is a least fixed point, called the stratified least fixed point.

We also studied models satisfying, together with each axiom, the dual axiom. We proved that such symmetric models are exactly the strong models, and in fact the strong symmetric models. For the future, it would be interesting to extend the representation theorem to more general classes of models introduced in [6], where the preorderings $\sqsubseteq_\alpha$ are not completely determined by the ordering $\leq$ and the equivalence relations $=_\alpha$.

Since the semantics of recursive definitions is usually captured by fixed points of functions, or functors, or other constructors, fixed point operations appear in almost all branches of computer science including automata and languages, semantics, concurrency, programming logics, the characterization of complexity classes using formal logic, etc. Our aim with this paper and its predecessors has been to contribute to the development of a novel general framework for solving fixed point equations involving non-monotone operations as an alternative of the bilattice based approach [3,8]. This method has already found applications in logic programming and Boolean context-free grammars, and it appears to be applicable in other situations including Boolean automata, fuzzy sets, and quantitative logics. A nice feature of the approach is that the stratified least fixed point operation over weakly monotonic functions also satisfies the standard equational laws, cf. [4].

References

1. Charalambidis, A., Ésik, Z., Rondogiannis, P.: Minimum model semantics for extensional higher-order logic programming with negation. Theor. Pract. Logic Programm. **14**, 725–737 (2014)
2. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order, 2nd edn. Cambridge University Press, New York (2002)
3. Denecker, M., Marek, V.W., Truszczyński, M.: Approximations, stable operations, well-founded fixed points and applications in nonmonotonic reasoning. In: Minker, J. (ed.) Logic-Based Artificial Intelligence, pp. 127–144. Kluwer, Norwell (2000)
4. Ésik, Z.: Equational properties of stratified least fixed points (Extended Abstract). In: Paiva, V., Queiroz, R., Moss, L.S., Leivant, D., Oliveira, A.G. (eds.) WoLLIC 2015. LNCS, vol. 9160, pp. 174–188. Springer, Heidelberg (2015). doi:10.1007/978-3-662-47709-0_13

5. Ésik, Z.: A representation theorem for stratified complete lattices, arXiv:1503.05124 (2015)
6. Ésik, Z., Rondogiannis, P.: A fixed-point theorem for non-monotonic functions. Theoret. Comput. Sci. **574**, 18–38 (2015)
7. Ésik, Z., Rondogiannis, P.: Theorems on pre-fixed points of non-monotonic functions with applications in logic programming and formal grammars. In: Kohlenbach, U., Barceló, P., Queiroz, R. (eds.) WoLLIC 2014. LNCS, vol. 8652, pp. 166–180. Springer, Heidelberg (2014). doi:10.1007/978-3-662-44145-9_12
8. Fitting, M.: Fixed point semantics for logic programming a survey. Theoret. Comput. Sci. **278**, 25–51 (2002)
9. Gierz, G., Hoffman, K.H., Keimel, K., Lawson, J.D., Mislove, M., Scott, D.S.: Continuous Lattices and Domains. Cambridge University Press, Cambridge (2003)
10. Rondogiannis, R., Wadge, W.W.: Minimum model semantics for logic programs with negation. ACM Trans. Comput. Logic **6**, 441–467 (2005)
11. Scott, D.: Continuous lattices. In: Lawvere, F.W. (ed.) Toposes, Algebraic Geometry and Logic. LNM, vol. 274, pp. 97–136. Springer, Heidelberg (1972). doi:10.1007/BFb0073967

Equilibrium Semantics for IF Logic and Many-Valued Connectives

Christian G. Fermüller[1(✉)] and Ondrej Majer[2]

[1] Theory and Logic Group, Vienna University of Technology, Vienna, Austria
chrisf@logic.at

[2] Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic
majer@flu.cas.cz

Abstract. We connect two different forms of game based semantics: Hintikka's game for Independence Friendly logic (IF logic) and Giles's game for Łukasiewicz logic. An interpretation of truth values in $[0, 1]$ as equilibrium values in semantic games of imperfect information emerges for a logic that extends both, Łukasiewicz logic and IF logic. We prove that already on the propositional level all rational truth values can be obtained as equilibrium values.

Keywords: Game semantics · IF logic · Fuzzy logic · Łukasiwicz logic · Giles's game

1 Introduction

Already in the 1960s Jaakko Hintikka [12] introduced a game based characterization of Tarski's central semantic notion of 'truth in a model'. The game features moves by two antagonistic players, one in the role of the verifier or proponent of a formula, the other one in the role of the falsifier or opponent. The game proceeds according to the outermost connective or quantifier of the formula currently at stake: disjunction and existential quantification trigger a move by the proponent, while conjunction and universal quantification elicit a move by the opponent; negation corresponds to a role switch. In this manner the formula currently in focus is replaced by one of its immediate sub-formulas in every round. At the atomic level a given model determines who won the actual run of the game. Other connectives, in particular implication, could be defined from the mentioned ones in classical logic, of course, but it is an important observation

C.G. Fermüller—Supported by Austrian Science Foundation (FWF) grant I1897-N25 (MoVaQ-MFL).

O. Majer—Supported by Czech Science Foundation grant GF15-34650L.

The original version of this chapter was revised. The affiliation and grant number of the second author have been corrected. The erratum to this chapter is available at DOI: 10.1007/978-3-662-54332-0_19

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 290–312, 2017.
DOI: 10.1007/978-3-662-54332-0_16

for our current endeavor that such a reduction of a richer set of connectives to just (this form of) conjunction, disjunction, and negation is no longer available in general, once we move on to a non-classical setting.

Hintikka's game-theoretic semantics deploys its full capacity when we consider imperfect information: the players may not be fully informed about previous moves during a run of the game. In particular *Independence Friendly logic* (IF logic) results from attaching "slash sets" to the quantifiers, containing those variables for which the current player is ignorant of corresponding assignments of domain elements that result from previous moves. For example $\forall x \exists y/\{x\} y = x$ corresponds to a game, where first the opponent chooses an arbitrary domain element for the variable x and then the proponent has to choose an element for y without knowing which element has been picked for x. This entails that (in contrast to the game for $\forall x \exists y \, y = x$) the opponent has no winning strategy if there are two or more domain elements. The fact that games for IF formulas are not determined in general leads to *equilibrium semantics* [18,21], which arises if one considers mixed Nash equilibria for corresponding strategic games. Indeed, if we identify losing a game with payoff 0 and winning with payoff 1, we may associate a unique value $v \in [0, 1]$ to every IF formula φ[1] with respect to any given finite model, such that v is the expected payoff for the proponent of φ in the corresponding game when the players employ mixed strategies that are in equilibrium (i.e., neither player has an incentive to unilaterally change her strategy).

Motivated by the challenge to model reasoning in physics, Robin Giles developed another game based approach to logic in the 1970s [7,8]. Giles was seemingly unaware of Hintikka's game-theoretic semantics, but referred to Paul Lorenzen's attempts to justify intuitionistic logic in terms of an idealized dialogue between a proponent and an opponent of a given formula. Giles's game consists of two components: first, the players stepwise reduce logically complex formulas to their sub-formulas, similar as in Hintikka's game, but not bound by the restriction that at any state of the game only a single formula is asserted by the proponent and attacked by the opponent. Rather a whole multiset, called tenet, of formulas is asserted by each of the players at any given state. The second stage of the game commences when only atomic formulas are left in both players' tenets. For each atomic assertion a corresponding experiment is performed. If the experiment fails, then the asserting player has to pay 1€ (one Euro) to the other player. If for every given atomic formula the corresponding experiment either always fails or always succeeds then Giles's game leads to an alternative characterization of classical logic. However, Giles stipulates that any experiment may be dispersive: it may yield different results when repeated—only a specific failure probability (risk value) is known for each experiment. A player's payoff at the final state of Giles's game is identified with the *expected* amount of money that she has to pay minus the expected amount that she receives from her opponent. Giles proved that

[1] In fact there are certain complications if one admits formulas corresponding to games where a player may not have access to her own previous moves. We will circumvent these problems by insisting on perfect recall. Moreover, we follow [18,21] in moving negations to the atomic level.

the payoffs enforceable by optimal strategies correspond to the (inverse of) truth values resulting from evaluating the initial formula according to the truth functions for Łukasiewicz logic Ł.

At first sight, Hintikka's and Giles's games seem to serve different purposes and moreover are quite different in detail as well as in their overall structure. Nevertheless we propose a combination of the two games that corresponds to a rather expressive logic, which we shall call Ł(IF). The formulas of Ł(IF) are two-tiered: they can be thought of as formulas of Ł where the atomic formulas are replaced by (arbitrarily complex) IF formulas. Accordingly, the combined game proceeds in two stages: first, Giles's game is played until only IF formulas occur in the players' tenets and then an instance of Hintikka's game is employed as a dispersive experiment (in the sense of Giles) for each IF formula. The overall evaluation is like for Giles's game. In this setting intermediate truth values turn out to correspond to equilibrium values for IF formulas that in turn may be combined to yield truth values for formulas of an expressive many-valued logic. The adequateness of the combined game for Ł(IF) emerges as a corollary to the adequateness of Giles's and Hintikka's games for Ł and for IF logic, respectively. The achieved gain arises on a conceptual level in two different directions:

(1) Skeptics of many-valued logics rightfully challenge their defenders by asking for an explanation of intermediate truth values and of corresponding truth functions in terms of first principles about reasoning. Giles's game only provides a partial answer by replacing classical (bivalent) interpretations with assignments of risk values (probabilities) to atomic formulas. Our combined game can be understood as an explanation of risk values as equilibrium values, arising from evaluations with respect to classical interpretations under imperfect information.
(2) Equilibrium semantics for IF logic supports a many-valued interpretation of disjunction and conjunction as maximum and minimum, respectively. Somewhat indirectly also the truth function $1-x$ for negation is justified. However the more general format of Giles's game is needed to interpret the considerably richer set of connectives (including implication, strong conjunction, and strong disjunction) of Łukasiewicz logic. From this perspective IF logic provides only a limited way of modeling the effects of imperfect information. At least some of these limitations are lifted in Ł(IF). For example, simple schematic Ł(IF) formulas (but not IF formulas) express that instances of Hintikka's game are always constant-sum, but not determined in general.

In the light of item 1 it is important to note that indeed *all* rationals in $[0,1]$ can be obtained as equilibrium values [21]. (For the related framework of Dependence Logic a similar result is shown in [6].) The corresponding constructions involve first-order formulas and particular models. This triggers the question whether one can obtain all rational truth values already on the propositional level. We provide a positive answer by showing that for every rational $r \in [0,1]$ there is constant propositional IF formula φ_r with equilibrium value r. (Propositional IF formulas arise from classical propositional formulas if there is imperfect information about the choice of conjuncts and disjuncts in Hintikkas game. The formula is constant if it is built up from the atomic formulas $\bot$ and $\top$ only.) In fact we will present two constructions: a simpler one for IF formulas with n-ary conjunction and disjunction for any $n \geq 2$ and a more involved one for ordinary binary connectives.

The rest of the paper is organized as follows. Section 2 reviews Hintikka's game for classical logic and moves on to explain equilibrium logic for (a particular type of) IF formulas. Section 3 is devoted to Giles's game for Łukasiewicz logic. The logic Ł(IF) and the corresponding combination the two types of semantic games is introduced in Sect. 4. The announced results regarding the realization of all rationals as equilibrium values are the topic of Sect. 5. We conclude in Sect. 6 with a short summary and some hints on directions for further research.

2 Hintikka's Game

Let us revisit Hintikka's game-theoretic semantics (cf. [12,13]). We will call the game that characterizes truth in a (classical) model the $\mathcal{H}$*-game*. There are two players, say *I* and *you*, who are either in the role of the *Proponent* **P** or the *Opponent* **O**[2]. Initially I am **P** and you are **O**. At each state of $\mathcal{H}$-game the player in role **P** seeks to defend the claim that a certain formula is true in a given model $\mathcal{I}$ under a given variable assignment ξ, while the player in role **O** aims at refuting this claim. We will use $D_{\mathcal{I}}$ to denote the domain of $\mathcal{I}$. Formulas are built up as usual from atomic formulas, including equalities, as well as $\top$ and $\bot$, using the propositional connectives $\wedge$, $\vee$, $\neg$, and the quantifiers $\forall$ and $\exists$. The game rules are symmetric in the sense that we only need to refer to the roles **P** and **O**, but not to the identity of the players. The formula φ together with the variable assignment ξ that is at stake at a given state is called the *current (augmented) formula*.[3] We will also say that **P** asserts the current formula $\varphi[\xi]$, while **O** attacks it.

($R^{\mathcal{H}}_{\wedge}$) If the current formula is $(\varphi \wedge \psi)[\xi]$, then **O** chooses whether the game continues with **P**'s assertion of $\varphi[\xi]$ or of $\psi[\xi]$.
($R^{\mathcal{H}}_{\vee}$) If the current formula is $(\varphi \vee \psi)[\xi]$, then **P** chooses whether the game continues with **P**'s assertion of $\varphi[\xi]$ or of $\psi[\xi]$.
($R^{\mathcal{H}}_{\neg}$) If the current formula is $\neg\varphi[\xi]$, then game continues with **P**'s assertion of $\varphi[\xi]$, except that the roles of the players are switched (i.e., **P** now is the player that attacked $\neg\varphi[\xi]$).
($R^{\mathcal{H}}_{\forall}$) If the current formula is $(\forall x\varphi)[\xi]$ then **O** chooses a $c \in D_{\mathcal{I}}$ and the game continues with **P**'s assertion of $\varphi[\xi[c/x]]$[4].

[2] Hintikka uses *Myself* and *Nature* as names for the players and *Verfier* and *Falisifer* for the two roles.

[3] It is more customary to attach the variable assignment to the interpretation instead of to the formula that is to evaluated. For the $\mathcal{H}$-game this does not make any difference. However we will later introduce games, where several formulas are to be evaluated over the same interpretation, but each with respect to a (possibly) different variable assignment.

[4] $\xi[c/x]$ denotes the variable assignment that is like ξ, except for assigning c to x.

($R^{\mathcal{H}}_{\exists}$) If the current formula is $\exists x\varphi[\xi]$ then **P** chooses a $c \in D_{\mathcal{I}}$ and the game continues with **P**'s assertion of $\varphi[\xi[c/x]]$.
($R^{\mathcal{H}}_{at}$) If the current formula is an atomic formula $A[\xi]$ then the game ends. **P** wins if A is true in $\mathcal{I}$ under assignment ξ, otherwise **O** wins.

We speak of the *$\mathcal{H}$-game for $\varphi[\xi]$ with respect to $\mathcal{I}$*, if the game starts with the augmented formula $\varphi[\xi]$. The adequateness of this game for classical logic is expressed as follows.

Theorem 1 (Hintikka). *I have a winning strategy in the $\mathcal{H}$-game for $\varphi[\xi]$ with respect to $\mathcal{I}$ iff φ is true in $\mathcal{I}$ under assignment ξ.*

Above, we have tacitly assumed that the players of the $\mathcal{H}$-game have perfect, complete and common knowledge. This means that they share knowledge not only about the rules, but also about all previous moves at each state of an instance of the game. A whole new branch of logic, called *I*ndependence *F*riendly logic (*IF* logic) arises by investigating the consequences of imperfect knowledge in the $\mathcal{H}$-game. Following [18], formulas of IF-logic are defined as follows.

Definition 1. *We fix a language with an infinite supply of constants and predicate symbols. Terms of the language are either constants or variables.*

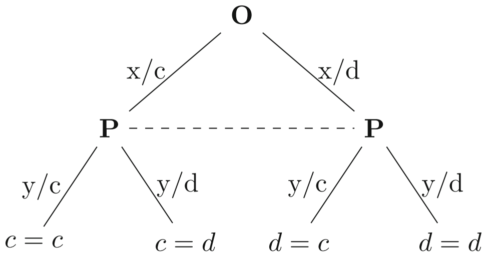
- $\top$ *and* $\bot$ *are IF formulas.*
- *If s and t are terms, then $s = t$ and $\neg(s = t)$, henceforth written as $s \neq t$, are IF formulas.*
- *If P is an n-ary predicate symbol and $t_1, \ldots, t_n$ are terms, then $P(t_1, \ldots, t_n)$ and $\neg P(t_1, \ldots, t_n)$ are IF formulas.*
- *If φ and ψ are IF formulas, then $\varphi \wedge \psi$ and $\varphi \vee \psi$ are IF formulas.*
- *If φ is an IF formula, x a variable, and W a finite set of variables not containing x, then $(\exists x/W)\varphi$ and $(\forall x/W)\varphi$ are IF formulas, where φ is called the* scope *of the exhibited quantifier occurrence and W is called a* slash set. *We abbreviate $(\exists x/\emptyset)\varphi$ by $\exists x\varphi$ and $(\forall x/\emptyset)\varphi$ by $\forall x\varphi$.*

The intended (game-theoretic) semantics of IF formulas is specified with respect to a version of the $\mathcal{H}$-game, where the players have to choose the witness elements for bound variables without knowing the choices that may have been made for the variables in the corresponding slash sets at earlier stages of the game (see [18] for details). Moreover, we modify rule $R^{\mathcal{H}}_{at}$ and let the game end when the current formula is a *literal*, i.e. either an atomic formula or a negation of an atomic formula, augmented by a variable assignment. Consequently the negation rule ($R^{\mathcal{H}}_{\neg}$) is no longer needed and hence no role change takes place during the run of an $\mathcal{H}$-game for an IF formula. Therefore we may now identify the players with their initial roles: I am **P** and you are **O** throughout every run of the game. **P** wins and **O** loses the game for an IF formula φ with respect to a given interpretation $\mathcal{I}$ if the literal with which the game ends is true in $\mathcal{I}$; otherwise **P** loses and **O** wins. We identify winning with payoff 1 and losing with payoff 0. The effect of imperfect information is rather dramatic: in contrast to the $\mathcal{H}$-game for classical formulas, it may be the case that none of the players has a winning strategy in an $\mathcal{H}$-game for an IF formula, in general. In other words the game is not determined.

Remark 1. As just discussed, negation is pushed to the atomic level in IF formulas. Accordingly, we may define the dual $\varphi^{\neg}$ of a given IF formula φ by interchanging all occurrences of $\vee$ and $\wedge$ as well as $\forall$ and $\exists$, respectively, and replacing negated atomic formulas with unnegated ones and vice versa.

Example 1. Consider the IF formula $\forall x(\exists y/\{x\})\, x = y$. If the formula is evaluated with respect to an interpretation $\mathcal{I}$ with two domain elements, then it is called MP (for *Matching Pennies*). In the $\mathcal{H}$-game for MP, **O** starts by assigning one of the elements of $\mathcal{I}$ to the variable x. In the second stage of the game **P** has to choose a domain element for y, without knowing **O**'s choice for x.

For $D_{\mathcal{I}} = \{c, d\}$ the $\mathcal{H}$-game for MP is represented by the following tree:



The dashed line between the two **P**-nodes indicates that the two nodes are in the same information set. Consequently, **P** (just like **O**) has only two possible strategies. In contrast, we obtain the (perfect information) game for the classical formula $\forall x \exists y\, x = y$, in which **P** has four possible strategies, by simply deleting the dashed line.

Because of her imperfect knowledge, **P** has no winning strategy in the $\mathcal{H}$-game for MP. Clearly **O** does not have a winning strategy either. The dual $\mathsf{MP}^{\neg} = \forall x(\exists y/\{x\})\, x \neq y$ of MP is called IMP (for *Inverse Matching Pennies*); its $\mathcal{H}$-game is undetermined as well, of course. But also, e.g., the $\mathcal{H}$-game for $\forall x(\exists y/\{x\})x = y \vee \forall u(\exists v/\{u\})v \neq u)$ is undetermined whenever $D_{\mathcal{I}}$ consists of more than one element.

Throughout the paper, we will assume that each player has *perfect recall.* This means that a player is always aware of her own previous choices in any run of the game. Moreover, each bound variable should refer to a unique quantifier occurrence. This motivates the following definition.

Definition 2. *An IF formula φ is called* recall regular *if the following conditions are satisfied:*

- *For each variable x there is at most one occurrence of $(\mathsf{Q}x/W)$ in φ, where $\mathsf{Q} \in \{\forall, \exists\}$.*
- *If $(\forall x/W)$ occurs in φ then for each $v \in W$ this occurrence is in the scope of a quantifier occurrence of the form $(\exists v/V)$.*

– *If* $(\exists x/W)$ *occurs in* φ *then for each* $v \in W$ *this occurrence is in the scope of a quantifier occurrence of the form* $(\forall v/V)$*.*

Note that all formulas in Example 1 are recall regular. In the rest of the paper all IF formulas are assumed to be recall regular, even when not stated explicitly.

We will restrict attention to finite models. Consequently the $\mathcal{H}$-game is always finite. While the $\mathcal{H}$-game is presented as an extensive game, we may as well consider its *strategic form* and will simply speak of the *strategic* $\mathcal{H}$-game for a given formula and (finite) interpretation.

Example 2. Consider an interpretation $\mathcal{I}$, where $D_{\mathcal{I}} = \{c, d\}$. Then the strategic $\mathcal{H}$-game for $\mathsf{MP} = \forall x(\exists y/\{x\})\, x = y$ corresponding to the extensive $\mathcal{H}$-game depicted in Example 1 is given by the following payoff matrix:

$$\begin{array}{c} \\ x/c \\ x/d \end{array} \begin{array}{c} \begin{array}{cc} y/c & y/d \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

The matrix entries denote the payoff for **P**, where **O** chooses a row, while **P** chooses a column. Since the payoff for **O** is $1 - x$ whenever x is the payoff for **P**, we refrain from specifying the payoff for **O** explicitly from now on.

The payoff matrix for the strategic form of the (perfect information extensive form) $\mathcal{H}$-game for $\forall x \exists y\, x = y$ can be specified as follows:

$$\begin{array}{c} \\ x/c \\ x/d \end{array} \begin{array}{c} \begin{array}{cccc} y/cc & y/cd & y/dc & y/dd \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

where we have used $y/\xi\rho$, for $\xi, \rho \in \{c, d\}$, to denote the following strategy of **P**: "if **O** assigned c to x then assign ξ to y, otherwise assign ρ to y." Note that in contrast to the game for MP, **P** now has a strategy (namely y/cd) that guarantees her the payoff 1.

Mixed strategies for a player **X** in an extensive game come in two versions: (1) behavior strategies, where for each information set of **X**, a probability distribution over all possible moves is attached; (2) strategies that consist in a single probability distribution over all pure strategies that are available to **X** in the game. In general, only in the second case the strategies directly correspond to those of the strategic form of the game and consequently lead to unique a equilibrium value in finite, constant-sum games, like (instances of) the $\mathcal{H}$-game. However by Kuhn's Theorem [15] the two types of strategies are in one-one correspondence in games where all players have perfect recall. This justifies our focus on recall regular IF formulas.

Since the strategic $\mathcal{H}$-game is finite and constant-sum, von Neumann's Minimax Theorem can be applied to obtain the following result. (Cf. [18,21], where the term *equilibrium semantics* is introduced in this context.)

Theorem 2. *For every finite interpretation $\mathcal{I}$, every IF formula φ, and every corresponding variable assignment there is a unique value $v \in [0,1]$ such that v is the expected payoff for $\boldsymbol{P}$ and $1-v$ is the expected payoff for $\boldsymbol{O}$ under the (unique) mixed Nash equilibrium of the strategic $\mathcal{H}$-game for $\varphi[\xi]$ and $\mathcal{I}$.*

We will call the value v mentioned in Theorem 2 the *equilibrium value* of φ with respect to $\mathcal{I}$ and ξ and use $v^{eq}_{\mathcal{I}}(\varphi[\xi])$ to denote it. If φ is a closed formula then the reference to the (empty) assignment ξ is dropped.

Example 3. Let n be the cardinality of the domain of the interpretation $\mathcal{I}$. In the corresponding strategic $\mathcal{H}$-game for $\forall x(\exists y/\{x\})\, x = y$ (see Example 2) the only Nash equilibrium arises if **P** and **O** both randomize uniformly over their n strategies, which consist in picking an element of the domain of $\mathcal{I}$. The same holds for the strategic $\mathcal{H}$-game for the dual formula $\forall x(\exists y/\{x\})\, x \neq y$. Consequently $v^{eq}_{\mathcal{I}}(\forall x(\exists y/\{x\})\, x = y) = 1/n$ and $v^{eq}_{\mathcal{I}}(\forall x(\exists y/\{x\})\, x \neq y) = (n-1)/n$.

As shown in [18,21], equilibrium semantics provides a link to some standard truth functions for many-valued logics in the following sense.

Theorem 3. *Let $\mathcal{I}$ be a finite interpretation, φ and ψ two IF formulas, and ξ a variable assignment. Moreover remember that $\xi[c/x]$ denotes the variable assignment that is like ξ, except for assigning the element c to the variable x. Each of the following statements holds:*

- $v^{eq}_{\mathcal{I}}((\varphi \wedge \psi)[\xi]) = \min\{v^{eq}_{\mathcal{I}}(\varphi[\xi]), v^{eq}_{\mathcal{I}}(\psi[\xi])\}$,
- $v^{eq}_{\mathcal{I}}((\varphi \vee \psi)[\xi]) = \max\{v^{eq}_{\mathcal{I}}(\varphi[\xi]), v^{eq}_{\mathcal{I}}(\psi[\xi])\}$,
- $v^{eq}_{\mathcal{I}}(\forall x F) = \inf\{v^{eq}_{\mathcal{I}}(\varphi[\xi[c/x]]) \mid c \in D_{\mathcal{I}}\}$,
- $v^{eq}_{\mathcal{I}}(\exists x F) = \sup\{v^{eq}_{\mathcal{I}}(\varphi[\xi[c/x]]) \mid c \in D_{\mathcal{I}}\}$.

Theorem 3 can be read as a justification of minimum, maximum, infimum, and supremum as truth functions for conjunction, disjunction, existential and universal quantification in a many-valued logic, where the set of truth values is identified with the unit interval $[0,1]$. While negation only occurs in front of atomic formulas for IF formulas, it is clear that also $\lambda x(1-x)$ as truth function for negation fits the picture provided by equilibrium semantics.

3 Giles's Game for Łukasiewicz logic

In the last section we have seen that equilibrium semantics relates IF logic to a propositional many-valued logic, where an assignment $\mathcal{M}$ of truth values in the real unit interval $[0,1]$ to atomic formulas—in the following just called *many-valued interpretation*—is extended to logically complex formulas as follows.

$$\begin{aligned}
&v_{\mathcal{M}}(\varphi \wedge \psi) = \min(v_{\mathcal{M}}(\varphi), v_{\mathcal{M}}(\psi)),\\
&v_{\mathcal{M}}(\varphi \vee \psi) = \max(v_{\mathcal{M}}(\varphi), v_{\mathcal{M}}(\psi)),\\
&v_{\mathcal{M}}(\neg\varphi) = 1 - v_{\mathcal{M}}(\varphi),\\
&v_{\mathcal{M}}(\bot) = 0,\\
&v_{\mathcal{M}}(\top) = 1.
\end{aligned}$$

This logic is sometimes simply identified with 'fuzzy logic', (e.g. in [19]). Following [1], we call it *Kleene-Zadeh logic*, or KZ for short. Using the notation for variable assignments introduced in Sect. 2, KZ is extended to the first order level by

$$v_{\mathcal{M}}(\forall x\varphi) = \inf\{v_{\mathcal{M}}(\varphi[\xi[c/x]]) \mid c \in D_{\mathcal{I}}\},$$
$$v_{\mathcal{M}}(\exists x\varphi) = \sup\{v_{\mathcal{M}}(\varphi[\xi[c/x]]) \mid c \in D_{\mathcal{I}}\}.$$

If we restrict attention to interpretations over finite domains, these clauses correspond to equilibrium semantics as well (cf. Theorem 3).

From the point of view of mathematical fuzzy logic, the logic KZ is rather unsatisfying. Following a paradigm developed by Petr Hájek [10], the connectives of KZ should be augmented at least by an *implication* $\supset$ and a so-called *strong conjunction* $\&$, where $\&$ is interpreted by a continuous t-norm[5] and $\supset$ by its residuum. Arguably the most important logic of that kind is Łukasiewicz logic Ł, which is obtained from KZ by adding the following truth functions ($\oplus$ is called *strong disjunction*):

$$v_{\mathcal{M}}(\varphi \supset \psi) = \min(1, (1 - v_{\mathcal{M}}(\varphi)) + v_{\mathcal{M}}(\psi)),$$
$$v_{\mathcal{M}}(\varphi \,\&\, \psi) = \max(0, v_{\mathcal{M}}(\varphi) + v_{\mathcal{M}}(\psi) - 1),$$
$$v_{\mathcal{M}}(\varphi \oplus \psi) = \min(1, v_{\mathcal{M}}(\varphi) + v_{\mathcal{M}}(\psi)).$$

In fact all other propositional connectives could by defined in Ł, e.g., from $\supset$ and $\bot$, or from $\&$ and $\neg$, alone. But neither $\supset$ nor $\&$ nor $\oplus$ can be defined in KZ.[6] The increased expressiveness of Ł over KZ is particularly prominent at the first-order level: while in KZ there are no valid formulas at all, except those involving truth constants in some obvious manner, the set of valid first-order formulas in Ł (with or without truth constants) is not even recursively enumerable due to a classic result of Scarpellini [20].

Independently of Hintikka, Robin Giles devised a game-theoretic interpretation of Łukasiewicz logic in the 1970s [7,8]. Rather than referring to Hintikka's game (which he seemingly was not aware of) Giles refers to the logical dialogue game suggested by Lorenzen [16,17] as a foundation for constructive reasoning. Initially Giles was interested in modeling logical reasoning within theories of physics and only later motivated his game for Ł explicitly as an attempt to provide "tangible meaning" for fuzzy logic [9].

We briefly review the essential features of Giles's game, called $\mathcal{G}$-game here, but refer to [5,7,8] for more detailed presentations, including adequateness proofs. Like in the $\mathcal{H}$-game, I and you are the players and we can both act in the roles **P** or **O** with respect to given formulas augmented by variable assignments. In contrast to the $\mathcal{H}$-game, there may be more than one formula at stake at any state of the $\mathcal{G}$-game. We say that an augmented formula $\varphi[\xi]$ is currently *asserted* by you, if you act as **P** and I act as **O** with respect to it; and vice versa for a formula asserted by me. Since it will matter how often a formula is asserted at a

---

[5] A t-norm is a commutative and associative function $\circ : [0,1]^2 \to [0,1]$ such that $x \circ 1 = x$ and $x < y$ implies $x \circ z \leq y \circ z$.

[6] KZ is sometimes called the 'weak fragment of Łukasiewicz logic'.

given state, we collect the formulas currently asserted by you in a *multiset*, called *your tenet*. Likewise, *my tenet* consists of the multiset of augmented formulas currently asserted by me. We denote a state by

$$\left[\varphi_1[\xi_1], \ldots, \varphi_m[\xi_m] \,\|\, \psi_1[\xi'_1], \ldots, \psi_n[\xi'_n]\right],$$

where $\{\varphi_1[\xi_1], \ldots, \varphi_m[\xi_m]\}$ is your tenet and $\{\psi_1[\xi'_1], \ldots, \psi_n[\xi'_n]\}$ is my tenet. At any given state an occurrence of a non-atomic augmented formula is picked randomly either from my or from your tenet and distinguished as *current formula*.[7]

States that only contain atomic formulas are called *elementary*. At non-elementary states the game proceeds according to the following rules. Like for the $\mathcal{H}$-game, we do not have to refer to the players' identity directly, but only to their roles with respect to the current formula (which by definition is an occurrence of some non-atomic augmented formula in **P**'s tenet).

($R^{\mathcal{G}}_{\wedge}$) If the current formula is $(\varphi \wedge \psi)[\xi]$ then **O** chooses whether to replace it by $\varphi[\xi]$ or by $\psi[\xi]$ in **P**'s tenet.

($R^{\mathcal{G}}_{\vee}$) If the current formula is $(\varphi \wedge \psi)[\xi]$ then **P** chooses whether to replace it by $\varphi[\xi]$ or by $\psi[\xi]$ in **P**'s tenet.

($R^{\mathcal{G}}_{\neg}$) If the current formula is $\neg\varphi[\xi]$ then it is replaced by $\perp$ in **P**'s tenet and $\varphi[\xi]$ is added to **O**'s tenet.

($R^{\mathcal{G}}_{\supset}$) If the current formula is $(\varphi \supset \psi)[\xi]$ then **O** chooses whether to remove it or else to replace it by $\psi[\xi]$ in **P**'s tenet and add $\varphi[\xi]$ to **O**'s tenet.

($R^{\mathcal{G}}_{\&}$) If the current formula is $(\varphi \,\&\, \psi)[\xi]$ then **P** chooses whether to replace it by both, $\psi[\xi]$ and $\varphi[\xi]$, or by $\perp$ in **P**'s tenet.

($R^{\mathcal{G}}_{\oplus}$) If the current formula is $(\varphi \oplus \psi)[\xi]$ then **O** chooses whether to remove it or to replace it by $\psi[\xi]$ and $\varphi[\xi]$ in **P**'s tenet while adding $\perp$ to **O**'s tenet.

($R^{\mathcal{G}}_{\forall}$) If the current formula is $(\forall x\varphi(x))[\xi]$ then it is replaced in **P**'s tenet by $\varphi(x)[\xi[c/x]]$, where $c \in D_{\mathcal{I}}$ is chosen by **O**.

($R^{\mathcal{G}}_{\exists}$) If the current formula is $(\exists x\varphi(x))[\xi]$ then it is replaced in **P**'s tenet by $\varphi(x)[\xi[c/x]]$, where $c \in D_{\mathcal{I}}$ is chosen by **P**.

Note that rules $R^{\mathcal{G}}_{\wedge}$, $R^{\mathcal{G}}_{\vee}$, $R^{\mathcal{G}}_{\forall}$, and $R^{\mathcal{G}}_{\exists}$ directly correspond to $R^{\mathcal{H}}_{\wedge}$, $R^{\mathcal{H}}_{\vee}$, $R^{\mathcal{H}}_{\forall}$, and $R^{\mathcal{H}}_{\exists}$, respectively. However the rules for implication ($R^{\mathcal{G}}_{\supset}$), negation ($R^{\mathcal{G}}_{\neg}$), strong conjunction ($R^{\mathcal{G}}_{\&}$), and strong disjunction ($R^{\mathcal{G}}_{\oplus}$), involve more than just one formula at the succeeding state and therefore cannot be formulated in the format of the $\mathcal{H}$-game, where only one formula is asserted at any state.

If there is no non-atomic formula left to pick as current formula, the game has reached an *elementary state*

$$\left[A_1[\xi_1], \ldots, A_m[\xi_m] \,\|\, B_1[\xi'_1], \ldots, B_n[\xi'_n]\right],$$

where the $A_i[\xi_i]$ and $B_i[\xi'_i]$ are augmented atomic formulas. To define the players' payoffs at an elementary state Giles introduces the concept of *dispersive*

[7] The powers of the players of a $\mathcal{G}$-game do not depend on the manner in which the current formula is picked at any state. In more formal presentations of the $\mathcal{G}$-game one may introduce the concepts of a regulation and of so-called internal states in formalizing state transitions. We refer to [5] for details.

elementary experiments. For each (augmented) atomic formula $A[\xi]$ there is a corresponding experiment[8] $E_{A[\xi]}$ that yields either 'yes' or 'no' at each trial. Dispersiveness refers to the fact that the same experiment may give different answers when repeated. However a fixed probability (*risk*) $\langle A[\xi]\rangle$ of yielding a negative answer is associated with $E_{A[\xi]}$. Experiment $E_\bot$ always yields a negative result and thus $\langle\bot\rangle = 1$; similarly $\langle\top\rangle = 0$. It is stipulated that at the end of any run of the game, i.e. at an elementary state, the experiment $E_{A[\xi]}$ is performed for each occurrence of an augmented atomic formula $A[\xi]$ in my tenet and that I have to pay a fixed amount of money, say 1€, to you if $E_{A[\xi]}$ yields 'no'. Likewise you have to pay 1€ to me for each assertion in your tenet, where the corresponding experiment yields a negative answer. Therefore the expected (average) total amount of money (in €) that I have to pay to you is given by

$$\sum_{1\le i\le n} \langle B_i[\xi_i]\rangle - \sum_{1\le i\le m} \langle A_i[\xi_i']\rangle\,.$$

We call this value *my (total) risk* in a run of the $\mathcal{G}$-game that ends at the elementary state $\big[A_1[\xi_1],\ldots,A_m[\xi_m]\,\|\,B_1[\xi_1'],\ldots,B_n[\xi_n']\big]$. (This amount could also be negative, indicating that the total risk associated with my assertions is smaller than that associated with your assertions. Moreover, remember that empty sums evaluate to 0, reflecting the fact that empty tenets carry no positive risk.) Risk value assignments can be seen as inverted many-valued interpretations. More precisely, given a many-valued interpretation $\mathcal{M}$, we define a corresponding assignment of risk values $\langle\cdot\rangle_\mathcal{M}$ to augmented atomic formulas by $\langle A[\xi]\rangle_\mathcal{M} = 1 - v_\mathcal{M}((A[\xi]))$.

Definition 3. *Given a formula φ, a variable assignment ξ, and an assignment $\langle\cdot\rangle$ of risk values $\in [0,1]$ to all augmented atomic formulas, an instance of the $\mathcal{G}$-game starting in state $\big[\,\|\,\varphi[\xi]\big]$, where final (elementary) states are evaluated with respect to $\langle\cdot\rangle$, is called a $\mathcal{G}$-game* for *$\varphi[\xi]$ under $\langle\cdot\rangle$.*

The value *of such a game is $1 - w$ if I have a strategy that guarantees that my risk at the final state is at most w, while you have a strategy that guarantees that my risk is at least $1 - w$.*

Remember that we insist on finite domains. Under this assumption, the adequateness of the $\mathcal{G}$-game for Łukasiewicz logic (Giles's Theorem) can be formulated as follows.[9]

Theorem 4. *Let φ be an Ł formula, ξ be a variable assignment, and $\mathcal{M}$ a (many-valued) interpretation. Then any $\mathcal{G}$-game for $\varphi[\xi]$ has value w under the risk value assignment $\langle\cdot\rangle_\mathcal{M}$ iff $v_\mathcal{M}(\varphi[\xi]) = w$.*

---

[8] The idea is that for each atomic formula A there is schematic experimental setup that turns into a concrete experiment if elements of the domain of discourse are assigned to the free variables in A.

[9] Giles actually never considered strong conjunction and strong disjunction. For a detailed proof including strong conjunction we refer to [5]. That paper also features a link between the $\mathcal{G}$-game and an analytic proof system for Ł based on hypersequents.

4 Connecting the $\mathcal{G}$-game and the $\mathcal{H}$-game

In Sect. 2 we have seen that equilibrium semantics for IF logic provides an interpretation of the connectives of the many-valued logic KZ. However, as discussed in Sect. 3, game semantics for (full) Łukasiewicz logic Ł calls for Giles's more general concept of a game state consisting of multisets of formulas currently asserted by you and me, respectively.[10] In this section we want combine equilibrium semantics for IF logic with Giles's game for Ł.

Arguably the most straightforward way to connect IF logic with Ł allows for imperfect information about the choice of witness elements for the quantifiers in Ł-formulas in the same manner as for (classical) IF formulas. We would just have to attach slash sets to the quantifiers and treat these in a corresponding version of the $\mathcal{G}$-game exactly as in the $\mathcal{H}$-game: the players' choices of assignments for quantified variables have to remain independent of any assignment to variables in corresponding slash sets. While the resulting "Independence Friendly Łukasiewicz logic" might well be worth studying, we think that the following alternative way to connect equilibrium semantics and Giles's game for Ł is actually more interesting.

The $\mathcal{G}$-game allows one to derive the truth functions for all connectives and quantifiers of Ł from principles of reasoning about logically complex statements as encoded in the rules of the game. Notice that this derivation is completely independent of Giles's interpretation of truth values for atomic formulas in terms of the risk involved in claiming that certain dispersive experiments will yield positive results. Indeed, if one insists that atomic formulas are either simply true or false in any given interpretation, then this does not affect the rules of the $\mathcal{G}$-game, but leads to a characterization of classical logic ($\wedge$ and $\&$ both collapse to classical conjunction in this version of the $\mathcal{G}$-game; likewise $\vee$ and $\oplus$ both turn into classical disjunction). On the other hand, the $\mathcal{H}$-game for IF formulas provides an interpretation of intermediate truth values without departing from classical evaluation at the atomic level: the interpretation $\mathcal{I}$, with respect to which a given IF formula is to be evaluated, assigns either 1 (true) or 0 (false) to each (augmented) atomic formula. We propose to combine these two different semantic concepts by replacing the atomic formulas of Ł and corresponding dispersive experiments of the $\mathcal{G}$-game by IF formulas and corresponding instances of the $\mathcal{H}$-game.

To implement the idea sketched above, we define a two-tiered syntax for a new logic Ł(IF), where atomic subformulas of Ł formulas are replaced by IF formulas.[11]

---

[10] As shown in [2] and in [4] one may in fact formulate alternative semantic games for Ł that, like the $\mathcal{H}$-game, keep a single formula in focus at any given state, if either an explicit truth value or a stack of formulas is added. These and related variants of semantic games are discussed in [3], but they hardly are relevant in our context.

[11] This is somewhat reminiscent of [11], where an inner language for representing events and an outer, many-valued language for expressing assertions about the probability of such events is combined.

Definition 4. *With respect to any language as specified in Definition 1, the set of* Ł(IF) *formulas is defined as follows:*

- *Every recall regular IF formula (see Definition 2) is an* Ł(IF) *formula.*
- *If F and G are* Ł(IF) *formulas then also $\underline{\neg}F$, $F\underline{\wedge}G$, $F\underline{\vee}G$, $F\underline{\&}G$, $F\underline{\oplus}G$, $F\underline{\supset}G$, $\underline{\forall}xF$, and $\underline{\exists}xF$ are* Ł(IF) *formulas.*

Note that the logical connectives and quantifiers of Ł are underlined in order to clearly separate the outer (many-valued) level of Ł(IF) formulas from the inner level of (classical) IF formulas. Ł(IF) formulas are not evaluated with respect to a many-valued interpretation (like Ł formulas), but with respect to a (finite) classical interpretation $\mathcal{I}$, as for IF formulas. The intended semantics of Ł(IF) is given by the following combination of the $\mathcal{G}$-game and the $\mathcal{H}$-game, which we call $\mathcal{GH}$-game. Suppose we want to evaluate an (augmented) Ł(IF) formula $\chi[\xi]$, then the corresponding $\mathcal{GH}$-game proceeds as follows:

Phase 1: The $\mathcal{G}$-game with initial state $\left[\,\|\,\chi[\xi]\right]$ is played until the game reaches a state $S = \left[\varphi_1[\xi_1], \ldots, \varphi_m[\xi_m] \,\|\, \psi_1[\xi_1'], \ldots, \psi_n[\xi_n']\right]$, in which all augmented formulas are (recall regular) IF formulas.

Phase 2: For each occurrence of an augmented IF formula $\varphi[\xi]$ in S a corresponding $\mathcal{H}$-game is played. If $\varphi[\xi]$ is in my tenet, then the $\mathcal{H}$-game starts with me as **P** and you as **O**, as usual. But if $\varphi[\xi]$ is in your tenet, then the initial roles are reversed: I act as **O** and you as **P**. No information about other instances of the $\mathcal{H}$-game initiated at state S is available to the players.

For the final evaluation we proceed like in the $\mathcal{G}$-game, where each instance $\varphi[\xi]$ of an augmented IF formula in S is treated like an atomic formula for which the corresponding dispersive experiment $E_{\varphi[\xi]}$ is the $\mathcal{H}$-game with initial formula $\varphi[\xi]$ as specified for Phase 2. If $\varphi[\xi]$ is in my tenet of S then I have to pay 1€ to you if you (initially acting as **O**) win the game. On the other hand, if $\varphi[\xi]$ is in your tenet of S then I initially act as **O** and you have to pay 1€ to me if I win the game. Assuming that we employ mixed strategies and play rationally, this setup guarantees that my risk associated with $\varphi[\xi]$ is equal to the inverse of my expected payoff at a Nash equilibrium of the $\mathcal{H}$-game for $\varphi[\xi]$. In other words, the risk value for $\varphi[\xi]$ is $1 - v_{\mathcal{I}}^{eq}(\varphi[\xi])$, where $v_{\mathcal{I}}^{eq}(\varphi[\xi])$ is the value of the $\mathcal{H}$-game for the IF formula $\varphi[\xi]$, as defined in Sect. 2.

Definition 5. *In analogy to Definition 3, we speak of a $\mathcal{GH}$-game for $\varphi[\xi]$ with respect to the (classical) interpretation $\mathcal{I}$ if the game starts in state $\left[\,\|\,\varphi[\xi]\right]$ and the evaluation is as indicated above: we compute the overall risk like for the $\mathcal{G}$-game that arises if each $\varphi'[\xi']$, where φ' is a largest sub-formula of φ that is an IF formula, is treated like an atomic formula for which the corresponding dispersive experiment consists in a run of the $\mathcal{H}$-game for $\varphi'[\xi']$ with respect to $\mathcal{I}$.*

The value of such a $\mathcal{GH}$-game is w if I have a strategy that guarantees that my overall risk evaluates to at most $1-w$, while you have a strategy that ensures that my risk is at least $1-w$.

Definition 6. *The truth value $v_{\mathcal{I}}(\varphi[\xi])$ of an augmented* Ł(IF) *formula $\varphi[\xi]$ under a classical interpretation $\mathcal{I}$ is defined as follows.*

- *If the outermost connective or quantifier of φ is an (underlined) connective or quantifier of* Ł *then $v_{\mathcal{I}}(\varphi[\xi])$ is obtained from the value(s) of the immediate sub-formula(s) just like specified for Ł at the beginning of Sect. 3.*
- *Otherwise, φ is a recall regular IF formula and we set $v_{\mathcal{I}}(\varphi[\xi]) = v_{\mathcal{I}}^{eq}(\varphi[\xi])$.*

The match between the game-theoretic semantics according to Definition 5 and the truth functional semantics specified in Definition 6 is obtained as a corollary to Giles's Theorem for Łukasiewicz logic (Theorem 4) and the adequateness of equilibrium semantics for IF formulas (Theorem 2).

Corollary 1. *Let φ be an* Ł(IF) *formula, $\mathcal{I}$ a classical interpretation, and ξ a variable assignment. Any $\mathcal{GH}$-game for $\varphi[\xi]$ with respect to $\mathcal{I}$ has value w iff $v_{\mathcal{M}}(\varphi[\xi]) = w$.*

Before analyzing some (schemes of) formulas in the light of Corollary 1, let us emphasize that neither implication nor strong disjunction can be expressed by IF formulas. Negation is represented, indirectly, by dualization. This latter fact can now be expressed in the object language by the (schematic) Ł(IF) formula $\underline{\neg}\varphi \underline{\leftrightarrow} \varphi^{\neg}$, where $\psi \underline{\leftrightarrow} \chi$ abbreviates $(\psi \underline{\supset} \chi) \wedge (\chi \underline{\supset} \psi)$. However, remember that $\psi \supset \chi$ is not equivalent to $\neg\psi \vee \chi$ in Ł (or in any other t-norm based logic for that matter). Therefore one should *not* define implication for IF formulas by $\psi^{\neg} \vee \chi$.

Example 4. Let φ be an arbitrary recall regular IF formula and consider the following Ł(IF) formulas:

(1) $\varphi \underline{\supset} \varphi$
(2) $\varphi \underline{\vee} \underline{\neg}\varphi$
(3) $\varphi \underline{\oplus} \underline{\neg}\varphi$

For (1), remember that $\psi \supset \psi$ is valid in Łukasiewicz logic for any Ł formula ψ. Consequently also (1) always evaluates to 1. In terms of the $\mathcal{GH}$-game starting in state $[\,\|\,\varphi \underline{\supset} \varphi]$ this can be seen as follows. According to the rule for implication, you (acting as **O**) can choose whether the next state of the game is the empty state $[\,\|\,]$ (resulting from removing $\varphi \underline{\supset} \varphi$ from my tenet) or else is $[\varphi \,\|\, \varphi]$. Clearly my risk is 0 in the first case. But it is also 0 in the second case, where we continue with two instances of the $\mathcal{H}$-game for φ: whatever amount of money I am expected to pay to you for the $\mathcal{H}$-game corresponding to the instance of φ in my tenet, it obviously equals the amount that you have to pay to me for the instance of φ in your tenet.

For (2), I can choose whether the $\mathcal{GH}$-game starting in state $[\,\|\,\varphi \underline{\vee} \underline{\neg}\varphi]$ will result in state $[\,\|\,\varphi]$ or in state $[\,\|\,\underline{\neg}\varphi]$, which further reduces to $[\varphi \,\|\, \bot]$ in Phase 1 of the game. In the first case my risk is $1 - v_{\mathcal{I}}^{eq}(\varphi)$, i.e., the inverse of the equilibrium value for the IF formula φ. In second case, I definitely have to pay 1€ to you, but expect to receive $(1 - v_{\mathcal{I}}^{eq}(\varphi))$€ from you, resulting from the $\mathcal{H}$-game for φ in your tenet where you are in role **P**. Consequently I have a strategy that limits my expected loss to $\min(1 - v_{\mathcal{I}}^{eq}(\varphi), v_{\mathcal{I}}^{eq}(\varphi))$. The value of the game is the inverse of that overall risk, i.e. $1 - \min(1 - v_{\mathcal{I}}^{eq}(\varphi), v_{\mathcal{I}}^{eq}(\varphi)) = \max(v_{\mathcal{I}}^{eq}(\varphi), 1 - v_{\mathcal{I}}^{eq}(\varphi))$, which matches the truth value calculated according to Definition 6.

In contrast to (2), formula (3) always evaluates to 1. This is obvious from Definition 6. To see it also for the corresponding $\mathcal{GH}$-game, recall that, by rule $R_{\oplus}^{\mathcal{G}}$, **O** can choose whether the initial state $[\,\|\,\varphi \underline{\oplus} \varphi]$ is succeeded by the empty state $[\,\|\,]$ or by the state $[\bot \,\|\, \varphi, \underline{\neg}\varphi]$, which reduces to $[\varphi, \bot \,\|\, \varphi\, \bot]$ in the next round. Clearly, I have no positive risk in either case. Therefore the value of the game is 1, as required.

Note that, since $\underline{\neg}\varphi$ is equivalent to $\varphi^{\neg}$ the validity of formula (3) corresponds to the fact that the $\mathcal{H}$-game is constant-sum. On the other hand, the fact that formula (2) is not valid corresponds to the indeterminateness of the $\mathcal{H}$-game, in general. Indeed, the value of $\varphi \,\underline{\vee}\, \underline{\neg}\varphi$ is below 1 iff I have neither a winning strategy in the $\mathcal{H}$-game for φ nor in the one for $\varphi^{\neg}$, where the players' roles are switched.

The above examples look at Ł(IF) as an extension of IF logic. The corresponding $\mathcal{GH}$-game widens the scope of equilibrium semantics by providing game based interpretations of a richer set of (truth functional) connectives for combining IF formulas.

On the other hand, one may understand the combined game as an extension of the original $\mathcal{G}$-game, where IF-formulas take the place of atomic Łukasiewicz formulas: results of previously unspecified dispersive experiments are now obtained as results of runs of Hintikka-styles games with imperfect information. This amounts to an interpretation of intermediate truth values as equilibria in games of imperfect information that involve only classical truth and falsity.

5 Propositional IF Logic and Realizable Truth Values

One of our motivations for introducing the $\mathcal{GH}$-game was to address philosophical worries about the nature of intermediate truth values by building a many-valued logic over classical models that evaluate all atomic formulas over $\{0, 1\}$. An important question in this context is, whether IF logic is rich enough to cover a sufficient range of truth values. Equilibrium semantics for IF logic refers to constant-sum, two-player games with 0 and 1 as the only possible payoff values. It is a well known game-theoretic fact that the value of every such game is rational [24], hence the values of IF formulas under equilibrium semantics must be rationals from the interval $[0, 1]$. As the functions of the connectives of Łukasiewicz logic are closed under rational numbers, we do not obtain the full real interval $[0, 1]$, usually understood as the standard set of truth values for fuzzy logics. But can we get at least all rational values in the interval $[0, 1]$? In particular, is there for any $q \in [0, 1] \cap \mathbb{Q}$ an IF formula φ such that the value of φ is q?

Mann, Sandu and Sevenster in [18] deal with this question within the framework of predicate IF logic and give two solutions [p. 184]. The first one is based on a random quantifier expressed by an IF formula, which over an interpretation with a domain of size n and a unary predicate satisfied by exactly m elements of the domain has the value m/n. The second one is more general—it shows how to construct an IF formula which has the value m/n over every domain with more than two objects.

We present two solutions of the same problem within the framework of propositional IF logic: for any rational $q \in [0, 1]$ we define a formula φ that evaluates to q according to equilibrium semantics under any (classical propositional) interpretation.

From a game-theoretic point of view there is no reason to limit imperfect information in semantic games to the quantifier moves: it is natural to consider independent choices already on the propositional level. This leads to propositional IF logic, discussed e.g. by Pietarinen [23] and Sandu and Pietarinen [22].

The minimal version of propositional IF logic introduces formulas expressing independence of disjunctions from immediately preceding conjunctions and, likewise, independence of conjunctions from immediately preceding disjunctions. The language of this logic is an extension of a standard propositional language by correspondingly slashed formulas.

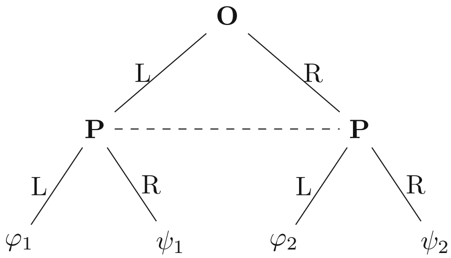
Definition 7. *The propositional IF formulas (IFP formulas) are built up over propositional variables and truth constants,* $\top, \bot$ *using* $\wedge, \vee, \neg$ *as usual. In addition we have two following clauses:*

- *If* $\varphi_1, \varphi_2, \psi_1, \psi_2$ *are IFP formulas, then* $(\varphi_1 \vee / \wedge \psi_1) \wedge (\varphi_2 \vee / \wedge \psi_2)$ *is an IFP formula*
- *If* φ *and* ψ *are IFP formulas, then* $(\varphi_1 \wedge / \vee \psi_1) \vee (\varphi_2 \wedge / \vee \psi_2)$ *is an IFP formula*

The interpretation of standard disjunction and conjunction remains the same: it consists of the choice by the Proponent **P** or the Opponent **O**, respectively. The slashed disjunction (conjunction) is, analogically to the first order case, interpreted as a game of imperfect information: one player chooses a disjunct (conjunct) without any information about the previous choice of the other player.

The moves for slashed disjunction (conjunction) cannot be labeled by the corresponding disjuncts (conjuncts), because perfect information would be recovered. If, in the semantic game for $(\varphi_1 \vee / \wedge \psi_1) \wedge (\varphi_2 \vee / \wedge \psi_2)$ **P** were asked to choose between φ_1 and ψ_1, she would know that in the previous move **O** must have chosen the left conjunct. Thus the players' choices are specified using labels ("Left disjunct", "Right disjunct", etc.).

The semantic game for the formula $(\varphi_1 \vee / \wedge \psi_1) \wedge (\varphi_2 \vee / \wedge \psi_2)$ has the following extensive form:



This game is the simplest (non-trivial) case of incomplete information as we can also see from its strategic form:

$$\begin{array}{lcc} \mathbf{O}\backslash\mathbf{P} & L & R \\ L & \multicolumn{2}{c}{\multirow{2}{*}{$\begin{pmatrix} \varphi_1 & \psi_1 \\ \varphi_2 & \psi_2 \end{pmatrix}$}} \\ R & & \end{array}$$

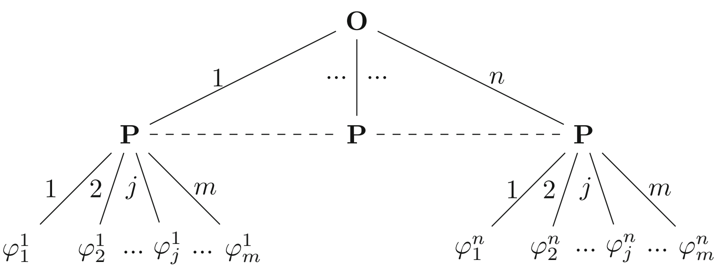
More general versions of propositional IF logic are discussed in the literature that, e.g., allow one to express that a disjunction is independent from any preceding conjunction. This requires a more substantial modification of syntax, that we will not introduce here, since the indicated minimal version is sufficient for our purposes. However, we will use a more concise notation suggested by Sandu and Pietarinen:

$$(\varphi_1 \vee / \wedge \psi_1) \wedge (\varphi_2 \vee / \wedge \psi_2) = W(\varphi_1, \psi_1, \varphi_2, \psi_2)$$

Our first solution for recovering arbitrary rationals in $[0, 1]$ as equilibrium values requires an extension of the syntax from binary to n-ary conjunctions and disjunctions. We replace the clause for (binary) slashed disjunction from Definition 7 by the following one for n-ary disjunction:

– let $m, n \geq 2$ and let φ_i^j for $i = 1, \ldots, m$, $j = 1, \ldots, n$ be IFP formulas, then $(\varphi_1^1 \vee / \wedge \varphi_2^1 \vee / \wedge \cdots \vee / \wedge \varphi_m^1) \wedge \cdots \wedge (\varphi_1^n \vee / \wedge \varphi_2^n \vee / \wedge \cdots \vee / \wedge \varphi_m^n)$ is an IFP formula

and a similarly in the case of slashed conjunction. Like in the binary case, the independence of the slashed connective is with respect to the immediately preceding connective. Thus n-ary slashed disjunction corresponds to the following game tree:



The corresponding strategic form is represented by the following $m \times n$ matrix:

$$\begin{array}{cc} \mathbf{P}\backslash\mathbf{O} & \begin{array}{ccc} 1 & \dots & n \end{array} \\ \begin{array}{c} 1 \\ \vdots \\ m \end{array} & \begin{pmatrix} \varphi_1^1 & \dots & \varphi_1^n \\ \vdots & & \vdots \\ \varphi_m^1 & \dots & \varphi_m^n \end{pmatrix} \end{array}$$

We will use the matrix form of the game interpreting the n-ary conjunction/disjunction in our first proof of realizability of rationals.

Theorem 5. *For every non-negative rational number q there is a strategic two-person, zero-sum game with payoffs in $\{0, 1\}$ such that:*

1. *the value of the game is q,*
2. *the equilibrium strategy for both players is the uniform distribution.*

Proof. For a given rational $q = m/n$, where $m, n \in \mathbb{N}, 0 \leq m < n$ we construct an $n \times n$ payoff matrix M with exactly m ones in each row and each column:

- $a_{i,j} = 1$ if $1 \leq i \leq j \leq i + m - 1$ and $j \leq n$
- $a_{i,j} = 1$ if $1 \leq j \leq (i + m - 1) \bmod n$ and $i + m - 1 > n$
- $a_{i,j} = 0$ otherwise

The pure strategies of both players consist in picking $i, j \in \{1, ..., n\}$. We denote their mixed strategies by $(p_1, ...p_n)$ and $(q_1, ..., q_n)$, respectively. The element $a_{i,j}$ is the payoff of the row player if she is playing i and the column player is playing j, the payoff of the column player for the same profile (couple) of pure strategies is $1 - a_{i,j}$.

It is clear that if both players play the mixed strategy corresponding to the uniform distribution ($p_i = q_j = 1/n$), the probability of each payoff $a_{i,j}$ is $1/n^2$. As there are m ones in each of the n rows, the payoff of the row player is $\frac{1}{n^2} m \cdot n \cdot 1 = m/n$.

It remains to check that the mixed strategy profile corresponding to uniform distributions for both players is an equilibrium pair. A standard characterization of equilibrium says that no player can improve her payoff by a unilateral deviation from her equilibrium strategy. As is well known, it is sufficient to check this condition with respect to *pure* strategies. If the first player deviates from uniform distribution playing i-th row against the uniform distribution played by the column player, her payoff is $1/n \cdot m$ (as there are exactly m ones in each row and each of them has probability $1/n$). This is the same as the equilibrium payoff, so no improvement is gained. The condition for the second player can be checked in a similar way.

The following theorem about realizability of rationals just translates the strategic game from Theorem 5 into the language of propositional IF logic.

Theorem 6. *For any $q \in [0, 1] \cap \mathbb{Q}$ there is an IFP formula ψ (using n-ary slashed disjunction) such that the value of ψ according to equilibrium semantics is q under any interpretation.*

Proof. Assume $q = m/n$, where $m, n \in \mathbb{N}, 0 \leq m < n$. From Theorem 5 we obtain an $n \times n$ matrix $(a_i^j), a_i^j \in \{0, 1\}$, representing a strategic game with the equilibrium value m/n. We can straightforwardly express this matrix in IF notation using n-ary slashed disjunction and the constants $\top$ and $\bot$ as follows: $\psi = (\varphi_1^1 \vee / \wedge \varphi_2^1 \vee / \wedge \cdots \vee / \wedge \varphi_n^1) \wedge ... \wedge (\varphi_1^n \vee / \wedge \varphi_2^n \vee / \wedge \cdots \vee / \wedge \varphi_n^n)$, where $\varphi_i^j = \top$ if $a_i^j = 1$ and $\varphi_i^j = \bot$ if $a_i^j = 0$.

We now present a second construction for realizing all rationals in $[0, 1]$ as equilibrium values, using IFP formulas logic with only binary connectives, as specified in Definition 7. It is based on iterating the connective W (encoding the simplest proper game of imperfect information). In analogy to the first-order case (in Sect. 2) we will denote by $v_I^{eq}(\varphi)$ the value of the formula φ according to equilibrium semantics. As the choice of a particular propositional interpretation I plays no role, we omit the index I. We also introduce the operator $\bar{W}$ of type $[0, 1]^4 \rightarrow [0, 1]$ corresponding to the connective W: $\bar{W}(x, y, z, u) = v^{eq}(W(\varphi_1, \varphi_2, \psi_1, \psi_2))$, where $v^{eq}(\varphi_1) = x, v^{eq}(\varphi_2) = y, v^{eq}(\psi_1) = z, v^{eq}(\psi_2) = u$.

Observe that W allows us to express random choice between two formulas φ, ψ. The equilibrium strategy of the game corresponding to $W(\varphi, \psi, \psi, \varphi)$ amounts to picking up with the same probability one of the elements φ, ψ. Consequently its equilibrium value is a mean of the values of φ and ψ: $v^{eq}(W(\varphi, \psi, \psi, \varphi)) = (v^{eq}(\varphi) + v^{eq}(\psi)/2)$. The resulting random choice connective is denoted by Π, where $\Pi(\varphi, \psi) = W(\varphi, \psi, \psi, \varphi)$, $\bar{\Pi}$ will be the corresponding function (operator), hence $v^{eq}(\Pi(\bot, \top)) = v^{eq}(W(\bot, \top, \top, \bot)) = \bar{W}(0, 1, 1, 0) = 1/2$. The corresponding game is the one of inverse Matching Pennies (IMP) with the following payoff matrix:

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

It is easy to see that iterating the Π-operator gives us powers of $1/2$: $\bar{\Pi}(0, \bar{\Pi}(0, 1)) = 1/4$, $\bar{\Pi}(0, \bar{\Pi}(0, \bar{\Pi}(0, 1))) = 1/8$ etc. We can represent this schematically as "pluggin-in" IMP into IMP :

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} 0 & \mathsf{IMP} \\ \mathsf{IMP} & 0 \end{pmatrix} \end{array}$$

The choices (L, L) and (R, R) lead to the payoff for the first player, while for the choices (L, R) and (R, L) the game continues by playing IMP. This corresponds to symmetric iterations of the $\bar{W}$-operator: for example $(\bar{\Pi}(0, \bar{\Pi}(0, 1))) = \bar{W}(0, \bar{W}(0, 1, 1, 0), W(0, 1, 1, 0), 0)$. What happens in the case of "asymmetric" iterations? Consider the game which is the result of the simplest case of an asymmetric plug-in of IMP:

$$\begin{array}{c c} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} 0 & \mathsf{IMP} \\ 1 & 0 \end{pmatrix} \end{array} \text{ with the payoff matrix } \begin{array}{c c} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix} \end{array}.$$

This corresponds to the substitution of the random choice operator at the second argument position of the $\bar{W}$-operator: $\bar{W}(0, \bar{\Pi}(0,1), 1, 0)$. We can easily check that the value of the game is 1/3 and the equilibrium strategy profile is $\langle(2/3, 1/3), (1/3, 2/3)\rangle$. We show that this simple kind of iteration in combination with negation is already sufficient to obtain all rationals. To simplify notation we introduce a unary connective O, defined by $O(\varphi) = W(\bot, \varphi, \top, \bot)$. Thus the formula corresponding to the above game can be written as $O(\varphi)$, where $v^{eq}(\varphi) = 1/2$ and $\bar{O}(1/2) = 1/3$. In fact we obtain $\bar{O}(1/n) = 1/(n+1)$ for every $n \in \mathbb{N}, n \geq 1$, as follows from the following Lemma.

Lemma 1. *The constant sum, two players strategic game represented by the payoff matrix*

$$\begin{array}{c c} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{pmatrix} 0 & k/n \\ 1 & 0 \end{pmatrix} \end{array}$$

where $n, k \in \mathbb{N}, n \geq 1$ and $0 \leq k \leq n$, has the unique Nash equilibrium (equilibrium strategy profile) $\langle(n/(n+k), k/(n+k)), (k/(n+k), n/(n+k))\rangle$ and the corresponding equilibrium value for the row player is $k/(n+k)$.

Proof. The case for $k = 0$ is trivial: the column player has a pure winning strategy R and the payoff of the row player is $0 = 0/(n+0)$. Except for this trivial case no pure strategy is an equilibrium, so every mixed equilibrium strategy is proper—both pure strategies will be played with a non-zero probability (i.e., both of them belong to the support of mixed equilibrium strategies). It is a well known game-theoretic fact that in this case an equilibrium strategy $(p, 1-p)$ of the first player must yield the same payoff in response to both pure strategies of the second player, which gives us the equation: $p \cdot \frac{k}{n} = 1 - p$. This allows us to calculate the required probability values: $p = \frac{n}{n+k}, 1-p = \frac{k}{n+k}$. A similar line of reasoning leads to the values for the second player: $q = \frac{k}{n+k}, 1-q = \frac{n}{n+k}$. The equilibrium value of the game is $\frac{n}{n+k} \cdot \frac{k}{n+k} \cdot 0 + (\frac{n}{n+k})^2 \cdot \frac{k}{n} + (\frac{k}{n+k})^2 \cdot 1 + \frac{k}{n+k} \cdot \frac{n}{n+k} \cdot 0 = (\frac{n}{n+k})^2 \cdot \frac{k}{n} + (\frac{k}{n+k})^2 \cdot 1 = \frac{n \cdot k + k^2}{n+k^2} = \frac{k}{n+k}$.

Lemma 1 shows that $\bar{O}(k/n) = k/(k+n)$. To obtain all rationals in $[0,1]$ as equilibrium values of IFP formulas we have to use negation in addition to the connective O. The following theorem shows that this is sufficient.

Theorem 7. *Every $q \in [0,1] \cap \mathbb{Q}$ can be obtained as the result of iteratively applying the functions $\bar{O}(x)$ and $1 - x$ to either 0 or 1.*

Proof. We show by induction on k that we can get all values $q = k/n$ for $0 \leq k \leq n$, where $n, k \in \mathbb{N}$ and $n \geq 1$. In fact it is sufficient to show we can get k/n for all $0 \leq k < n/2$ because the rest of the values is obtained by applying $1 - x$.

Base step: For the cases $k = 0, k = n$ we obtain from Lemma 1 that $\bar{O}(0) = 0$ and $\bar{O}(n/n) = \bar{O}(1) = 1/2$.

Induction step: Assume that we have all the values k'/n' where $1 \leq n' < n$ and $1 \leq k' \leq n'$. We show, that we can get k/n for all $k, 1 \leq k < n/2$. It follows from Lemma 1 that $\frac{k}{n} = \frac{k}{m+k} = \bar{O}(\frac{k}{m})$ for $m = n - k$. As we only need $k < \frac{n}{2}$, it holds that $2k < n$ and $k < n - k$. But then $k < n - k = m < n$ and the value $\frac{k}{m}$ is guaranteed by the induction hypothesis.

We obtain an expression of the form $\pm\bar{O}(\pm\bar{O}(\cdots \pm \bar{O}(x)))$, where $+\bar{O}(\_)$ is $\bar{O}(\_)$, $-\bar{O}(\_)$ is $(1 - \bar{O}(\_))$ and x equals 0 or 1. In fact we only need $x = 0$ in the case our $q = 0$, so we get either $\bar{O}(0)$ or $\pm\bar{O}(\pm\bar{O}(\cdots \pm \bar{O}(1)))$ The corresponding formula is $\pm O(\pm O(\ldots \pm O(\top)))$, or $O(\bot)$ where $+O(\_)$ is $O(\_)$, $-O(\_)$ is $(\neg O(\_))$.

The only remaining step is to translate the iterated $\bar{O}$-operator back to the language of propositional IF logic.

Theorem 8. *For every $q \in [0,1] \cap \mathbb{Q}$ there is an IFP formula ψ built up from $\top$ and $\bot$ using only binary slashed disjunction and negation such that the value of ψ according to equilibrium semantics is q under any interpretation.*

Proof. Remember that $O(\varphi) = W(\bot, \varphi, \top, \bot) = (\bot \vee / \wedge \varphi) \wedge (\top \vee / \wedge \bot)$. Therefore the claim immediately follows from Theorem 7.

Example 5. We illustrate the previous results by constructing a propositional IF formula the value of which is $2/5$. We start by expressing this value in the terms of the operator $\bar{O}$ using the formula $k/(n + k) = \bar{O}(k/n)$ from Lemma 1 iteratively. Our initial value can be expressed as $2/5 = \bar{O}(2/3)$. In the second step we need the value $2/3$. As it is bigger than $1/2$ we obtain it by complementation: $1 - \bar{O}(1/2) = 1 - 1/3$. We already know that $\bar{O}(1) = 1/2$. Putting together all these expressions we get $2/5 = \bar{O}(1 - \bar{O}(\bar{O}(1)))$. The translation to IF propositional logic is less compact, but it straightforwardly encodes the corresponding game tree. Using the connective O, corresponding to the $\bar{O}$-operator, we get $O(\neg O(O(\top)))$, which we can expand using $O(\varphi) = W(\bot, \varphi, \top, \bot) = (\bot \vee / \wedge \varphi) \wedge (\top \vee / \wedge \bot)$:

$$\begin{aligned} O(\neg O(O(\top))) &= W(\bot, \neg W(\bot, W(\bot, \top, \top, \bot), \top, \bot), \top, \bot) \\ &= (\bot \vee / \wedge \neg(\bot \vee / \wedge (\bot \vee / \wedge \top) \wedge (\top \vee / \wedge \bot)) \wedge (\top \vee / \wedge \bot)) \wedge (\top \vee / \wedge \bot) \end{aligned}$$

6 Conclusion

We have revisited two different types of semantic games: On the one hand, there is Hintikka's game-theoretic characterization of classical truth in a model, generalized by Hintikka and Sandu to IF logic that incorporates imperfect information, syntactically encoded by slashed quantifiers and connectives. Equilibrium semantics for IF logic provides an interpretation in which intermediate truth values arise from equilibrium strategies in the corresponding $\mathcal{H}$-game. On the other hand, there is Giles's game ($\mathcal{G}$-game) for Łukasiewicz logic, an expressive many-valued logic that, e.g., features two different forms of conjunction and disjunction. The $\mathcal{H}$-game and the $\mathcal{G}$-game are quite different, not only regarding their respective target logic, but also in their basic structure. Nevertheless they nicely fit together from a certain perspective. We introduced the $\mathcal{GH}$-game and the corresponding logic $\mathsf{Ł(IF)}$, which allows one to combine IF formulas with the connectives and quantifiers of Łukasiewicz logic. In this manner intermediary truth values retain their interpretation in terms of equilibria in imperfect information games, while featuring a set of propositional connectives and corresponding truth functions that reaches well beyond just min (weak conjunction), max (weak disjunction), and $1-x$ (negation). Thus $\mathsf{Ł(IF)}$ generalizes both, Łukasiewicz logic $\mathsf{Ł}$ as well as IF logic.

We have also addressed an interesting issue that already arises for IF logic: Can one represent all rational truth values already at the propositional level? We provided a positive answer in two different manners. If one allows for "slashed" conjunction or disjunction with arbitrary finite arity, formulas of minimal nesting depth built up from $\top$ and $\bot$ are sufficient to represent all rational truth values. If one insists on binary disjunction and conjunction a more elaborate construction, involving unbounded nesting of slashed connectives, is needed for this purpose.

We conclude by listing a number of possible directions for further investigations triggered by our considerations. As already indicated at the beginning of Sect. 4, it might be worthwhile to work out an independence friendly version of Łukasiewicz logic, which calls for a different generalization of the $\mathcal{H}$- and the $\mathcal{G}$-game. Yet another combination and generalization of the underlying games will arise if one considers arbitrary nestings of slashed (classical) connectives and Łukasiewicz connectives, instead of the strictly two-tiered syntax suggested here. One might also consider other many-valued logics for combining and/or generalizing IF formulas, e.g., Gödel or Product logic. Finally, we want to hint at subtle connections to Japaridze's Computability Logic (see, e.g., [14]). While Japaridze's game model of computation is quite different in several respects, there emerges some similarity in the options for representing various forms of combining (sub-)games by corresponding connectives, at least if one is willing to go beyond the truth functional setting induced by the two-tired syntax of $\mathsf{Ł(IF)}$.

References

1. Aguzzoli, S., Gerla, B., Marra, V.: Algebras of fuzzy sets in logics based on continuous triangular norms. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS, vol. 5590, pp. 875–886. Springer, Heidelberg (2009). doi:10.1007/978-3-642-02906-6_75
2. Cintula, P., Majer, O.: Towards evaluation games for fuzzy logics. In: Majer, O., Pietarinen, A.-V., Tulenheimo, T. (eds.) Games: Unifying Logic, Language, and Philosophy, pp. 117–138. Springer, Amsterdam (2009)
3. Fermüller, C.G.: Semantic games for fuzzy logics. In: Cintula, P., Fermüller, C.G., Noguerra, C. (eds.) Handbook of Mathematical Fuzzy Logic, vol. 3, pp. 969–1028 (2016)

4. Fermüller, C.G.: Semantic games with backtracking for fuzzy logics. In: 44th International Symposium on Multiple-Valued Logic (ISMVL), pp. 38–43. IEEE Press (2014)
5. Fermüller, C.G., Metcalfe, G.: Giles's game and the proof theory of Łukasiewicz logic. Stud. Log. **92**, 27–61 (2009)
6. Galliani, P.: Game values and equilibria for undetermined sentences of dependence logic. Master thesis, Universiteit van Amsterdam (2008)
7. Giles, R.: A non-classical logic for physics. Stud. Log. **4**(33), 399–417 (1974)
8. Giles, R.: A non-classical logic for physics. In: Wojcicki, R., Malinkowski, G. (eds.) Selected Papers on Łukasiewicz Sentential Calculi, pp. 13–51. Polish Academy of Sciences (1977)
9. Giles, R.: Semantics for fuzzy reasoning. Int. J. Man Mach. Stud. **17**(4), 401–415 (1982)
10. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer, Dordrecht (1998)
11. Hájek, P., Godo, L., Esteva, F.: Fuzzy logic and probability. In: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, pp. 237–244. Morgan Kaufmann Publisher (1995)
12. Hintikka, J.: Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic. Clarendon Press, Oxford (1973)
13. Hintikka, J., Sandu, G.: Game-theoretical semantics. In: Handbook of Logic and Language. Elsevier (2010)
14. Japaridze, G.: Introduction to computability logic. Ann. Pure Appl. Logic **123**(1), 1–99 (2003)
15. Kuhn, H.W.: Extensive games and the problem of information. Contrib. Theor. Games **2**, 193–216 (1953)
16. Lorenzen, P.: Logik und Agon. In: Atti Congr. Internaz. di Filosofia, Venezia, 12–18 Settembre 1958, vol, IV. Sansoni (1960)
17. Lorenzen, P.: Dialogspiele als semantische Grundlage von Logikkalkülen. Archiv für mathemathische Logik und Grundlagenforschung **11**, 32–55, 73–100 (1968)
18. Mann, A.L., Sandu, G., Sevenster, M.: Independence-Friendly Logic: A Game-Theoretic Approach. Cambridge University Press, Cambridge (2011)
19. Nguyêñ, H.T., Walker, E.A.: A First Course in Fuzzy Logic. CRC Press, Boca Raton (2006)
20. Scarpellini, B.: Die Nichtaxiomatisierbarkeit des unendlichwertigen Prädikatenkalküls von Łukasiewicz. J. Symbolic Logic **27**(2), 159–170 (1962)
21. Sevenster, M., Sandu, G.: Equilibrium semantics of languages of imperfect information. Ann. Pure Appl. Logic **161**(5), 618–631 (2010)
22. Sandu, G., Pietarinen, P.: A partiality and games: propositional logic. Logic J. IGPL **9**(1), 101–121 (2001)
23. Pietarinen, A.V.: Propositional logic of imperfect information: foundations and applications. Notre Dame J. Formal Logic **42**(4), 193–210 (2001)
24. Raghavan, T.E.S.: Zero-sum two person games. In: Aumann, R.J., Hart, S. (eds.) Handbook of Game Theory with Economic Applications, vol. 2, pp. 736–759. Elsevier, Amsterdam (1994)

Quasi-Realization

Melvin Fitting(✉)

City University of New York, New York City, USA
melvin.fitting@gmail.com

Abstract. Justification logics connect with modal logics via *Realization Theorems*. The first such theorem was proved constructively by Artemov, [1]. It showed how to translate an S4 sequent proof, as a whole, into an LP proof. We present a different algorithmic Realization proof for LP/S4, proceeding step by step instead of working on the entire proof, and dividing the argument into two natural parts, one specific to LP/S4, the other widely applicable to justification/modal pairs. This structure makes an implementation easier, and we provide a link to one in Prolog.

Keywords: Justification logic · Modal logic · Realization · Quasi-Realization · Tableau · Prolog

1 Introduction

Justification logics are similar to modal logics, but with modal operators replaced by an infinite family of *justifications*. The first justification logic, LP (logic of proofs), was introduced by Artemov [1]. It played an essential role in Artemov's arithmetic completeness result for intuitionistic logic, finishing a line of research that began with Gödel, [13]. A step in that work shows that LP has a direct connection with modal S4, via a *Realization Theorem*. This says that every S4 theorem has a Realization, a replacement of modal operators by justification terms, that is a theorem of LP. In a sense, a Realization represents the flow of information hidden in the modal operators. One sometimes sees references to S4 as a logic of *implicit* knowledge, while LP *explicitly* represents that knowledge.

Since Artemov's work, many justification logics have been created, and many proofs of Realization have been developed. It is now known that the family of modal/justification pairs is infinite, and much work has gone into the investigation of justification counterparts of familiar modal logics such as K, T, K4, S5, and so on. See [2,8] for a discussion of the family of justification logics. Recently quantification has been added, but this is another story, see [3,9].

The first proof of a Realization theorem can be found in [1]. It is constructive. Constructive proofs commonly use cut-free Gentzen sequent systems but prefixed tableaus/nested sequents have also been used, providing a modular approach applicable to a basic family of modal logics, [14]. There are non-constructive proofs based on semantics, [6]. Recently there is a non-constructive proof using the model existence theorem, [11].

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 313–332, 2017.
DOI: 10.1007/978-3-662-54332-0_17

Here we give a new constructive proof of Realization. We present it for LP and S4, but the argument is clearly more general. We use semantic tableaus rather than a sequent calculus, but there is a well-known connection between them. The central fact is that our Realization algorithm proceeds step by step, rather than working with a tableau proof as a whole, and this makes implementation easier. A link to a Prolog implementation can be found at [5]. In addition, our algorithm divides into two parts. First a *Quasi-Realization Theorem* is shown (formulated as Theorem 3 in Sect. 6 and finally proved in Sect. 9). This extracts a Quasi-Realization from a modal tableau proof—a simpler thing to do than getting a Realization proper. This part depends on details of S4, and needs modification for other modal logics. Algorithmic conversion from Quasi-Realization to Realization is independent of the particular logic, and can be found in [12]. This part is only sketched here (in Sect. 10).

2 The Logic LP

This section contains a brief formulation of LP axiomatically, which comes from [1]. A semantics will not be needed in this paper.

Justification terms are built up from justification variables, v_1, v_2, ..., and justification constants, c_1, c_2, ..., using the function symbols $\cdot$, $+$, and !. If t is a justification term, so is $!t$, and if t and u are justification terms, so are $(t + u)$ and $(t \cdot u)$. (We may omit some parenthesis when no harm is done.) *Ground* justification terms are those without variables.

Formulas are built up from propositional variables, P, Q, ..., and the propositional constant $\bot$ using $\supset$ (with other connectives defined in the usual way), and an extra rule of formation: if t is a justification term and X is a formula then $t{:}X$ is a formula.

The formula $t{:}X$ can be read: "t is a justification of X." Justification constants represent justifications of basic, assumed truths—axioms. Justification variables are thought of as representing arbitrary justifications. If t is a justification of $X \supset Y$ and u is a justification of X, think of $t \cdot u$ as a justification of Y. The operation ! is a checker: if t is a justification of X then $!t$ is a verification that t is such a justification. The operation $+$ combines justifications, in that $t + u$ justifies all the things that t justifies plus all the things that u justifies.

The following axiom system for LP is from [1]. Axioms are specified by giving axiom schemas and rules, and these are:

A0. **Classical**	Enough classical propositional axiom schemes
A1. **Application**	$t{:}(X \supset Y) \supset (s{:}X \supset (t{\cdot}s){:}Y)$
A2. **Factivity**	$t{:}X \supset X$
A3. **Justification Checker**	$t{:}X \supset\, !t{:}(t{:}X)$
A4. **Weakening**	$s{:}X \supset (s{+}t){:}X$
	$t{:}X \supset (s{+}t){:}X$
R1. **Modus Ponens**	$\vdash Y$ provided $\vdash X$ and $\vdash X \supset Y$
R2. **Axiom Necessitation**	$\vdash c{:}X$ where X is an axiom *A0 – A4* and c is a justification constant.

A proof is a finite sequence of formulas each of which is an axiom or comes from earlier terms by one of the rules of inference. *Derivations* can be introduced either directly or indirectly by defining $\Gamma \vdash_{\mathsf{LP}} X$ to mean that $(G_1 \wedge \ldots \wedge G_n) \supset X$ is a theorem for some finite subset $\{G_1, \ldots, G_n\}$ of Γ.

Which constants are associated with which axioms for rule *R2* applications is called a *constant specification*. More formally, a constant specification is a set $\mathcal{C}$ whose members are of the form $c{:}A$ where c is a justification constant and A is an axiom. A proof uses constant specification $\mathcal{C}$ if each instance of Axiom Necessitation is in $\mathcal{C}$. A constant specification can be given ahead of time, or created during the course of a proof. We will assume all constant specifications are *axiomatically appropriate*: each instance of one of the axiom schemes is assigned at least one constant. In addition, all such assignments will be *injective*, no justification constant is used for more than one axiom. Many other conditions have been investigated, but we are not interested in constant specification details here.

If Z is any theorem of LP, and we replace every proof polynomial by $\Box$ (the *forgetful* projection), the result is a theorem of S4. This is easy to see: it is the case for each axiom of LP, and is preserved by the LP rules of derivation. The Artemov Realization Theorem, from [1], is a converse to this.

Theorem 1 (Realization Theorem). *If Z is a theorem of S4, there is some replacement of $\Box$ symbols with justification terms to produce a theorem of LP, provable using an injective, axiomatically appropriate constant specification). This can be done so that negative occurrences of $\Box$ in Z are replaced with distinct justification variables, and positive occurrences by justification terms that may involve those variables.*

Negative occurrences of justification variables can be thought of as inputs, and positive justification terms as outputs. Thus theorems of S4 carry implicit constructive functional content which their LP Realizations make explicit.

A fundamental result is the Lifting Lemma, from [1,2], not proved here, which says that proofs and derivations in LP can be internalized. We present a somewhat simplified version, which is enough for our purposes in this paper.

Theorem 2 (Lifting Lemma). *Assume we have an axiomatically appropriate constant specification. Suppose $s_1 : X_1, \ldots, s_n : X_n \vdash_{LP} Z$. Then there is a justification term $t(s_1, \ldots, s_n)$ such that $s_1{:}X_1, \ldots, s_n{:}X_n \vdash_{LP} t(s_1, \ldots, s_n){:}Z$.*

Corollary 1. *With an axiomatically appropriate constant specification, if Z has an LP proof, then for some ground proof polynomial t, $t{:}Z$ will have an LP proof.*

3 Tableaus

Tableaus are refutation proof systems. Informally, one assumes a formula X could be false under some circumstances and derives a syntactic contradiction. Classical formulas are built up from propositional letters and $\bot$ using $\wedge$, $\vee$, $\supset$, and $\neg$, though other binary connectives could also be admitted. Smullyan's *uniform notation* is useful here, [15,16], both for theoretical purposes and to simplify tableau implementations. We use *signed* formulas. Two special symbols, T and F, are introduced and $T\,X$ and $F\,X$ are signed formulas if X is a formula. The intended reading is that X is true, or false respectively. Signed formulas involving binary connectives divide into α cases, conjunctive, and β cases, disjunctive. For each case, two components are also specified. This is given in Fig. 1.

Conjunctive			Disjunctive		
α	α_1	α_2	β	β_1	β_2
$T\,X \wedge Y$	$T\,X$	$T\,Y$	$F\,X \wedge Y$	$F\,X$	$F\,Y$
$F\,X \vee Y$	$F\,X$	$F\,Y$	$T\,X \vee Y$	$T\,X$	$T\,Y$
$F\,X \supset Y$	$T\,X$	$F\,Y$	$T\,X \supset Y$	$F\,X$	$T\,Y$

Fig. 1. α- and β-formulas and components

A tableau proof is a special labeled binary tree. A proof of X begins with a tree having only a root node, labeled $F\,X$. Then a tree is 'grown' using the *branch extension* rules, given in Fig. 2. All trees produced this way are tableaus.

$$\frac{T\,\neg X}{F\,X} \qquad \frac{F\,\neg X}{T\,X} \qquad \frac{\alpha}{\begin{array}{c}\alpha_1\\ \alpha_2\end{array}} \qquad \frac{\beta}{\beta_1 \mid \beta_2}$$

Fig. 2. Classical branch extension rules

Tableaus are displayed as downward branching trees. Think of a tree as representing the disjunction of its branches, and a branch as representing the conjunction of the signed formulas on it. The members of a tableau branch can be thought of as constituting a *set*, or a *multi-set*, or even a *sequence*. We treat branches as sets. Tableau rules are *non-deterministic*. At each stage we choose a signed formula occurrence on a branch and apply a rule to it. Since the order of choice is arbitrary, there can be many tableaus for a single signed formula. A tableau branch is *closed* if it contains $T\,A$ and $F\,A$ for some formula A, or if it contains $T\,\bot$. If each branch is closed, the tableau is *closed.* A closed tableau for $F\,X$ is a tableau proof of X. A branch is *atomically* closed if it contains $T\,P$ and $F\,P$ where P is atomic. If a tableau can be closed, it can continued to closure at the atomic level, so we will require atomic closure. Classical branch extension rules can be restricted to *single use*: a classical tableau rule is never applied to a signed formula occurrence on a branch more than once. (This does not work for all logics, however.)

An example of a classical tableau proof is given in Fig. 3. Numbers are for reference purposes only. In it, 2 and 3 are from 1 by α; 4 and 5 are from 3 by α; 6 and 7 are from 5 by α; 8 and 9 are from 2 by β. 10 and 11 are from 4 by β.

1. $F\,[(P \supset (Q \supset R)) \supset ((P \vee S) \supset ((Q \supset R) \vee S))]$
2. $T\,P \supset (Q \supset R)$
3. $F\,(P \vee S) \supset ((Q \supset R) \vee S)$
4. $T\,P \vee S$
5. $F\,(Q \supset R) \vee S$
6. $F\,Q \supset R$
7. $F\,S$

8. $F\,P$ | 9. $T\,Q \supset R$

10. $T\,P$ | 11. $T\,S$

Fig. 3. Classical proof of $(P \supset (Q \supset R)) \supset ((P \vee S) \supset ((Q \supset R) \vee S))$

Reading from left to right, the branches are closed because of 8 and 10, 7 and 11, and 6 and 9. Notice that on one of the branches closure is on a non-atomic formula. This branch can be continued to yield atomic closure.

Justification logics make little use of possibility, so there is no advantage now to modal uniform notation, and we do not use it. Syntactically $\Box$, but not $\Diamond$, is added to the classical language. We present what are called *destructive* tableaus. The name comes from the fact that certain modal rules cause branch information to disappear. Such tableaus exist for K, T, D, D4, K4, S4, among others, but not for S5. Here we only give rules for S4.

Definition 1. *Let S be a set of signed formulas. $S^\sharp = \{T\,\Box X \mid T\,\Box X \in S\}$.*

The destructive tableau rules for S4 are the classical tableau rules together with those in Fig. 4. The first rule embodies reflexivity in an obvious way. The second rule is different, and is destructive, indicated by the double line. Suppose we have a branch containing $F\,\Box X$, with S being the set of other formulas on the branch. *The entire branch can be replaced with a new branch consisting of the members of $S^\sharp$, and $F\,X$.* Note that information is lost passing from S to $S^\sharp$, hence the name *destructive*.

$$\frac{T\,\Box X}{T\,X} \qquad \frac{S, F\,\Box X}{\overline{S^\sharp, F\,X}}$$

Fig. 4. S4 branch extension rules

With classical propositional tableaus, *any* order of rule application is acceptable. This is not the case for S4. If both $F\,\Box X$ and $F\,\Box Y$ are present, applying a rule to one eliminates the other, and it may be that only one of the two possibilities will lead to a proof. Now *backtracking* becomes critical to proof search.

Figure 5 shows a proof, using the S4 rules, of $\Box X \supset \Box(\Box X \vee Y)$. We have indicated branch replacement with horizontal lines. Lines 2 and 3 are from 1 by α. Next, the second S4 modal rule from Fig. 4 is applied to $F\,\Box(\Box X \vee Y)$, adding 4 while replacing S by $S^\sharp$ eliminates 1 and 3. Now an α-rule application to 4 adds 5 and 6, and produces a closed tableau, though not an atomically closed one. Continuing, we apply the S4 modal rule again, to 5, adding 7 while eliminating 4, 5, and 6. Applying the first S4 modal rule from Fig. 4 to 2 adds 8, and we have atomic closure.

1. $F\,\Box X \supset \Box(\Box X \vee Y)$
2. $T\,\Box X$
3. $F\,\Box(\Box X \vee Y)$

---

2. $T\,\Box X$
4. $F\,\Box X \vee Y$
5. $F\,\Box X$
6. $F\,Y$

---

2. $T\,\Box X$
7. $F\,X$
8. $T\,X$

Fig. 5. S4 proof of $\Box X \supset \Box(\Box X \vee Y)$

The rule $S, F\,\Box X \Rightarrow S^\sharp, F\,X$ is automatically single use since applying it with $F\,\Box X$ eliminates the formula. The rule $T\,\Box X \Rightarrow T\,X$ is trickier. For S4, if $T\,\Box X \Rightarrow T\,X$ is applied to a signed formula occurrence it need not be applied again, *until the rule* $S, F\,\Box X \Rightarrow S^\sharp, F\,X$ *has been applied.* The intuition is simple: the destructive rule might eliminate the conclusion of $T\,\Box X \Rightarrow T\,X$ but for S4 it will not eliminate the premise, so a new application may be useful.

4 Annotated Formulas and Tableaus

Mapping modal formulas to formulas of justification logic requires that we keep track of the *occurrences* of $\Box$. In [7] we introduced *annotated formulas* for this; we use a simpler version here.

Definition 2. *An* annotated modal formula *is like a standard modal formula, except that instead of a single modal operator* $\Box$ *there is an infinite family,* $\Box_1$*,* $\Box_2$*, ..., of* indexed *modal operators. In an annotated formula,* no index may occur twice.

If A *is an annotated formula and* A' *is the result of replacing all indexed modal operators,* $\Box_n$*, with* $\Box$*, regardless of index, we say* A *is an* annotated version *of* A'*, and* A' *is an* unannotated version *of* A.

Annotations are purely for bookkeeping purposes. The α/β classification is exactly as with unannotated formulas, as is the definition of components. For instance, $T\,\Box_1 P \wedge \Box_2 Q$ counts as an α, with $\alpha_1 = T\,\Box_1 P$ and $\alpha_2 = T\,\Box_2 Q$. In tableau constructions, branch extension rules apply to annotated formulas exactly as to unannotated ones. The annotated version of the $\sharp$ operation is $S^\sharp = \{T\,\Box_i X \mid T\,\Box_i X \in S\}$. Since we are requiring atomic closure, closure conditions are not affected by annotations.

Figure 6 is an annotated version of the proof shown in Fig. 5. Every S4 tableau proof can be turned into an annotated proof by annotating the modal operators appearing in the root, and then propagating these annotations downward through the tree.

1. $F\,\Box_1 X \supset \Box_2(\Box_3 X \vee Y)$
2. $T\,\Box_1 X$
3. $F\,\Box_2(\Box_3 X \vee Y)$

---

2. $T\,\Box_1 X$
4. $F\,\Box_3 X \vee Y$
5. $F\,\Box_3 X$
6. $F\,Y$

---

2. $T\,\Box_1 X$
7. $F\,X$
8. $T\,X$

Fig. 6. Annotated S4 proof of $\Box_1 X \supset \Box_2(\Box_3 X \vee Y)$

5 Changing the Tableau Representation

So far tableaus have been trees, and formula occurrences could be common to multiple branches. While this has advantages for some purposes, it does not when our Quasi-Realization algorithm is introduced. We will be associating a set of Quasi-Realizers with each signed formula occurrence in a tableau. How that is done depends on the history of the branch containing a given occurrence. If an occurrence is common to more than one branch, it is part of more than one history and things become ambiguous. Our solution is to change the way tableaus are represented, something that also brings us much closer to the data structure used in our Prolog implementation.

From now on a tableau is not a tree, but instead it is the set of its branches, where each branch is the set of signed formulas on it. When we write a branch as $\mathcal{B}, Z$, or more graphically $\begin{array}{c}\mathcal{B}\\ Z\end{array}$, we mean it is the set whose members are those of $\mathcal{B}$, together with signed formula Z. *This notation assumes that Z is not part of $\mathcal{B}$.* We reformulate the S4 tableau rules in this style, building in the notion of single-use for tableau rules. Here are the formal details, which apply equally well to tableaus of signed formulas or of annotated signed formulas.

Definition 3 (Classical Tableau Revised). *A classical tableau is a finite set of finite sets (called branches) of signed formulas. A branch is closed if $T\,P$ and $F\,P$ are members for some atomic P, or if $T\,\bot$ is a member. A tableau is closed if each of its branches is closed. We say a signed formula is on a branch if it is a member of it, and a branch is in a tableau if it is a member of it. A* tableau proof *of X is a sequence of tableaus, beginning with a single branch tableau where that branch contains only $F\,X$, continuing using the Branch Extension Rules given in* Figs. 7 and 8*, and ending with a closed tableau.*

$$
\frac{\begin{array}{c}\mathcal{B}\\ T\,\neg X\end{array}}{\begin{array}{c}\mathcal{B}\\ F\,X\end{array}}
\qquad
\frac{\begin{array}{c}\mathcal{B}\\ F\,\neg X\end{array}}{\begin{array}{c}\mathcal{B}\\ T\,X\end{array}}
\qquad
\frac{\begin{array}{c}\mathcal{B}\\ \alpha\end{array}}{\begin{array}{c}\mathcal{B}\\ \alpha_1\\ \alpha_2\end{array}}
\qquad
\frac{\begin{array}{c}\mathcal{B}\\ \beta\end{array}}{\begin{array}{c|c}\mathcal{B} & \mathcal{B}\\ \beta_1 & \beta_2\end{array}}
$$

Fig. 7. Classical branch extension rules revised

As an example, the β rule in Fig. 7 is to be read as follows. If a tableau has $\mathcal{B}, \beta$ as a branch, then the result of removing the branch from the tableau and replacing it with two branches, $\mathcal{B}, \beta_1$ and $\mathcal{B}, \beta_2$ is another tableau, which we call a *successor* of the original tableau. Similarly for the other rules. Note that the new branches do not contain β, but have β_1 and β_2 respectively instead. This is our general strategy for enforcing single use. That branches do not share any common parts is essential here.

There is one misleading aspect to the notation above. In the rule for $T\,\neg$ for instance, it may happen that $F\,X$ already occurs in $\mathcal{B}$, in which case the display of $\mathcal{B}, F\,X$ below the line is not correct—it should be simply $\mathcal{B}$. We allow this mild abuse, rather than complicating notation.

Single use is trickier for modal rules. As noted earlier, single use for the $F\,\Box$ rule is automatic, but for the $T\,\Box$ rule of S4 single use only applies until the next application of the $F\,\Box$ rule. We build this into Fig. 8 by crossing off an occurrence of $T\,\Box X$ when a rule has been applied to it, and providing no rule that has a crossed off signed formula as a trigger. A cross off mark is removed, as part of the definition of $\mathcal{B}^\sharp$, when an $F\,\Box$ rule is applied.

$$
\frac{\begin{array}{c}\mathcal{B}\\ T\,\Box X\end{array}}{\begin{array}{c}\mathcal{B}\\ \cancel{T\,\Box X}\\ T\,X\end{array}}
\qquad
\frac{\begin{array}{c}\mathcal{B}\\ F\,\Box X\end{array}}{\overline{\begin{array}{c}\mathcal{B}^\sharp\\ F\,X\end{array}}}
$$

where $\mathcal{B}^\sharp = \{T\,\Box X \mid T\,\Box X \in \mathcal{B} \text{ or } \cancel{T\,\Box X} \in \mathcal{B}\}$

Fig. 8. S4 Modal branch extension rules revised

1. $\{\{F\,\Box(P \supset Q) \supset (\Box P \supset \Box Q)\}\}$
2. $\{\{T\,\Box(P \supset Q), F\,\Box P \supset \Box Q\}\}$
3. $\{\{\cancel{T\,\Box(P \supset Q)}, T\,P \supset Q, F\,\Box P \supset \Box Q\}\}$
4. $\{\{\cancel{T\,\Box(P \supset Q)}, T\,P \supset Q, T\,\Box P, F\,\Box Q\}\}$
5. $\{\{T\,\Box(P \supset Q), T\,\Box P, F\,Q\}\}$
6. $\{\{\cancel{T\,\Box(P \supset Q)}, T\,P \supset Q, T\,\Box P, F\,Q\}\}$
7. $\{\{\cancel{T\,\Box(P \supset Q)}, F\,P, T\,\Box P, F\,Q\}, \{\cancel{T\,\Box(P \supset Q)}, T\,Q, T\,\Box P, F\,Q\}\}$
8. $\{\{\cancel{T\,\Box(P \supset Q)}, F\,P, \cancel{T\,\Box P}, T\,P, F\,Q\}, \{\cancel{T\,\Box(P \supset Q)}, T\,Q, T\,\Box P, F\,Q\}\}$

Fig. 9. Tableau as set of sets

Figure 9 shows a revised tableau proof of $\Box(P \supset Q) \supset (\Box P \supset \Box Q)$. It is not the shortest, by the way.

6 Quasi-Realizations

Our algorithm for computing Realizations divides into two halves. The first half constructs an intermediate object, a *Quasi-Realization*, from a tableau proof. The second half converts a Quasi-Realization into a proper Realization. The construction of Quasi-Realizations is logic dependent—we present an algorithm for S4 only. Input to this algorithm is a tableau proof. The input to the Quasi-Realization to Realization algorithm is a Quasi-Realization, not a formal proof, and the construction is independent of the particular logic involved. Because space is limited, the Quasi-Realization to Realization algorithm is not given here. It can be found in both [10,12].

Informally the goal is to associate a Quasi-Realization with each signed formula occurrence in a tableau proof. This Quasi-Realization is constructed according to how the branch on which the signed formula appears is continued to closure. A problem is that a particular signed formula occurrence can be on more than one branch, appearing before the branch splits. A Quasi-Realization computed on one branch might be different than a Quasi-Realization computed on another. If we were dealing with Realizations, a merging solution to this problem would involve the + operation and substitution, and would be of some complexity since the entire nested structure of the realizing formulas would need to be taken into consideration. Our approach here bypasses this problem by allowing a *set* of Quasi-Realizations rather than insisting on a single one. In fact + does not appear until we reach the Quasi-Realization to Realization algorithm.

From now on we assume that $v_1, v_2, \ldots$ is an enumeration of all justification variables of LP with no variable repeated, fixed once and for all. Case 4 of the definition below always uses v_k in Quasi-Realizations where the sign is T and $\Box_k$ is involved. It does not matter whether a formula $T\,\Box_n A$ occurs crossed out or not, and crossing out is suppressed in our current notation. In case 3 two conjunctive, or α, signed formulas are mentioned. For one we use α with α_1 and α_2 as components. For the other we use α' with α'_1 and α'_2 as components. Similarly for disjunctive, or β, signed formulas.

Definition 4 (Quasi-Realization Function). *The mapping* $\langle\langle\cdot\rangle\rangle$ *is defined recursively on the set of signed annotated modal formulas.*

1. *If* A *is atomic,* $\langle\langle T\,A\rangle\rangle = \{T\,A\}$ *and* $\langle\langle F\,A\rangle\rangle = \{F\,A\}$.
2. $\langle\langle T\,\neg A\rangle\rangle = \{T\,\neg U \mid F\,U \in \langle\langle F\,A\rangle\rangle\}$.
   $\langle\langle F\,\neg A\rangle\rangle = \{F\,\neg U \mid T\,U \in \langle\langle T\,A\rangle\rangle\}$.
3. $\langle\langle \alpha\rangle\rangle = \{\alpha' \mid \alpha'_1 \in \langle\langle \alpha_1\rangle\rangle$ *and* $\alpha'_2 \in \langle\langle \alpha_2\rangle\rangle\}$.
   $\langle\langle \beta\rangle\rangle = \{\beta' \mid \beta'_1 \in \langle\langle \beta_1\rangle\rangle$ *and* $\beta'_2 \in \langle\langle \beta_2\rangle\rangle\}$.
4. $\langle\langle T\,\Box_n A\rangle\rangle = \{T\,v_n{:}U \mid T\,U \in \langle\langle T\,A\rangle\rangle\}$.
   $\langle\langle F\,\Box_n A\rangle\rangle = \{F\,t : (U_1 \vee \ldots \vee U_k) \mid F\,U_1, \ldots, F\,U_k \in \langle\langle F\,A\rangle\rangle$ *and* t *is any justification term*$\}$.
5. *The mapping is extended to* sets *of signed annotated formulas by letting* $\langle\langle S\rangle\rangle = \cup\{\langle\langle Z\rangle\rangle \mid Z \in S\}$.

Members of $\langle\langle Z\rangle\rangle$ *are called* Quasi-Realizers *of* Z.

As an example, suppose t, u, and w are justification terms and P and Q are atomic formulas. Here are some Quasi-Realization calculations, leading up to $F\,\Box_1(\Box_2 P \vee \neg\Box_3 Q)$. We do not produce *all* Quasi-Realizations, an infinite set. Here's the reasoning for one case. $F\,\Box_2 P \vee \neg\Box_3 Q$, in item 5, is an α, with $\alpha_1 = F\,\Box_2 P$ and $\alpha_2 = F\,\neg\Box_3 Q$. By items 2 and 4, we can take $\alpha'_1 = F\,t{:}P$ and $\alpha'_2 = F\,\neg v_3{:}Q$, and then $\alpha' = F\,t{:}P \vee \neg v_3{:}Q$, which is taken to be one of the members of $\langle\langle F\,\Box_2 P \vee \neg\Box_3 Q\rangle\rangle$.

1. $\{F\,P\} = \langle\langle F\,P\rangle\rangle$ and $\{T\,Q\} = \langle\langle T\,Q\rangle\rangle$
2. $\{F\,t{:}P, F\,u{:}P\} \subseteq \langle\langle F\,\Box_2 P\rangle\rangle$
3. $\{T\,v_3{:}Q\} = \langle\langle T\,\Box_3 Q\rangle\rangle$
4. $\{F\,\neg v_3{:}Q\} = \langle\langle F\,\neg\Box_3 Q\rangle\rangle$
5. $\{F\,t{:}P \vee \neg v_3{:}Q, F\,u{:}P \vee \neg v_3{:}Q\} \subseteq \langle\langle F\,\Box_2 P \vee \neg\Box_3 Q\rangle\rangle$
6. $\{F\,t{:}((t{:}P \vee \neg v_3{:}Q) \vee (u{:}P \vee \neg v_3{:}Q)), F\,w{:}(u{:}P \vee \neg v_3{:}Q)\} \subseteq \langle\langle F\,\Box_1(\Box_2 P \vee \neg\Box_3 Q)\rangle\rangle$

We can now formulate the main contribution of this paper. It will be proved in Sects. 7 and 9 using an algorithm given in Sect. 8.

Theorem 3. *Let* X *be an annotated modal formula. Given a tableau proof of* X *in* S4*, a finite set* $\{F\,Q_1, \ldots, F\,Q_k\}$ *of quasi-realizers for* $F\,X$ *can be constructed so that* $Q_1 \vee \ldots \vee Q_k$ *is a theorem of* LP.

7 Mixed Tableaus

We now introduce what we call mixed tableaus, which unite modal features with justification logic features. They are based on tableaus as defined in Sect. 5, using a set of sets representation. Informally, a mixed tableau expands an S4 tableau by associating a set of Quasi-Realizers to each signed annotated modal formula appearing in it.

Definition 5 (Mixed Tableau). *A* mixed $\mathsf{S4}$ tableau *is like a tableau except that members of branches are pairs* (M, S) *where* M *is a signed annotated modal formula and* S *is a finite, non-empty set of signed justification formulas, meeting the following requirements.*

1. *If* (M, S) *occurs in a mixed tableau, it is required that* $S \subseteq \langle\!\langle M \rangle\!\rangle$.
2. *If, in a mixed tableau, we replace each entry* (M, S) *by just* M*, the result must be an annotated* $\mathsf{S4}$ *tableau.*

In a mixed tableau, we refer to M *as the* modal part *of* (M, S)*, and to* S *as the* justification part *of* (M, S)*. We say a mixed tableau* $\mathcal{T}^{mix}$ *is an* expansion *of an* $\mathsf{S4}$ *tableau* $\mathcal{T}$ *if* $\mathcal{T}$ *results from* $\mathcal{T}^{mix}$ *by eliminating the justification parts of node labels, as in item 2 of Definition 5.*

Remark: A referee for this paper observed that while an empty S cannot produce a justification-sound mixed tableau, allowing it leaves open the possibility of using it for trivial expansions. We have not explored this suggestion, but believe it might plausibly make a simplification in the presentation of the main algorithm of this paper.

Definition 6 (Justification Sound). *Let* $\mathcal{B}$ *be a branch of a mixed tableau. By the* associated justification formula *for* $\mathcal{B}$ *we mean* $\bigwedge \mathcal{B}_T^{just} \supset \bigvee \mathcal{B}_F^{just}$ *where* $\mathcal{B}_T^{just}$ *is the set of all justification formulas* X *such that* $T\,X$ *occurs in the justification part of some member of* $\mathcal{B}$ *and* $\mathcal{B}_F^{just}$ *is the set of* X *such that* $F\,X$ *occurs in the justification part of some member of* $\mathcal{B}$.

We say a mixed $\mathsf{S4}$ *tableau branch is* justification sound *provided that its associated justification formula is provable in axiomatic* LP*. We say a mixed* $\mathsf{S4}$ *tableau is justification sound if each branch is.*

The heart of our Quasi-Realization work is the following theorem, proved in Sect. 9, which immediately gives us a proof of Theorem 3.

Theorem 4. *Let* $\mathcal{T}$ *be an annotated* $\mathsf{S4}$ *tableau that can be continued to one that is closed (or is closed already). Then* $\mathcal{T}$ *has a mixed tableau expansion* $\mathcal{T}^{mix}$ *that is justification sound, where* $\mathcal{T}^{mix}$ *can be algorithmically constructed from any closed modal tableau extending* $\mathcal{T}$.

Proof (of Theorem 3). Suppose X is an annotated modal formula, and we have a closed $\mathsf{S4}$ tableau proof for X. The construction of that proof begins with the single-branch modal tableau consisting of just a root node, labeled $F\,X$. Since this trivial tableau can be continued to a closed tableau, by Theorem 4 it can be expanded to a mixed tableau that is justification sound. Such a mixed tableau must consist of just a root node, labeled $(F\,X, \{F\,Q_1, \ldots, F\,Q_k\})$, where $\{F\,Q_1, \ldots, F\,Q_k\} \subseteq \langle\!\langle F\,X \rangle\!\rangle$. Since this expanded tableau is justification sound, the formula $\bigwedge \emptyset \supset \bigvee \{Q_1, \ldots, Q_k\}$ is axiomatically LP provable. That is, $Q_1 \vee \ldots \vee Q_k$ is provable, where $F\,Q_1, \ldots, F\,Q_k$ are quasi-realizers for $F\,X$.

8 The Quasi-Realization Algorithm

Figure 10 contains an algorithm to construct justification-sound mixed tableaus from closed S4 tableaus, followed by an example in Figs. 11 and 12. In Sect. 9 a proof of the correctness of the algorithm is given, and this establishes Theorem 4. The construction is a kind of 'backward induction'. Suppose $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k$ is a sequence of annotated S4 tableaus, in which each arises from the preceding by a single application of an S4 branch extension rule, as given in Sect. 5. Suppose also that $\mathcal{T}_k$ is closed. We show $\mathcal{T}_k$ has a mixed tableau expansion that is justification sound. Then, using this, we show the same for $\mathcal{T}_{k-1}$, then for $\mathcal{T}_{k-2}$, and so on back to $\mathcal{T}_1$. A bit more properly, the algorithm produces a mixed tableau expansion for each $\mathcal{T}_i$; the correctness proof in the following section shows that it must be justification sound.

A branch extension rule application modifies only one branch—all others remain unchanged. Consequently the algorithm is stated in terms of branch extension rules applied to single branches. The rest of the mixed tableau being constructed does not change, so unaffected branches are not explicitly displayed.

S4 tableaus are understood as sets of branches, with branches being sets of signed annotated formulas, as in Sect. 5. We make use of the notion convention introduced there, where $\mathcal{B}, Z$, or $\overset{\mathcal{B}}{Z}$, is a branch consisting of the members of $\mathcal{B}$, and Z (which is understood not to occur in $\mathcal{B}$).

The idea is to *expand* branches of an annotated S4 tableau so they become branches of a mixed tableau. If $\mathcal{B}$ is an S4 tableau branch, we will write $\mathcal{B}^E$ to denote an expansion of it to a mixed tableau branch. Each signed annotated formula M in $\mathcal{B}$ is transformed into a pair (M, S) in $\mathcal{B}^E$ so that $S \subseteq \langle\!\langle M \rangle\!\rangle$. Of course $\mathcal{B}^E$ is not unique—it is simply some expansion. In one case of the algorithm more than one branch expansion must be referenced, and we use $\mathcal{B}^{E_1}$ and $\mathcal{B}^{E_2}$ as notation. We write $\mathcal{B} \xrightarrow{exp} \mathcal{B}^E$ to indicate that annotated S4 tableau branch $\mathcal{B}$ expands to mixed tableau branch $\mathcal{B}^E$.

If $\mathcal{B}^{E_1}$ and $\mathcal{B}^{E_2}$ are both expansions of the same branch, $\mathcal{B}$, by $\mathcal{B}^{E_1} \dot{\cup} \mathcal{B}^{E_2}$ we mean the mixed tableau branch consisting of all $(M, S_1 \cup S_2)$ where $(M, S_1) \in \mathcal{B}^{E_1}$ and $(M, S_2) \in \mathcal{B}^{E_2}$.

In a few of the algorithm cases we refer to a *trivial expansion.* A trivial expansion of a signed formula M is (M, S) where S is *any* finite set such that $S \subseteq \langle\!\langle M \rangle\!\rangle$. A trivial expansion of an S4 branch replaces each member with a trivial expansion. In our Prolog implementation a particular easily computed trivial expansion is used, but the details don't matter here.

The algorithm is stated schematically below. We give a reading of the α Case as a representative example of how the algorithm notation should be understood. The idea is, we say how to expand the S4 tableau branch $\mathcal{B}, \alpha$ provided we already know how to expand $\mathcal{B}, \alpha_1, \alpha_2$. So, assume we have an expansion for S4 tableau branch $\mathcal{B}, \alpha_1, \alpha_2$, where $\mathcal{B}$ expands to $\mathcal{B}^E$, α_1 expands to (α_1, S_1), and α_2 expands to (α_2, S_2). Then S4 tableau branch $\mathcal{B}, \alpha$ expands to $\mathcal{B}^E, (\alpha, S)$, where S consists of all α signed formulas for which $\alpha_1 \in S_1$ and $\alpha_2 \in S_2$. (In the schematic we used α', α'_1, and α'_2 in characterizing S, simply because α, α_1, and α_2 were already in use to designate members of $\mathsf{S4}$ tableau branches.)

Recall that $v_1, v_2, \ldots$ is a fixed enumeration of all justification variables of LP with no variable repeated.

In the $F\,\Box$ case of Fig. 10, $\bigwedge \mathcal{A} \supset \bigvee S$ appears as the associated justification formula for the branch $(\mathcal{B}^\sharp)^E, (F\,X, S_0)$. In fact, $\mathcal{A} = ((\mathcal{B}^\sharp)^E)^{just}_T$, using the notation of Definition 6, but such detailed notation distracts from the basic idea and we have suppressed it here. The existence of a term t such that $\vdash_{\mathsf{LP}} \bigwedge \mathcal{A} \supset t{:}\bigvee S$ will be guaranteed by the Lifting Lemma. Also note that the combination $\mathcal{B}^\sharp$ and $\mathcal{B} - \mathcal{B}^\sharp$ below the line simply amounts to $\mathcal{B}$, though the separation is useful for our purposes.

We give an example to illustrate how the Quasi-Realization Algorithm works. Figure 11 shows an $\mathsf{S4}$ proof of the annotated formula $(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B)$ where A and B are atomic, using the representation of tableaus described in Sect. 5. Each numbered item should be thought of as the set of signed formulas making up a tableau branch. A detailed description follows. 1 is the initial single-branch tableau. Single-branch tableau 2 follows from 1 by α. A β rule application creates a tableau with two branches, 3 and 4. Modal rule applications on $F\,\Box_3(A \vee B)$ in 3 and 4 produce the two-branched tableau having branches 5 and 6. Modal rule applications on $T\,\Box_1 A$ and $T\,\Box_2 B$ in these give the two branches 7 and 8. Finally, α rule applications give the two branches 9 and 10, both of which are atomically closed.

Next, the proof created in Fig. 11 is converted to a mixed tableau, displayed in Fig. 12. The work is from bottom up. In Fig. 11, 9 is an atomically closed branch. For this, the algorithm makes use of a *trivial* expansion, giving the corresponding 9 of Fig. 12, and similarly for 10. Branch 7 in Fig. 11 yields branch 9 by an α rule. Since 9 in Fig. 11 expands to 9 in Fig. 12, 7 of Fig. 11 converts to 7 of Fig. 12 by the α case of the Algorithm. Similarly for 8 and 10. Then branch 5 of Fig. 11 converts to 5 of Fig. 12 because of the 7 conversion, and the $T\,\Box$ case of the Algorithm, and similarly for 6 and 8. Branch 3 of Fig. 11 yields branch 5 by the $F\Box$ rule. The associated justification formula for branch 5 is $v_1{:}A \supset (A \vee B)$. Justification term t, in 3, is such that $v_1 : A \supset t : (A \vee B)$ is provable in LP. Existence is guaranteed by the Lifting Lemma 2. Similarly u in branch 4 is such that $v_2{:}B \supset u{:}(A \vee B)$ is provable in LP. Branch 2 yields branches 3 and 4 using the β rule. Note that in branch 2 in Fig. 12, the justification part associated with $F\,\Box_3(A \vee B)$ is the union of those parts from branches 3 and 4. Finally 1 is a straightforward application of the α rule.

Then, according to the algorithm, $\{F\,(v_1{:}A \vee v_2{:}B) \supset t{:}(A \vee B), F\,(v_1{:}A \vee v_2{:}B) \supset u{:}(A \vee B)\}$ is a set of quasi-realizers for $F\,(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B)$. In fact, the following is provable in LP.

$$\bigvee\{(v_1{:}A \vee v_2{:}B) \supset t{:}(A \vee B), (v_1{:}A \vee v_2{:}B) \supset u{:}(A \vee B)\}$$

Atomic Cases

$$\begin{array}{ccc} \mathcal{B} & & \mathcal{B}^E \\ T\,P & \xrightarrow{exp} & (T\,P,\{T\,P\}) \\ F\,P & & (F\,P,\{F\,P\}) \end{array} \quad \text{where } \mathcal{B}^E \text{ trivially expands } \mathcal{B}$$

$$\begin{array}{ccc} \mathcal{B} & & \mathcal{B}^E \\ T\,\bot & \xrightarrow{exp} & (T\,\bot,\{T\,\bot\}) \end{array} \quad \text{where } \mathcal{B}^E \text{ trivially expands } \mathcal{B}$$

α Cases

$$\dfrac{\begin{array}{ccc} \mathcal{B} & & \mathcal{B}^E \\ \alpha_1 & \xrightarrow{exp} & (\alpha_1, S_1) \\ \alpha_2 & & (\alpha_2, S_2) \end{array}}{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ \alpha & & (\alpha, S) \end{array}} \quad \text{where } S = \{\alpha' \mid \alpha_1' \in S_1 \text{ and } \alpha_2' \in S_2\}$$

β Cases

$$\dfrac{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^{E_1} \\ \beta_1 & & (\beta_1, S_1) \end{array} \qquad \begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^{E_2} \\ \beta_2 & & (\beta_2, S_2) \end{array}}{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ \beta & & (\beta, S) \end{array}} \quad \begin{array}{l} \text{where } S = \{\beta' \mid \beta_1' \in S_1 \\ \text{and } \beta_2' \in S_2\} \text{ and } \mathcal{B}^E = \mathcal{B}^{E_1} \dot\cup \mathcal{B}^{E_2} \end{array}$$

Negation Cases

$$\dfrac{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ F\,X & & (F\,X, S_0) \end{array}}{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ T\,\neg X & & (T\,\neg X, S) \end{array}} \quad \text{where } S = \{T\,\neg Z \mid F\,Z \in S_0\}$$

$$\dfrac{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ T\,X & & (T\,X, S_0) \end{array}}{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ F\,\neg X & & (F\,\neg X, S) \end{array}} \quad \text{where } S = \{F\,\neg Z \mid T\,Z \in S_0\}$$

$T\,\Box$ Case

$$\dfrac{\begin{array}{ccc} \mathcal{B} & & \mathcal{B}^E \\ \cancel{T\,\Box_k X} & \xrightarrow{exp} & (\cancel{T\,\Box_k X}, S_0) \\ T\,X & & (T\,X, S_1) \end{array}}{\begin{array}{ccc} \mathcal{B} & \xrightarrow{exp} & \mathcal{B}^E \\ T\,\Box_k X & & (T\,\Box_k X, S) \end{array}} \quad \text{where } S = S_0 \cup \{T\,v_k{:}Z \mid TZ \in S_1\}$$

$F\,\Box$ Case

$$\dfrac{\begin{array}{ccc} \mathcal{B}^\sharp & \xrightarrow{exp} & (\mathcal{B}^\sharp)^E \\ F\,X & & (F\,X, S_0) \end{array}}{\begin{array}{ccc} \mathcal{B}^\sharp & & (\mathcal{B}^\sharp)^E \\ \mathcal{B} - \mathcal{B}^\sharp & \xrightarrow{exp} & (\mathcal{B} - \mathcal{B}^\sharp)^E \\ F\,\Box_n X & & (F\,\Box_n X, \{F\,t{:}\bigvee S\}) \end{array}} \quad \begin{array}{l} \text{where } (\mathcal{B} - \mathcal{B}^\sharp)^E \text{ trivially expands } \mathcal{B} - \mathcal{B}^\sharp, \\ S = \{Z \mid F\,Z \in S_0\}, \\ \bigwedge \mathcal{A} \supset \bigvee S \text{ is the associated justification} \\ \text{formula for branch } (\mathcal{B}^\sharp)^E, (F\,X, S_0) \\ \text{and } \vdash_{\mathsf{LP}} \bigwedge \mathcal{A} \supset t{:}\bigvee S \end{array}$$

Fig. 10. Quasi-Realization Algorithm

1. $F\,(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B)$

2. $T\,\Box_1 A \vee \Box_2 B$
   $F\,\Box_3(A \vee B)$

3. $T\,\Box_1 A$
   $F\,\Box_3(A \vee B)$

4. $T\,\Box_2 B$
   $F\,\Box_3(A \vee B)$

5. $T\,\Box_1 A$
   $F\,A \vee B$

6. $T\,\Box_2 B$
   $F\,A \vee B$

7. $\cancel{T\,\Box_1 A}$
   $F\,A \vee B$
   $T\,A$

8. $\cancel{T\,\Box_2 B}$
   $F\,A \vee B$
   $T\,B$

9. $\cancel{T\,\Box_1 A}$
   $T\,A$
   $F\,A$
   $F\,B$

10. $\cancel{T\,\Box_2 B}$
    $T\,B$
    $F\,A$
    $F\,B$

Fig. 11. S4 tableau proof (to be expanded)

1. $(F\,(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B), \{F\,(v_1{:}A \vee v_2{:}B) \supset t{:}(A \vee B),$
   $F\,(v_1{:}A \vee v_2{:}B) \supset u{:}(A \vee B)\})$

2. $(T\,\Box_1 A \vee \Box_2 B, \{T\,v_1{:}A \vee v_2{:}B\})$
   $(F\,\Box_3(A \vee B), \{F\,t{:}(A \vee B), F\,u{:}(A \vee B)\})$

3. $(T\,\Box_1 A, \{T\,v_1{:}A\})$
   $(F\,\Box_3(A \vee B), \{F\,t{:}(A \vee B)\})$

4. $(T\,\Box_2 B, \{T\,v_2{:}B\})$
   $(F\,\Box_3(A \vee B), \{F\,u{:}(A \vee B)\})$

5. $(T\,\Box_1 A, \{T\,v_1{:}A\})$
   $(F\,A \vee B, \{F\,A \vee B\})$

6. $(T\,\Box_2 B, \{T\,v_2{:}B\})$
   $(F\,A \vee B, \{F\,A \vee B\})$

7. $(\cancel{T\,\Box_1 A}, \{T\,v_1{:}A\})$
   $(F\,A \vee B, \{F\,A \vee B\})$
   $(T\,A, \{T\,A\})$

8. $(\cancel{T\,\Box_2 B}, \{T\,v_2{:}B\})$
   $(F\,A \vee B, \{F\,A \vee B\})$
   $(T\,B, \{T\,B\})$

9. $(\cancel{T\,\Box_1 A}, \{T\,v_1{:}A\})$
   $(T\,A, \{T\,A\})$
   $(F\,A, \{F\,A\})$
   $(F\,B, \{F\,B\})$

10. $(\cancel{T\,\Box_2 B}, \{T\,v_2{:}B\})$
    $(T\,B, \{T\,B\})$
    $(F\,A, \{F\,A\})$
    $(F\,B, \{F\,B\})$

Justification term t, in 3, is such that $v_1{:}A \supset t{:}(A \vee B)$ is provable in LP. Similarly u in 4 is such that $v_2{:}B \supset u{:}(A \vee B)$ is LP provable.

Fig. 12. S4 tableau proof (expanded)

9 Quasi-Realization Algorithm Correctness Proof

This section is devoted to showing the correctness of the Quasi-Realization Algorithm, and hence proving Theorem 4. It is straightforward that the algorithm produces a mixed tableau expansion. We concentrate on showing the resulting mixed tableau must be justification sound, Definition 6. To do this, we show it for the Atomic Cases, and show that each rule of the algorithm preserves justification soundness.

Proof (Correctness for Quasi-Realization Algorithm)

Atomic Cases. Consider the first of the two atomic cases—the second is similar. The mixed tableau branch produced is $\mathcal{B}^E, (T\,P, \{T\,P\}), (F\,P, \{F\,P\})$. The associated justification formula is $[\bigwedge(\mathcal{B}^E)^{just}_T \wedge P] \supset [\bigvee(\mathcal{B}^E)^{just}_F \vee P]$, and this is trivially an LP theorem, so the branch is justification sound.

α **Case.** Assume that $\mathcal{B}^E, (\alpha_1, S_1), (\alpha_2, S_2)$ is a mixed tableau branch that is justification sound. We must show the same for $\mathcal{B}^E, (\alpha, S)$ where $S = \{\alpha' \mid \alpha'_1 \in S_1 \text{ and } \alpha'_2 \in S_2\}$. Since $S_1 \subseteq \langle\!\langle \alpha_1 \rangle\!\rangle$ and $S_2 \subseteq \langle\!\langle \alpha_2 \rangle\!\rangle$, it is easy to see from Definition 4 that $S \subseteq \langle\!\langle \alpha \rangle\!\rangle$. After this case we leave such arguments to the reader. We must show the branch is justification sound.

Since we only consider $\wedge$, $\vee$, and $\supset$, there are three possibilities for α. We look at one of them, with $\alpha = F\,A \supset B$; the other two cases are similar. All three could be condensed into a single argument by making use of uniform notation, but this would be a bit of a diversion just now. So, assume $\mathcal{B}^E, (T\,A, S_1), (F\,B, S_2)$ is justification sound; we show the same for $\mathcal{B}^E, (F\,A \supset B, S)$.

Let us say $S_1 = \{T\,A_1, \ldots, T\,A_m\}$ and $S_2 = \{F\,B_1, \ldots, F\,B_n\}$. Then the associated justification formula for $\mathcal{B}^E, (T\,A, S_1), (F\,B, S_2)$ is the following.

$$\left[\bigwedge(\mathcal{B}^E)^{just}_T \wedge \bigwedge\{A_1, \ldots, A_m\}\right] \supset \left[\bigvee(\mathcal{B}^E)^{just}_F \vee \bigvee\{B_1, \ldots, B_n\}\right]$$

By classical logic we also have provability of the following, where i ranges over $1, \ldots, m$ and j ranges over $1, \ldots, n$.

$$\bigwedge(\mathcal{B}^E)^{just}_T \supset \left[\bigvee(\mathcal{B}^E)^{just}_F \vee \bigvee_{i,j}(A_i \supset B_j)\right]$$

Thus the associated justification formula for $\mathcal{B}, (F\,A \supset B, S)$ is provable.

β **Case.** Assume that $\mathcal{B}^{E_1}, (\beta_1, S_1)$ and $\mathcal{B}^{E_2}, (\beta_2, S_2)$ are justification sound. We show this also to be the case for $\beta^E, (\beta, S)$, where $S = \{\beta' \mid \beta'_1 \in S_1 \text{ and } \beta'_2 \in S_2\}$ and $\mathcal{B}^E = \mathcal{B}^{E_1} \dot\cup \mathcal{B}^{E_2}$. As with α there are three cases, and we only consider one of them, where $\beta = T\,A \supset B$. So, assume that $\mathcal{B}^{E_1}, (F\,A, S_1)$ and $\mathcal{B}^{E_2}, (T\,B, S_2)$ are justification sound.

Suppose $S_1 = \{F\,A_1, \ldots, F\,A_m\}$ and $S_2 = \{T\,B_1, \ldots, T\,B_n\}$. Then the provable associated justification formulas for $\mathcal{B}^{E_1}, (F\,A, S_1)$ and $\mathcal{B}^{E_2}, (T\,B, S_2)$ are the following.

$$\bigwedge(\mathcal{B}^{E_1})_T^{just} \supset \left[\bigvee(\mathcal{B}^{E_1})_F^{just} \vee \bigvee\{A_1, \ldots, A_m\}\right]$$

$$\left[\bigwedge(\mathcal{B}^{E_2})_T^{just} \wedge \bigwedge\{B_1, \ldots, B_n\}\right] \supset \bigvee(\mathcal{B}^{E_2})_F^{just}$$

$\mathcal{B}^E = \mathcal{B}^{E_1} \dot{\cup} \mathcal{B}^{E_2}$, and it follows easily that $(\mathcal{B}^E)_T^{just} = (\mathcal{B}^{E_1})_T^{just} \cup (\mathcal{B}^{E_2})_T^{just}$ and $(\mathcal{B}^E)_F^{just} = (\mathcal{B}^{E_1})_F^{just} \cup (\mathcal{B}^{E_2})_F^{just}$. Then we have provability of the following.

$$\bigwedge(\mathcal{B}^{E})_T^{just} \supset \left[\bigvee(\mathcal{B}^{E})_F^{just} \vee \bigvee\{A_1, \ldots, A_m\}\right]$$

$$\left[\bigwedge(\mathcal{B}^{E})_T^{just} \wedge \bigwedge\{B_1, \ldots, B_n\}\right] \supset \bigvee(\mathcal{B}^{E})_F^{just}$$

By classical logic this gives provability of the following, where i ranges over $1, 2, \ldots, m$ and j ranges over $1, 2, \ldots, n$.

$$\left[\bigwedge(\mathcal{B}^{E})_T^{just} \wedge \bigwedge_{i,j}(A_i \supset B_j)\right] \supset \bigvee(\mathcal{B}^{E})_F^{just}$$

Thus the associated justification formula for $\mathcal{B}, (T\, A \supset B, S)$ is provable.

Negation Cases. These cases are similar to the α and β cases, but are simpler and are left to the reader.

$T\,\Box$ **Case.** Assume that $\mathcal{B}^E, (\cancel{T\,\Box_k X}, S_0), (T\,X, S_1)$ is a justification-sound mixed tableau branch. Then $\mathcal{B}^E, (T\,\Box_k X, S)$ is a mixed tableau branch, where $S = S_0 \cup \{T\,v_k{:}Z \mid T Z \in S_1\}$. We show it is justification sound.

Suppose $S_0 = \{T\,v_k{:}W_1, \ldots, T\,v_k{:}W_m\}$ and $S_1 = \{T\,Z_1, \ldots, T\,Z_h\}$. Then the provable associated justification formula for $\mathcal{B}^E, (\cancel{T\,\Box_k X}, S_0), (T\,X, S_1)$ is the following.

$$\left[\bigwedge(\mathcal{B}^{E})_T^{just} \wedge \bigwedge\{v_k{:}W_1, \ldots, v_k{:}W_m\} \wedge \bigwedge\{Z_1, \ldots, Z_h\}\right] \supset \bigvee(\mathcal{B}^{E})_F^{just}$$

Using Factivity, Axiom A2, we have LP provability of the following.

$$\left[\bigwedge(\mathcal{B}^{E})_T^{just} \wedge \bigwedge\{v_k{:}W_1, \ldots, v_k{:}W_m, v_k{:}Z_1, \ldots, v_k{:}Z_h\}\right] \supset \bigvee(\mathcal{B}^{E})_F^{just}$$

This is the associated justification formula for $\mathcal{B}^E, (T\,\Box_k X, S)$.

$F\,\Box$ **Case.** Assume that $(\mathcal{B}^\sharp)^E, (F\,X, S_0)$ is a justification-sound mixed tableau branch. Then $(\mathcal{B}^\sharp)^E, (\mathcal{B}-\mathcal{B}^\sharp)^E, (F\,\Box_n X, \{F\,t{:}\bigvee S\})$ is a mixed tableau branch, where $(\mathcal{B} - \mathcal{B}^\sharp)^E$ trivially expands $\mathcal{B} - \mathcal{B}^\sharp$, $S = \{Z \mid F\,Z \in S_0\}$, and t is any justification term. We show $(\mathcal{B}^\sharp)^E, (\mathcal{B} - \mathcal{B}^\sharp)^E, (F\,\Box_n X, \{F\,t{:}\bigvee S\})$ is justification sound, given the right choice of t.

Note that since all members of $\mathcal{B}^\sharp$ are T-signed, the LP-provable associated justification formula for $(\mathcal{B}^\sharp)^E, (F\,X, S_0)$ is simply $\bigwedge((\mathcal{B}^\sharp)^E)_T^{just} \supset \bigvee S$, where $S = \{Z \mid F\,Z \in S_0\}$. Also members of $\mathcal{B}^\sharp$ are necessitated, so by the Lifting Lemma 2, for some justification term t, $\vdash_{\mathsf{LP}} \bigwedge((\mathcal{B}^\sharp)^E)_T^{just} \supset t{:}\bigvee S$. Then, trivially, the following is also LP-provable

$$\left[\bigwedge((\mathcal{B}^\sharp)^E)_T^{just} \wedge \bigwedge((\mathcal{B} - \mathcal{B}^\sharp)^E)_T^{just}\right] \supset \left[t{:}\bigvee S \vee \bigvee((\mathcal{B} - \mathcal{B}^\sharp)^E)_F^{just}\right]$$

and this is the associated justification formula for $(\mathcal{B}^\sharp)^E, (\mathcal{B} - \mathcal{B}^\sharp)^E$, $(F\,\Box_n X, \{F\,t{:}\bigvee S\})$ as specified by the algorithm.

10 Realizations

Quasi-Realizations convert to Realizations. There is an algorithm for doing this in [12] that does not depend on tableau proofs, but only on the structure of a Quasi-Realization formula. It applies uniformly to a wide range of justification logics, not just to LP. Because of space limitations we omit the algorithm, and just state what it gives us.

We begin with a definition of Realization equivalent to the usual one, but following the lines of Definition 4. Differences are confined to case 4 where a disjunction appearing in the definition of quasi-realizer is folded into a justification term by using the + operator. We still assume that $v_1, v_2, \ldots$ is an enumeration of all justification variables of LP, with no justification variable repeated.

Definition 7. *The mapping* $[\![\cdot]\!]$ *is defined recursively on the set of signed annotated modal formulas.*

1. *If* A *is atomic,* $[\![T\,A]\!] = \{T\,A\}$ *and* $[\![F\,A]\!] = \{F\,A\}$.
2. $[\![T\,\neg A]\!] = \{T\,\neg U \mid F\,U \in [\![F\,A]\!]\}$.
   $[\![F\,\neg A]\!] = \{F\,\neg U \mid T\,U \in [\![T\,A]\!]\}$.
3. $[\![\alpha]\!] = \{\alpha' \mid \alpha'_1 \in [\![\alpha_1]\!]$ *and* $\alpha'_2 \in [\![\alpha_2]\!]\}$.
   $[\![\beta]\!] = \{\beta' \mid \beta'_1 \in [\![\beta_1]\!]$ *and* $\beta'_2 \in [\![\beta_2]\!]\}$.
4. $[\![T\,\Box_n A]\!] = \{T\,v_n{:}U \mid T\,U \in [\![T\,A]\!]\}$.
   $[\![F\,\Box_n A]\!] = \{F\,t{:}U \mid F\,U \in [\![F\,A]\!]$ *and* t *is any justification term*$\}$.
5. *The mapping is extended to* sets *of signed annotated formulas by letting* $[\![S]\!] = \cup\{[\![Z]\!] \mid Z \in S\}$.

Members of $[\![Z]\!]$ *are* Realizers *of* Z*, where* Z *is a signed, annotated modal formula. A* normal Realization *of annotated modal* A *is any justification formula* U *where* $F\,U \in [\![F\,A]\!]$*. For a modal formula* A *without annotations, a normal Realization for* A *is any normal Realization for* A'*, where* A' *is an annotated version of* A*.*

Substitutions are fundamental. A substitution σ replaces justification variables with justification terms. For a justification formula A the result of applying a substitution σ is denoted $A\sigma$. It is easy to show that substitutions turn LP theorems into LP theorems, though generally the constant specification will change. Substitution σ meets the *no new variable* condition if, for every v_k in the domain of σ, the justification term $v_k\sigma$ contains no variables other than v_k. σ *lives on* an annotated modal formula A if, for every justification variable v_k in the domain of σ, $\Box_k$ occurs in A.

Definition 8. *Let* A *be an annotated modal formula,* $\mathcal{A}$ *be a set of justification formulas,* A' *be a single justification formula, and* σ *be a substitution.*

1. $\mathcal{A} \xrightarrow{T\,A} (A', \sigma)$ *means:* σ *lives on* A *and meets the no new variable condition;* $T\,\mathcal{A} \subseteq \langle\!\langle T\,A \rangle\!\rangle$*;* $T\,A' \in [\![T\,A]\!]$*; and* $\vdash_{LP} A' \supset (\bigwedge \mathcal{A})\sigma$.
2. $\mathcal{A} \xrightarrow{F\,A} (A', \sigma)$ *means:* σ *lives on* A *and meets the no new variable condition;* $F\,\mathcal{A} \subseteq \langle\!\langle F\,A \rangle\!\rangle$*;* $F\,A' \in [\![F\,A]\!]$*; and* $\vdash_{LP} (\bigvee \mathcal{A})\sigma \supset A'$.

One can read $\mathcal{A} \xrightarrow{T\,A} (A', \sigma)$ as saying that the set of quasi-realizers $\mathcal{A}$ for $T\,A$ *condenses* to the single realizer $T\,A'$ using substitution σ, and similarly for $\mathcal{A} \xrightarrow{F\,A} (A', \sigma)$.

Theorem 5 (Condensing). *Let A be an annotated modal formula. For each finite set $\mathcal{A}$ of justification formulas:*

1. *If $T\,\mathcal{A} \subseteq \langle\langle T\,A \rangle\rangle$ then there are A' and σ so that $\mathcal{A} \xrightarrow{T\,A} (A', \sigma)$.*
2. *If $F\,\mathcal{A} \subseteq \langle\langle F\,A \rangle\rangle$ then there are A' and σ so that $\mathcal{A} \xrightarrow{F\,A} (A', \sigma)$.*

As has been said several times, the proof of Theorem 5 is algorithmic, and [12] can be consulted for details.

Corollary 2 (Realization). *Every formula provable in* S4 *has a normal Realization that is provable in* LP.

Proof. Suppose X is a theorem of S4. Let A be an annotated version of X, any one will do. Then from Theorem 3, proved using the algorithm given in Fig. 10, there are $Q_1, \ldots, Q_k$ with $\{F\,Q_1, \ldots, F\,Q_k\} \subseteq \langle\langle F\,A \rangle\rangle$ such that $Q_1 \vee \ldots \vee Q_k$ is a theorem of LP. By part 2 of Theorem 5 there is a substitution σ and a formula A' with $F\,A' \in [\![F\,A]\!]$ such that $(Q_1 \vee \ldots \vee Q_k)\sigma \supset A'$ is a theorem of LP. Since $(Q_1 \vee \ldots \vee Q_k)\sigma$ must also be provable in LP, so is A', and this is a normal Realization of A, and hence of X.

At the end of Sect. 8 we presented an example showing that the annotated modal formula $(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B)$, provable in S4, has the following Quasi-Realization set, $\{(v_1 : A \vee v_2 : B) \supset t : (A \vee B), (v_1 : A \vee v_2 : B) \supset u : (A \vee B)\}$, where $\vdash_{\mathsf{LP}} v_1 : A \supset t : (A \vee B)$ and $\vdash_{\mathsf{LP}} v_2 : B \supset u : (A \vee B)$. Then $\vdash_{\mathsf{LP}} [(v_1{:}A \vee v_2{:}B) \supset t{:}(A \vee B)] \vee [(v_1{:}A \vee v_2{:}B) \supset u{:}(A \vee B)]$. Applying the (unstated) algorithm converting Quasi-Realizers to Realizers, we obtain that, $(v_1 : A \vee v_2 : B) \supset (c \cdot t + c \cdot u) : (A \vee B)$ is a provable normal Realization of $(\Box_1 A \vee \Box_2 B) \supset \Box_3(A \vee B)$, where c internalizes a proof of $(A \vee B) \supset (A \vee B)$.

11 What Next?

Here is a brief summary of work that remains undone. It is extensive.

We have given a constructive proof of Quasi-Realization from modal S4 to justification LP. As noted several times, a uniform algorithmic conversion from Quasi-Realizers to Realizers is available, [12]. The ideas extend directly to any modal logic having a similar destructive tableau system. Other cut-free proof methods extend things to a still richer variety of logics. In [12] it is shown that the family of modal logics having justification logic counterparts is infinite, which is somewhat surprising. But the proof is non-constructive. It is not known whether something similar can be shown constructively.

Recently LP has been extended to admit quantifiers, [3,9], with a Realization theorem connecting it with first-order S4. But a *monotonicity* condition is assumed. Work is in progress on a constant domain version, but this is incomplete. Varying domain assumptions have not yet been considered. No modal logic except S4 has been examined.

Hybrid logic and paraconsistent modal logics are largely unexplored as far as justification counterparts are concerned.

There are still things to do.

References

1. Artemov, S.N.: Explicit provability and constructive semantics. Bull. Symbolic Logic **7**(1), 1–36 (2001)
2. Artemov, S.N.: The logic of justification. Rev. Symbolic Logic **1**(4), 477–513 (2008)
3. Artemov, S.N., Yavorskaya (Sidon), T.: First-order logic of proofs. Technical report TR-2011005, City University of New York, May 2011
4. Feferman, S., Dawson Jr., J.W., Kleene, S.C., Moore, G.H., Solovay, R.M., van Heijenoort, J., Goldfarb, W.D., Parsons, C., Sieg, W. (eds.) Kurt Gödel Collected Works, vol. 5. Oxford (1986–2003)
5. Fitting, M.C.: Prolog code for S4 realization. See Realization Implemented. http://melvinfitting.org/bookspapers/techreports.html
6. Fitting, M.C.: The logic of proofs, semantically. Ann. Pure Appl. Logic **132**, 1–25 (2005)
7. Fitting, M.C.: Realizations and LP. Ann. Pure Appl. Logic **161**(3), 368–387 (2009). doi:10.1016/j.apal.2009.07.010
8. Fitting, M.C.: Reasoning with justifications. In: Makinson, D., Malinowski, J., Wansing, H. (eds.) Towards Mathematical Philosophy, Chap. 6. Trends in Logic, vol. 28, pp. 107–123. Springer, Dordrecht (2009)
9. Fitting, M.C.: Possible world semantics for first order LP. Technical report TR-2011010, CUNY Ph.D. Program in Computer Science, September 2011
10. Fitting, M.C.: Realization implemented. Technical report TR-2013005, CUNY Ph.D. Program in Computer Science, May 2013
11. Fitting, M.C.: Realization using the model existence theorem. J. Logic Comput. **26**, 213–234 (2013)
12. Fitting, M.C.: Modal logics, justification logics, and realization. Ann. Pure Appl. Logic **167**, 615–648 (2016)
13. 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 [4] I, 296–301
14. Goetschi, R., Kuznets, R.: Realization for justification logics via nested sequents: modularity through embedding. Ann. Pure Appl. Logic **163**(9), 1271–1298 (2012)
15. Smullyan, R.M.: A unifying principle in quantification theory. Proc. Natl. Acad. Sci. **49**(6), 828–832 (1963)
16. Smullyan, R.M., Logic, F.-O.: First-Order Logic. Springer, Heidelberg (1968). (Revised Edition, Dover Press, New York)

Subintuitionistic Logics with Kripke Semantics

Dick de Jongh[1(✉)] and Fatemeh Shirmohammadzadeh Maleki[2]

[1] Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands
D.H.J.deJongh@uva.nl
[2] Department of Mathematics, Shahid Beheshti University, G.C., Evin, Tehran, Iran
f.shmaleki2012@yahoo.com

Abstract. The subintuitionistic logics introduced by Corsi and Restall are developed in a uniform manner. In this way Restall's contributions are clarified. Hilbert type proof systems are given for derivations without and with assumptions. The results are applied to give conservation theorems for intuitionistic logic IPC over Corsi's system F. For Visser's basic logic additional conservation results are obtained.

Keywords: Subintuitionistic logic · Intuitionistic logic · Basic logic · Conservativity · Completeness · Modus ponens

1 Introduction

G. Corsi, in [3], introduced sublogics of intuitionistic propositional logic IPC which are characterized by classes of Kripke models in which no assumption of truth preservation is made. She made a systematic study of these systems, considering the finite model property, disjunction property and translations into modal logic as well as proving strong completeness for some of them. Her basic system was called F. G. Restall [6] made a similar study, also considering truth preservation, using different methods and a somewhat different notion of validity and proved a special type of completeness theorem. His basic system was called SJ and is considered to be equivalent to F (see Remark 2). The proofs in his paper are somewhat sketchy.

In 1981, A. Visser [10] had already introduced Basic Logic, BPC, an extension of F with truth preservation, in the natural deduction form, and proved completeness of BPC for finite and transitive Kripke models. In 1997, Suzuki and Ono [9] introduced a Hilbert style proof system for BPC as an extension of Corsi's system [3]. They proved a weak completeness theorem.

Another line of research was initiated by K. Došen [4] in 1994. As Corsi he considered translations and obtained a system K_σ, the result of a translation of K, with regard to theorems equivalent to F. He also studied the correspondent of the logic D. These investigations were continued by S. Celani and R. Jansana [2] in 2001. They concentrated on the types of consequence relations, classified the logics according to the hierarchy of abstract algebraic logic and proved strong completeness theroems, e.g. for BPC.

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, pp. 333–354, 2017.
DOI: 10.1007/978-3-662-54332-0_18

The structure of this paper is as follows. In Sect. 2 we leisurely introduce the logic F, provide a Hilbert type proof system without and with assumptions and prove a weak and strong completeness theorem. The results here are not new, they have been proved before in [2–4].

It is important to note that we consider the logics discussed strictly as their sets of theorems. The consequence relation we use, $\Gamma \vdash B$, can be defined in terms of theoremhood as $\vdash A_1, \ldots, A_n \to B$ for some $A_1, \ldots, A_n \in \Gamma$ (see Corollary 1). The strong completeness theorems we prove are w.r.t. this definable consequence relation. This should particularly be kept in mind in the case of BPC because this system was introduced by Visser with its own different consequence relation.

In Sect. 3 we will focus on Restall's validity notion [6] using the methods and the type of proof systems set up in Sect. 2, and we make his notions more explicit. We show that any prime theory Π satisfying some specific good properties can be treated in much the same way as F with the same proofs. From this the form of strong completeness of F due to Restall is shown. This relates to the consequence relations studied by Celani and Jansana [2], in particular to the system K_σ, although this is not quite the same. We then apply the results to logics stronger than F. In all of this we expose the role of the rules of modus ponens, conjunction and a fortiori.

In Sect. 4 we will introduce two special classes of formulas and show that IPC is conservative over F with respect to these classes. This clarifies what theorems of IPC can be proved in F. We will prove that IPC is in addition conservative over BPC with respect to the NNIL formulas of [11]. This clarifies what more BPC can prove than F. We relate the second result to the bounded translation for IPC into BPC given by M. Aghaei and M. Ardeshir in [1].

2 Subintuitionistic Logic

The Kripke models of subintuitionistic logics have a relation R that lacks the properties of reflexivity, transitivity and truth preservation of intuitionistic Kripke models.

Definition 1. *A* ***rooted subintuitionistic Kripke frame*** *is a triple* $\langle W, g, R\rangle$. *R is a binary relation on W; $g \in W$, the* ***root*** *is* ***omniscient****, i.e. gRw for each $w \in W$. A* ***rooted subintuitionistic Kripke model*** *is a quadruple* $\langle W, g, R, V\rangle$ *with $V : P \to 2^W$ a valuation function on the set of propositional variables P. The binary relation $\Vdash$ is defined on $w \in W$ as follows.*

1. $w \Vdash p \quad \Leftrightarrow \quad w \in V(p)$, *for any* $p \in P$,
2. $w \Vdash A \wedge B \quad \Leftrightarrow \quad w \Vdash A$ *and* $w \Vdash B$,
3. $w \Vdash A \vee B \quad \Leftrightarrow \quad w \Vdash A$ *or* $w \Vdash B$,
4. $w \Vdash A \to B \quad \Leftrightarrow \quad$ *for each v with wRv, if $v \Vdash A$ then $v \Vdash B$.*

The constant f *representing the* ***contradiction*** *is treated as a propositional variable. $M \Vdash A$ if, for all $w \in W$, $M, w \Vdash A$, and if all models force A, we write $\Vdash A$ and call A* ***valid****.*

This validity notion is Corsi's. We use Restall's omniscient roots (also called *base points*) because they play an essential role in his validity notion. We will discuss Restall's notion in Sect. 3.2, and we will show immediately that the two validities coincide. This follows from the next proposition.

Proposition 1. *If for all M, $M, g \Vdash A$, then for all M, w, $M, w \Vdash A$.*

Proof. It is not difficult to see that for each A' there exists a with respect to forcing equivalent A which is a conjunction of disjunctions of implications and atoms (compare Theorem 16). So, we can assume $M, g \Vdash A$ for an A of this form.

Let us first assume A is a disjunctions of implications and atoms, $(C_1 \to D_1) \vee \cdots \vee (C_n \to D_n) \vee p_1 \vee \cdots \vee p_k$, and let $M, w \nVdash A$. Then for each i $(1 \leq i \leq n)$, there exists w_i with $wRw_i \Vdash C_i, \nVdash D_i$. We now add an omniscient root g to M, and define the atoms $p_1, \ldots, p_k$ to be false in g. Then $g \nVdash A$.

In case A is a conjunction of such formulas we start with a countermodel to one of the conjuncts of A and continue as above. □

Definition 2. F *is the logic given by the following axioms and rules,*

1. $A \to A \vee B$
2. $B \to A \vee B$
3. $A \wedge B \to A$
4. $A \wedge B \to B$
5. $\frac{A \quad B}{A \wedge B}$
6. $\frac{A \quad A \to B}{B}$
7. $A \wedge (B \vee C) \to (A \wedge B) \vee (A \wedge C)$
8. $(A \to B) \wedge (B \to C) \to (A \to C)$
9. $(A \to B) \wedge (A \to C) \to (A \to B \wedge C)$
10. $A \to A$
11. $(A \to C) \wedge (B \to C) \to (A \vee B \to C)$
12. $\frac{A}{B \to A}$

As Restall we haven't included Corsi's $\mathsf{f} \to A$ as an axiom, which we can safely do because negation doesn't play any role in our discussions. The rules are to be applied in such a way that, if the formulas above the line are theorems of F, then the formula below the line is a theorem as well. We may write $\vdash$ for $\vdash_{\mathsf{F}}$. We will call rule 5 the *conjunction rule* and, after Corsi, rule 12 the *a fortiori rule* (it is also called *weakening rule*). We return to the rules when we discuss deduction from hypotheses. In [6] SJ has different rules and axioms but has been considered to be the same system as F though it misses the a fortiori rule 12 (see Remark 2 after Proposition 3).

For clarity's sake we prefer to prove weak completeness first, using only direct deduction without hypotheses.

Proposition 2 (Soundness of F). *In any rooted subintuitionistic Kripke model $\langle W, g, R, V \rangle$, for each $w \in W$ and each formula A, if $\vdash_{\mathsf{F}} A$ then $w \Vdash A$.*

Proof. We only check the cases 6 and 12 of F. The other cases are simpler.

(6) Let for all M and for all $w \in W$, $w \Vdash A$ and also for all M and for all $w \in W$, $w \Vdash A \to B$. We want to show that for all M and for all $w \in W$, $w \Vdash B$. By assumption, $M, g \Vdash A \to B$. We know that g is omniscient, so gRw and $w \Vdash B$, because $w \Vdash A$.

(12) Let $\Vdash A$. We have to show that for all models M and $w \in M$, $M, w \Vdash B \to A$. let wRv and $v \Vdash B$. By assumption we have $M, v \Vdash A$. That is $M, w \Vdash B \to A$. □

By soundness the following example shows that $\nvdash p \to (q \to p)$.

Example 1. Let $W = \{g, w_0, w_1\}$ and define $M = \langle W, g, R, V\rangle$ as follows:
$R = \{(g, g), (g, w_0), (g, w_1), (w_0, w_1)\}$.
$V(p) = \{w_0\}$, $V(q) = \{g, w_1\}$.
In this model $M, g \nVdash p \to (q \to p)$.

Next we will show F to be complete. Similarly to Došen [4] and by the same method we will first show that F has the disjunction property.

Definition 3. [5] We define $|A$ by induction on A, as follows

1. $|p$ iff $\vdash p$,
2. $|A \wedge B$ iff $|A$ and $|B$,
3. $|A \vee B$ iff $|A$ or $|B$,
4. $|A \to B$ iff $\vdash A \to B$ and (if $|A$ then $|B$).

Theorem 1. $|A \;\Leftrightarrow\; \vdash A$

Proof. The proof is a trivial modification of the standard one for IPC. □

Theorem 2. *If* $\vdash A \vee B$ *then* $\vdash A$ *or* $\vdash B$*.*

Proof. Let $\vdash A \vee B$. By Theorem 1($\Leftarrow$), $|A \vee B$. So $|A$ or $|B$. By Theorem 1($\Rightarrow$), $\vdash A$ or $\vdash B$. □

Remark 1. Now that we have the disjunction property the following rules adopted by Restall follow from the corresponding rules without $\vee$.

$$\frac{A \vee C \quad (A \to B) \vee C}{B \vee C} \qquad \textit{and} \qquad \frac{(A \to B) \vee E \quad (C \to D) \vee E}{((B \to C) \to (A \to D)) \vee E}$$

Because let $\vdash A \vee C$ and $\vdash (A \to B) \vee C$. By Theorem 2, $\vdash A$ or $\vdash C$, and $\vdash A \to B$ or $\vdash C$. If $\vdash C$ then $\vdash B \vee C$. So, let $\nvdash C$. Then $\vdash A$ and $\vdash A \to B$. By rule 6 of F we conclude that $\vdash B$ and hence $\vdash B \vee C$. The proof of the other rule is similar to this.

We show that we do not need Restall's rule $\frac{(A \to B) \quad (C \to D)}{(B \to C) \to (A \to D)}$, because it follows from the a fortiori rule.

Proposition 3. *Let* $\vdash A \to B$ *and* $\vdash C \to D$ *then* $\vdash (B \to C) \to (A \to D)$*.*

Proof. Let $\vdash A \to B$ and $\vdash C \to D$ then,

1. $\vdash (B \to C) \to (A \to B)$ rule 12
2. $\vdash (B \to C) \to (B \to C)$
3. $\vdash ((B \to C) \to (A \to B)) \wedge ((B \to C) \to (B \to C))$
4. $\vdash (B \to C) \to (A \to B) \wedge (B \to C)$ From 3 using axiom 9 and rule 6
5. $\vdash (A \to B) \wedge (B \to C) \to (A \to C)$
6. $\vdash (B \to C) \to (A \to C)$ From 4, 5 using axiom 8

7. $\vdash (B \to C) \to (C \to D)$	From assumption and rule 12
8. $\vdash (B \to C) \to (A \to C) \wedge (C \to D)$	From 6, 7 using axiom 9
9. $\vdash (B \to C) \to (A \to D)$	Frome 8 using axiom 8. ☐

Remark 2. We haven't been able to go the other way around, i.e. to derive the a fortiori rule from Restall's rules and axioms. We do need this rule in Sect. 3 to prove Restall's theorems for F. So, strictly speaking our proofs do not apply to SJ. The simplest solution seems to be as we do to assume the a fortiori rule to be part of SJ. It is valid in the intended semantics as we have shown by Propositions 1 and 2. All the other rules and axioms of F are easily derivable in SJ and vice versa. In any case, our theorems do apply to the logic F.

To show weak completeness of F we need some definitions.

Definition 4. *1. A set of sentences Δ is a* ***theory*** *if and only if*
  (a) $A, B \in \Delta \;\Rightarrow\; A \wedge B \in \Delta$,
  (b) $\vdash A \to B \;\Rightarrow$ (if $A \in \Delta$, then $B \in \Delta$),
  (c) F is contained in Δ.
2. For theories Γ, Δ, $\Gamma R \Delta$ iff, for all $A \to B \in \Gamma$, $A \in \Delta \Rightarrow B \in \Delta$.
3. A set of sentences Δ is ***prime*** *if and only if*

$$\text{if } A \vee B \in \Delta, \text{ then } A \in \Delta \text{ or } B \in \Delta.$$

Theorem 3. *Let Γ be a prime theory and $C \to D \notin \Gamma$. Then there is a prime theory Δ such that $\Gamma R \Delta$, $C \in \Delta$ and $D \notin \Delta$.*

Proof. Enumerate all formulas, with infinitely many repetitions: $B_0, B_1, ...$ and define

$\Delta_0 = \{E \mid C \to E \in \Gamma\}$,
$\Delta_{n+1} = \Delta_n \cup \{B_n\}$ if for no $\bar{B}_1, ..., \bar{B}_m \in \Delta_n,\; \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B_n \to D \in \Gamma$,
$\Delta_{n+1} = \Delta_n$ otherwise.

Take Δ to be the union of all Δ_n.

1. First we will show that Δ is a theory.

(a) Assume that $F \in \Delta,\; G \in \Delta$ and $F \wedge G \notin \Delta$. Let $F = B_i,\; G = B_j$ and $F \wedge G = B_n$ such that, $i \geq n$ and $j \geq n$. So there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$ such that

$$(\bar{B}_1 \wedge ... \wedge \bar{B}_m) \wedge (F \wedge G) \to D \in \Gamma \tag{1}$$

W.l.o.g. let $i \geq j$, then $\bar{B}_1, ..., \bar{B}_m, G \in \Delta_i$. By (1) we conclude that $F \notin \Delta$ and this is in contradiction with our assumption.

(b) Let $\vdash A \to B$ and $A \in \Delta$. We must show that $B \in \Delta$. Let $B = B_n$ and $B \notin \Delta$. So there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, such that

$$\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B \to D \in \Gamma$$

We know $\vdash A \to B$. We conclude by axiom 9 and Modus Ponens that

$$\vdash (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A) \to (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B)$$

and so $(\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A) \rightarrow (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B) \in \Gamma$. Now we have

$$((\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A) \rightarrow (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B)) \wedge ((\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B)) \rightarrow D) \in \Gamma \quad (2)$$

Γ is a theory, so by (2) and axiom 8 we have $(\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow D \in \Gamma$ and this is a contradiction, because $A \in \Delta$.

(c) Assume that $\vdash F$, we want to show that $F \in \Delta$. Assume $F = B_n$ and $F \notin \Delta$, then for some $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, we have

$$(\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge F) \rightarrow D \in \Gamma \quad (3)$$

We have $\vdash \bar{B}_1 \wedge ... \wedge \bar{B}_m \rightarrow F$, so

$$(\bar{B}_1 \wedge ... \wedge \bar{B}_m) \rightarrow (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge F) \in \Gamma. \quad (4)$$

Γ is a theory therefore by (3) and (4) and axiom 8 we conclude that

$$\bar{B}_1 \wedge ... \wedge \bar{B}_m \rightarrow D \in \Gamma$$

and this is a contradiction. Hence $F \in \Delta$.

2. We will show that $\Gamma R \Delta$. Let $A \rightarrow B \in \Gamma$ and $A \in \Delta$. Let $B \notin \Delta$ and $B = B_n$. Then there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$ such that

$$\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B \rightarrow D \in \Gamma \quad (5)$$

We call $\bar{B}_1 \wedge ... \wedge \bar{B}_m = C$. We have

$$\vdash (C \wedge A \rightarrow A) \wedge (A \rightarrow B) \rightarrow (C \wedge A \rightarrow B) \quad (6)$$

We know that $(C \wedge A \rightarrow A) \wedge (A \rightarrow B) \in \Gamma$ and Γ is a theory. So by (6) $C \wedge A \rightarrow B \in \Gamma$. On the other hand we have

$$\vdash (C \wedge A \rightarrow B) \wedge (C \wedge A \rightarrow C) \rightarrow (C \wedge A \rightarrow B \wedge C) \quad (7)$$

We know that $(C \wedge A \rightarrow B) \wedge (C \wedge A \rightarrow C) \in \Gamma$ and Γ is a theory, so by (7) $C \wedge A \rightarrow B \wedge C \in \Gamma$. That is

$$\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B \in \Gamma \quad (8)$$

Γ is a theory so by (5) and (8) and axiom 8 we have $\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow D \in \Gamma$ and this is a contradiction, because $A \in \Delta$. So $\Gamma R \Delta$.

3. Assume that $F \vee G \in \Delta$, and $F \notin \Delta$, $G \notin \Delta$. Let $F = B_n$ and $G = B_k$. Then there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$ such that $\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge F \rightarrow D \in \Gamma$ and also there exist $B_1^{'}, ..., B_{m'}^{'} \in \Delta_k$ such that $B_1^{'} \wedge ... \wedge B_{m'}^{'} \wedge G \rightarrow D \in \Gamma$. W.l.o.g. take $n \geq k$, then $\bar{B}_1, ..., \bar{B}_m, B_1^{'}, ..., B_{m'}^{'} \in \Delta_n$. Thus by axiom 11 and some steps we will have

$$(\bar{B}_1 \wedge ... \wedge \bar{B}_m) \wedge (B_1^{'} \wedge ... \wedge B_{m'}^{'}) \wedge (F \vee G) \rightarrow D \in \Gamma$$

But that cannot be true since $F \vee G \in \Delta$. So Δ is prime.

Finally, we know that $C \rightarrow C \in \Gamma$, so by definition $C \in \Delta_0$ and hence $C \in \Delta$. Also we have $D \rightarrow D \in \Gamma$, so $D \notin \Delta$. □

Definition 5. *We call* $\{A \mid \vdash A\}$ *the* ***empty theory***.

Proposition 4. *The empty theory* Δ *is a prime theory.*

Proof. First we will show that Δ is a theory.

(a) Let $A, B \in \Delta$, then $\vdash A$ and $\vdash B$, so $\vdash A \wedge B$. By definition, $A \wedge B \in \Delta$.
(b) Let $\vdash A \rightarrow B$ and $A \in \Delta$. Then $\vdash A$, so $\vdash B$. By definition $B \in \Delta$.
(c) Trivial. And that Δ is prime follows from Theorem 2. □

Definition 6. *The* ***Canonical Model*** $M_{\mathsf{F}} = \langle W_{\mathsf{F}}, \Delta, R, \Vdash \rangle$ *of* F *is defined by:*

1. Δ *is the empty theory,*
2. W_{F} *is the set of all prime theories,*
3. *The canonical valuation is defined by* $\Gamma \Vdash p$ *iff* $p \in \Gamma$.

In the canonical model $M_{\mathsf{F}} = \langle W_{\mathsf{F}}, \Delta, R, \Vdash \rangle$, Δ is omniscient. Because let $\Gamma \in W_{\mathsf{F}}$. If $A \rightarrow B \in \Delta$, then $\vdash A \rightarrow B$. So, if $A \in \Gamma$, then $B \in \Gamma$.

Lemma 1 (Truth lemma). *For each* $\Gamma \in W_{\mathsf{F}}$ *and for every formula* C,

$$\Gamma \Vdash C \text{ iff } C \in \Gamma.$$

Proof. By induction on C. We skip the proof because it is standard. We just mention that Theorem 3 is used for the $\rightarrow$-case. □

Theorem 4 (Weak Completeness). *For any formula* A *if* $\Vdash A$, *then* $\vdash A$.

Proof. Let $\nvdash A$ and let Δ be the empty theory. By the definition of empty theory $A \notin \Delta$. So, we have $M_{\mathsf{F}}, \Delta \nVdash A$. That is, $\nVdash A$. □

Next we prove strong completeness with the semantics as in Corsi [3]. First we introduce a notion of derivation from hypotheses.

Definition 7

(a) We define $\Gamma \vdash A$ *if there is a derivation of* A *from* Γ *and theorems of* F *using the rules* $\frac{A \quad B}{A \wedge B}$, *and* $\frac{A \quad A \rightarrow B}{B}$ (*only if* $\vdash_{\mathsf{F}} A \rightarrow B$).
(b) We define $\Gamma \Vdash A$ *iff for all* $M, w \in M$, *if* $M, w \Vdash \Gamma$ *then* $M, w \Vdash A$.

Remark 3. Note that if $\Gamma \vdash A$ then it does not follow that $\Gamma \vdash B \rightarrow A$. For example if we assume that $\Gamma = \mathsf{F} \cup \{p\}$, then $\Gamma \vdash p$ and $\Gamma \nvdash q \rightarrow p$.

Surprisingly, the weak Deduction Theorem holds for F and $\vdash$.

Theorem 5 (Weak Deduction Theorem). $A \vdash B$ *if and only if* $\vdash A \rightarrow B$.

Proof. $\Rightarrow$: By induction on the length of the proof.
If B is a theorem of F. Then $\vdash B$, so by rule 12, $\vdash A \rightarrow B$.
$A \vdash A$ is covered by $\vdash A \rightarrow A$.
If $A \vdash B$ and $A \vdash C$. By induction hypothesis $\vdash A \rightarrow B$ and $\vdash A \rightarrow C$, so $\vdash A \rightarrow B \wedge C$.
If $A \vdash B$ and $\vdash B \rightarrow C$. Then by induction hypothesis $\vdash A \rightarrow B$, so $\vdash A \rightarrow C$.
$\Leftarrow$: By definition this direction is straightforward. □

Corollary 1. *1.* $A_1, ..., A_n \vdash B$ *iff* $\vdash A_1 \wedge ... \wedge A_n \to B$.
2. $\Delta \vdash B$ *iff* $A_1 \wedge ... \wedge A_n \vdash B$ *for some* $A_1, ..., A_n \in \Delta$.

Proof. The proof is easy. □

Proposition 5. Δ *is a theory* $\Longleftrightarrow$ $\Delta \vdash A$ *if and only if* $A \in \Delta$.

Proof. $\Rightarrow$: The proof from right to left is immediate. The other direction is by induction on the length of the derivation. If $A \in \Delta$ there is nothing to prove. If A is a theorem of F, then by definition of theory $A \in \Delta$.

If $\Delta \vdash A$ and $\Delta \vdash B$, by induction hypothesis $A \in \Delta$ and $B \in \Delta$. So, by the definition of theory $A \wedge B \in \Delta$.

If $\vdash A \to B$ and $\Delta \vdash A$, by induction hypothesis $A \in \Delta$, and by definition of theory $B \in \Delta$.
$\Leftarrow$: This is straightforward. □

Theorem 6. *If* $\Sigma \nvdash D$ *then there is a prime theory* Δ *such that* $\Delta \supseteq \Sigma$, $D \notin \Delta$.

Proof. By assumption and by definition of provability we conclude that $D \notin \Sigma$. Enumerate all formulas, with infinitely many repetitions: $B_0, B_1, ...$ and define

$\Delta_0 = \Sigma \cup \mathsf{F}$,
$\Delta_{n+1} = \Delta_n \cup \{B_n\}$ if $\Delta_n, B_n \nvdash D$,
$\Delta_{n+1} = \Delta_n$ otherwise.

Take Δ to be the union of all Δ_n. The proof now runs as for Theorem 3. □

Theorem 7 (Strong Completeness). *For any formula* A, $\Sigma \vdash A$ *if and only if* $\Sigma \Vdash A$.

Proof. Left to right is easy. For the other direction, Let $\Sigma \nvdash A$. Then by Theorem 6, there is a prime theory $\Gamma \supseteq \Sigma$ such that $A \notin \Gamma$. So, we will have $M_{\mathsf{F}}, \Gamma \Vdash \Sigma$ and $M_{\mathsf{F}}, \Gamma \nVdash A$. That is $\Sigma \nVdash A$. □

We will not discuss the finite model property in this paper, or translations into modal logic. We have no new results in that area and refer the reader to Corsi [3], Došen [4] and Sano and Ma [7].

3 Π-Provability and Restall's Strong Completeness

In this section we reprove Restall's completeness theorem in a clearer form and use its concepts to prove a very general completeness theorem. We disentangle the notions of Π-provability and strong provability from a set of assumptions, and we clarify the obscure role of the a fortiori rule, first using a restricted form.

In 3.1 we develop the notion of Π-provability. This is a notion of proof from a theory Π which allows use of Π in a strong way that includes a restricted a fortiori rule. One may say that Π is considered as a true axiom system. We assume that Π has good properties, here called 'adequate'. A completeness theorem is proved for Π-provability in much the same way as for provability in F.

In 3.2 the notion of Π-provability is developed into a more general notion $\vdash_r$ of r-provability from arbitrary hypothesis sets Σ. It is based on Restall's validity notion. He considers validity in a model as truth in the root, and validity of a consequence accordingly. We develop the two notions of provability and consequence along the lines of Sect. 2 still relying on the restricted a fortiori rule. We prove the form of strong completeness connected with $\vdash_r$, which can be called Restall's completeness theorem. It states that if a set of assumptions doesn't $\vdash_r$-prove A, then the set of assumptions can be extended to an adequate theory that still doesn't prove A and is the root of its canonical model. Strong completeness in the ordinary sense does not follow for adequate theories.

In 3.3 we strengthen the concept of adequate theory to fully adequate theory satisfying the full a fortiori rule, and prove a very general completeness theorem showing that such theories can occur as the root of canonical models in such a way that strong completeness follows for such theories. The method is quite the same as developed in Sect. 2. We sketch how this can be applied to provability in logics stronger than F. We in particular consider BPC.

3.1 Π-Provability

We start by introducing properties for a theory Π which will guarantee that provability from Π can be treated to a large extent just as provability in F itself so that we can proceed in proving completeness of Π-provability in much the same way as for provability in F (e.g. in modus ponens $A \to B$ is assumed to be in Π, and Π-theories behave much as theories).

Definition 8. *Δ is a Π-theory if and only if:*

1. *If $A, B \in \Delta$, then $A \wedge B \in \Delta$,*
2. *If $A \to B \in \Pi$ and $A \in \Delta$, then $B \in \Delta$,*
3. *The set $\Pi_{\to}$ of members of Π of the form $A \to B$ is contained in Δ,*
4. *F is contained in Δ.*

Definition 9. *Π is an* ***adequate theory*** *if Π is a prime Π-theory closed under the* ***restricted a fortiori rule****, if $A \in \Pi_{\to}$, then for all B, $B \to A \in \Pi$.*

Lemma 2. *Π is an adequate theory iff Π is a prime theory closed under modus ponens and the restricted a fortiori rule.*

Proof. Obvious. □

In all of Sect. 3, Π will be assumed to be an adequate theory. This implies that Π is closed under its own Π-provability rules (see Corollary 3) and will turn out to make Π suitable to be the set of formulas true in the root of a model.

Definition 10. *We define $\Gamma \vdash_\Pi A$ as: there is a derivation of A from $\Gamma \cup \Pi_{\to} \cup \mathsf{F}$ using the rules $\frac{A \quad B}{A \wedge B}$, $\frac{A \quad A \to B}{B}$ with $A \to B \in \Pi$ in the latter case.*

Proposition 6. *Δ is a Π-theory $\Longleftrightarrow$ $\Delta \vdash_\Pi A$ if and only if $A \in \Delta$.*

Proof. $\Rightarrow$: From right to left is trivial. The other direction is by induction on the length of the proof. If $A \in \Delta \cup \mathsf{F} \cup \Pi_{\rightarrow}$ there is nothing to prove.

Let $\Delta \vdash_{\Pi} A$ and $\Delta \vdash_{\Pi} B$. By induction hypothesis $A \in \Delta$ and $B \in \Delta$. So, by definition of Π-theory $A \wedge B \in \Delta$.

If $A \rightarrow B \in \Pi$ and $\Delta \vdash_{\Pi} A$. By induction hypothesis $A \in \Delta$ and by definition of Π-theory $B \in \Delta$.
$\Leftarrow$: Straightforward. □

So, under the assumption that Π is an adequate theory:

Corollary 2. *$\Pi \vdash_{\Pi} A$ iff $A \in \Pi$.*

Theorem 8. *If Γ is a prime Π-theory with $C \rightarrow D \notin \Gamma$, then there is a prime Π-theory Δ with $\Gamma R \Delta$, $C \in \Delta$ and $D \notin \Delta$.*

Proof. Enumerate all formulas, with infinitely many repetitions: $B_0, B_1, ...$ and define

$\Delta_0 = \{E \mid C \rightarrow E \in \Gamma\} \cup \mathsf{F}$,
$\Delta_{n+1} = \Delta_n \cup \{B_n\}$ if for all $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, $\Gamma \nvdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B_n \rightarrow D$,
$\Delta_{n+1} = \Delta_n$ otherwise.

Take Δ to be the union of all Δ_n. We show that

1. $\Gamma R \Delta$,
2. Δ is a prime Π-theory,
3. $C \in \Delta$,
4. $D \notin \Delta$.

1. Let $A \rightarrow B \in \Gamma$ and $A \in \Delta$. We must show that $B \in \Delta$. Let $B = B_n$ and $B \notin \Delta$. So there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$ such that

$$\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B \rightarrow D \tag{9}$$

$A \rightarrow B \in \Gamma$ so, $\Gamma \vdash_{\Pi} A \rightarrow B$ and $\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow A$ then,

$$\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow B$$

Also we have $\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow \bar{B}_1 \wedge ... \wedge \bar{B}_m$, therefore

$$\Gamma \vdash_{\Pi} (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow \bar{B}_1 \wedge ... \wedge \bar{B}_m) \wedge (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow B). \tag{10}$$

By (10) and axiom 9 we conclude that

$$\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B. \tag{11}$$

By (9) and (11), we have

$$\Gamma \vdash_{\Pi} (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B) \wedge (\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B \rightarrow D). \tag{12}$$

Again by (12) and axiom 8 we have $\Gamma \vdash_{\Pi} \bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge A \rightarrow D$ and this is a contradiction, because $A \in \Delta$. So $B \in \Delta$ and hence $\Gamma R \Delta$.
2. Let $A \in \Delta$, $A \rightarrow B \in \Pi$. So, also $A \rightarrow B \in \Gamma$ because Γ is a Π-theory. As in (1) $B \in \Delta$ follows.

Let $F \in \Pi_{\rightarrow}$. We want to show that $F \in \Delta$. We know that Π is closed under the restricted a fortiori rule, so $C \rightarrow F \in \Pi_{\rightarrow}$ and therefore $C \rightarrow F \in \Gamma$. So by definition of Δ_0, $F \in \Delta_0$ and thus $F \in \Delta$.

As in the proof of Theorem 3, we can conclude that Δ is prime and closed under conjunction.
3. We know that $C \rightarrow C \in \Gamma$, so $C \in \Delta_0$ and then $C \in \Delta$.
4. $D \notin \Delta$, since $\Gamma \vdash_{\Pi} D \rightarrow D$. □

Definition 11. *In the Π-**canonical model** $M^{\Pi} = \langle W_{\Pi}, \Pi, R, \Vdash \rangle$ of* F

1. *W_{Π} is the set of all prime Π-theories,*
2. *The canonical valuation is defined by $\Gamma \Vdash p$ if and only if $p \in \Gamma$.*

Lemma 3 (Truth lemma). *For each $\Gamma \in W_{\Pi}$ and for every formula C,*

$$\Gamma \Vdash C \quad \text{iff} \quad \mathrm{C} \in \Gamma.$$

Proof. By induction on C exactly as in the proof of Lemma 1. □

This truth lemma lead as usual to completeness of Π-provability.

Theorem 9 (Weak Completeness of Π-provability).
$\Pi \vdash_{\Pi} A \;\Leftrightarrow\; M^{\Pi}, \Pi \Vdash A.$

Proof. Immediate by Proposition 6 and Lemma 3. □

This is of course only a weak completeness theorem. To get a strong form we need the unrestricted a fortiori rule for Π. This we will do in 3.3. But first we prove Restall's form of completeness in the next subsection because our proof there uses the restricted rule.

3.2 Restall's Form of Strong Completeness

We first introduce a stronger notion of proof $\Gamma \vdash_r$ from a set of assumptions Γ. It will include full modus ponens as well as the restricted a fortiori rule. We do not assume that the set of assumptions is a prime theory or even a theory. The fact that as in Restall [6] no disjunction property is assumed means that the proof rules will have to be more complex. The idea is that it will be ultimately be possible to extend any set of assumptions to an adequate theory which is the essential step leading to Restall's completeness theorem.

As far as we can see we do need to restrict the a fortiori rule to implications to be able to prove Lemma 5 which is crucial for the Completeness Theorem 10. The consequence relation $\vdash_r$ is close to the consequence relation of the system K_{σ} introduced by Celani and Jansana [2]. The logic K_{σ} is one of the variations of F introduced in [2] with different consequence relations. But K_{σ} has the full a fortiori rule (called (W) for weakening in that article) contrary to $\vdash_r$. Specifically, $p \vdash_{\mathsf{K}_{\sigma}} q \rightarrow p$ but $p \nvdash_r q \rightarrow p$.

Definition 12. (*a*) *We define $\Gamma \vdash_r A$ if there is a derivation of A from Γ and theorems of* F *using the rules*

$$\frac{A \quad B}{A \wedge B}\ ,\quad \frac{A \quad A \to B}{B}\ ,\quad \frac{A \vee C \quad (A \to B) \vee C}{B \vee C}$$

and

$$\frac{A}{B \to A}\quad ,\quad \frac{A \vee C}{(B \to A) \vee C}$$

with in the latter two cases the restriction that A has to be an implication.
(*b*) *We define, $\Gamma \vDash_r A$ iff for all $M = \langle W, g, R, V\rangle$, if $M, g \Vdash \Gamma$ then $M, g \Vdash A$.*

We now first clarify the relationship between $\vdash_r$ and $\vdash_\Pi$.

Proposition 7. *$\Pi \vdash_r A \;\Leftrightarrow\; A \in \Pi$.*

Proof. $\Rightarrow$: By induction on the length of the proof.

If $A \in \Pi$ there is nothing to prove.

If $\Pi \vdash_r A$ and $\Pi \vdash_r B$, then, by induction hypothesis, $A \in \Pi$ and $B \in \Pi$. So, by the definition of Π-theory $A \wedge B \in \Pi$.

If $\Pi \vdash_r A \to B$ and $\Pi \vdash_r A$, then, by induction hypothesis, $A \in \Pi$ and $A \to B \in \Pi$. So, by the definition of Π-theory $B \in \Pi$.

If $\Pi \vdash_r (A \to B) \vee C$ and $\Pi \vdash_r A \vee C$, then, by induction hypothesis, $(A \to B) \vee C \in \Pi$ and $A \vee C \in \Pi$. So, $A \to B \in \Pi$ or $C \in \Pi$, and $A \in \Pi$ or $C \in \Pi$, since Π is prime. Therefore $C \in \Pi$, or $A \to B \in \Pi$ and $A \in \Pi$. In the latter case, by definition of Π-theory, $B \in \Pi$. So, in both cases $B \vee C \in \Pi$.

If $\Pi \vdash_r A$ and A is an implication, then, by the induction hypothesis and the closure of Π under the a restricted a fortiori rule, for all B, $B \to A \in \Pi$.

If $\Pi \vdash_r A \vee C$ and A is an implication, then, by the induction hypothesis $A \vee C \in \Pi$. So $A \in \Pi$ or $C \in \Pi$, since Π is prime. If $A \in \Pi$, then for all B, $B \to A \in \Pi$, since Π is closed under the a restricted a fortiori rule. So, in both cases $(B \to A) \vee C \in \Pi$.

$\Leftarrow$: If $A \in \Pi$, then by definition, $\Pi \vdash_r A$. □

Corollary 3. *$\Pi \vdash_r A \;\Leftrightarrow\; \Pi \vdash_\Pi A$.*

Proof. Immediate from Propositions 6 and 7. □

Next, following Restall [6], we show how to reason with $\vee$ in case we do not have the disjunction property for the assumptions.

Lemma 4. *If $A \vdash_r B$ then $C \vee A \vdash_r C \vee B$.*

Proof. The proof is easy by induction on the length of the proof. □

Proposition 8. *If $A \vdash_r C$ and $B \vdash_r C$, then $A \vee B \vdash_r C$.*

Proof. By Lemma 4, $A \vee B \vdash_r C \vee B$, and also, $C \vee B \vdash_r C \vee C$. It is simple to show that $\vdash_r C \vee C \to C$, so $A \vee B \vdash_r C$. □

Now we have shown that reasoning from disjunctions can be executed properly we have reached the point at which we can show that an arbitrary set of formulas not proving a certain formula A can be extended to an adequate theory not proving A, the point of which is that an adequate theory is the root of its own canonical model (Theorem 9).

Lemma 5. *If $\Sigma \nvdash_r A$, then there is a $\Pi \supseteq \Sigma$ such that Π is an adequate theory and $\Pi \nvdash_r A$.*

Proof. Enumerate all formulas, with infinitely many repetitions: $B_0, B_1, ...$ and define

$\Pi_0 = \{B \mid \Sigma \vdash_r B\}$,
$\Pi_{n+1} = \Pi_n \cup \{B_n\}$ if for no $\bar{B}_1, ..., \bar{B}_m \in \Pi_n,\ \bar{B}_1, ..., \bar{B}_m, B_n \vdash_r A$,
$\Pi_{n+1} = \Pi_n$ otherwise.

Take Π to be the union of all Π_n. By definition of Π, it is clear that $\Pi \nvdash_r A$.

We must show that Π is a Π-theory. Assume that $E \in \Pi,\ F \in \Pi$ and $E \wedge F \notin \Pi$. Let $E = B_i,\ F = B_j$ and $E \wedge F = B_n$ such that, $i \geq n$ and $j \geq n$. So there exist $\bar{B}_1, ..., \bar{B}_m \in \Pi_n$, such that

$$\bar{B}_1, ..., \bar{B}_m, E \wedge F \vdash_r A$$

and so,

$$\bar{B}_1, ..., \bar{B}_m, E, F \vdash_r A$$

But $E, F, \bar{B}_1, ..., \bar{B}_m \in \Pi_j$, so $\bar{B}_1, ..., \bar{B}_m, E, F \nvdash_r A$, a contradiction.

Now let $C \to D \in \Pi$ and $C \in \Pi$ we must show that $D \in \Pi$. Let $D = B_n$ and $D \notin \Delta$. So there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, such that

$$\bar{B}_1, ..., \bar{B}_m, D \vdash_r A$$

and so

$$\bar{B}_1, ..., \bar{B}_m, C \to D, C \vdash_r A$$

This is a contradiction.

Assume that $E \vee F \in \Pi$, and $E \notin \Pi,\ F \notin \Pi$. Let $E = B_n$ and $F = B_k$. Then there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, such that $\bar{B}_1, ..., \bar{B}_m, E \vdash_r A$ and therefore

$$\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge E \vdash_r A \tag{13}$$

and also there exist $B_1^{'}, ..., B_{m'}^{'} \in \Delta_k$, such that $B_1^{'}, ..., B_{m'}^{'}, F \vdash_r A$ and therefore

$$B_1^{'} \wedge ... \wedge B_{m'}^{'} \wedge F \vdash_r A \tag{14}$$

By (13) and (14), the distributive law and Proposition 8 we can conclude $\bar{B}_1 \wedge ... \wedge \bar{B}_m \wedge B_1^{'} \wedge ... \wedge B_{m'}^{'} \wedge (E \vee F) \vdash_r A$ and hence

$$\bar{B}_1, ..., \bar{B}_m, B_1^{'}, ..., B_{m'}^{'}, E \vee F \vdash_r A \tag{15}$$

But this is a contradiction, since $E \vee F \in \Pi$. So Π is a prime Π-theory.

Finally let $E \in \Pi$ be an implication. We need to show that for all B, $B \to E \in \Pi$. Let $B \to E \notin \Pi$, then there exist $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, such that

$$\bar{B}_1, ..., \bar{B}_m, B \to E \vdash_r A \tag{16}$$

So, by (16) we have $\bar{B}_1, ..., \bar{B}_m, E \vdash_r A$, since from E we can derive $B \to E$. But this is a contradiction, hence $B \to E \in \Pi$. □

Theorem 10 (Restall's Completeness Theorem). *$\Sigma \vdash_r A$ if and only if $\Sigma \vDash_r A$.*

Proof. ⇒: Suppose $\Sigma \vdash_r A$. We use induction on the length of the derivation of A from Σ to prove that $\Sigma \vDash_r A$. We only check one case.

Let A be an implication and $\Sigma \vDash_r A$. We want to show that for all formulas B, $\Sigma \vDash_r B \to A$. Let $M = \langle W, g, R, \Vdash \rangle$ and $M, g \Vdash \Sigma$. By assumption $M, g \Vdash A$, the root g is omniscient and A is an implication formula, so let $A = C \to D$. We will show that for all $v \in W$, $M, v \Vdash A$. For this purpose let vRz and $M, z \Vdash C$. Then $M, z \Vdash D$, since gRz and $M, g \Vdash C \to D$. Hence for all $v \in W$, $M, v \Vdash A$ and then $M, g \Vdash B \to A$.

⇐: Let $\Sigma \nvdash_r A$. By Lemma 5 there is a prime Π-theory, $\Pi \supseteq \Sigma$ such that $A \notin \Pi$. So, in the canonical model $M^\Pi = \langle W_\Pi, \Pi, R, \Vdash \rangle$, $M^\Pi, \Pi \Vdash \Sigma$ and $M^\Pi, \Pi \nVdash A$, since $A \notin \Pi$. So $\Sigma \nvDash_r A$. □

3.3 *Π*-Provability and Stronger Logics

In this subsection we strengthen the conditions on Π to ensure that it satisfies the full a fortiori rule. This makes it possible to prove strong completeness for Π-provability very much again along the lines of proof of Sect. 2. Note though that the consequence relation $\vdash_r$ sticks to the restricted a fortiori rule.

Definition 13. *Π is a **fully adequate theory** if Π is an adequate theory containing no formulas without implications.*

A fully adequate theory can be said to make no purely local statements.

Lemma 6. *A fully adequate theory Π is closed under the (unrestricted) a fortiori rule.*

Proof. We have to prove $A \in \Pi \Rightarrow D \to A \in \Pi$. We prove it by induction on the complexity of A. Note that the base case is that A is an implication. Statements without implication are not in Π.

If $A \in \Pi$ is an implication, then $D \to A \in \Pi$ by assumption.

If $A \in \Pi$ is $B \wedge C$, then, by axioms 3 and 4 and modus ponens $B \in \Pi$ and $C \in \Pi$. By induction hypothesis, $D \to B \in \Pi$ and $D \to C \in \Pi$. Then, by the conjunction rule, axiom 9 and modus ponens, $D \to B \wedge C \in \Pi$.

If $A \in \Pi$ is $B \vee C$, then, since Π is prime, $B \in \Pi$ or $C \in \Pi$. By induction hypothesis $D \to B \in \Pi$ or $D \to C \in \Pi$, and by axiom 1 or 2, axiom 8 and modus ponens, $D \to B \vee C \in \Pi$. □

Theorem 11. *If Π is a fully adequate theory and $\Sigma \nvdash_\Pi D$, then there is a prime Π-theory $\Delta \supseteq \Sigma$ such that $D \notin \Delta$.*

Proof. Enumerate all formulas, with infinitely many repetitions: $B_0, B_1, ...$ and define

$\Delta_0 = \Sigma \cup \Pi_{\rightarrow} \cup \mathsf{F}$,
$\Delta_{n+1} = \Delta_n \cup \{B_n\}$ if for no $\bar{B}_1, ..., \bar{B}_m \in \Delta_n$, $\vdash_\Pi \bar{B}_1 \wedge \bar{B}_m \wedge ... \wedge B_n \rightarrow D$,
$\Delta_{n+1} = \Delta_n$ otherwise.

Take Δ to be the union of all Δ_n. By assumption $D \notin \Delta_0$ and also we have $\vdash_\Pi D \rightarrow D$, so $D \notin \Delta$.

We show that Δ is a prime Π-theory. This simply goes exactly as in the proof of Theorems 3 and 6, Π has all the relevant properties of F that were used in these proofs. □

Theorem 12. *If Π is a fully adequate theory then $\Sigma \vdash_\Pi A$ if and only if for all Γ in the Π-Canonical model M^Π, if $\Gamma \Vdash \Sigma$ then $\Gamma \Vdash A$.*

Proof. Left to right is easy by induction on the length of the proof. The other direction follows by Theorem 11. □

Definition 14. *We define $\Delta \vDash_\Pi A$ iff for all $M = \langle W, g, R, V \rangle$ such that $M, g \Vdash \Pi$, and all $w \in W$, if $M, w \Vdash \Delta$, then $M, w \Vdash A$.*

This now allows us to state a very general completeness theorem.

Theorem 13 (Π-completeness theorem). *If Π is a fully adequate theory, then $\Delta \vdash_\Pi A \;\Leftrightarrow\; \Delta \vDash_\Pi A$.*

Proof. Left to right is easy, the other direction follows by Theorem 12. □

This theorem can be applied to any logic extending F as long as it has the rules of modus ponens, conjunction and the a fortiori rule. Of course, a logic will usually be closed under substitution but there is no need for this. To get useful completeness theorems we of course will have to prove that the canonical model of the logic has the desired properties. As an example consider the case of BPC. Strong completeness has been proved before for a similar proof system in [2], but for the next section it is good to get an idea of this system.

BPC is interpreted in Kripke models similarly to intuitionistic propositional logic except that the accessibility relation is not necessarily reflexive. Suzuki and Ono [9] introduced a Hilbert style proof system for BPC. Their axiomatization is an extension of the logic F by the axioms, $A \rightarrow (B \rightarrow A)$, $\mathsf{f} \rightarrow A$ and $A \rightarrow (B \rightarrow A \wedge B)$.

It is to be remarked that $A \rightarrow (B \rightarrow A \wedge B)$ follows from $\mathsf{F} + (A \rightarrow (B \rightarrow A))$ using the conjunction rule (rule 5):

1. $\vdash A \rightarrow (B \rightarrow A)$
2. $\vdash B \rightarrow B$

3. $\vdash (B \to B) \to (A \to (B \to B))$	Axiom
4. $\vdash A \to (B \to B)$	From 2, 3 by modus ponens
5. $\vdash (A \to (B \to A)) \wedge (A \to (B \to B))$	From 1, 4 using rule 5 (conjunction)
6. $\vdash A \to (B \to A) \wedge (B \to B)$	Follows from 5 using axiom 9
7. $\vdash (B \to A) \wedge (B \to B) \to (B \to A \wedge B)$	
8. $\vdash A \to (B \to A \wedge B)$	Follows from 6, 7 using axiom 8

In the step from 2 to 4 one sees how a fortiori can be circumvented by using $A \to (B \to A)$. In reasoning without assumptions the conjunction rule (rule 5) is superfluous if $A \to (B \to A \wedge B)$ is present; it then follows by modus ponens. But in reasoning with assumptions one would need unrestricted modus ponens.

The next Lemma and the strong completeness theorem for BPC can be found in Restall [6] and Celani and Jansana [2].

Lemma 7. *Let $M^{\Pi} = \langle W^{\Pi}, \Pi, R, \Vdash \rangle$ be the Π-canonical model for some Π containing $A \to (B \to A)$ for all A, B. Then the relation R is transitive and satisfies preservation of truth.*

Proof. First we will show that R is transitive. Let Γ, Δ and Σ are in W^{Π} and let $\Gamma R \Delta$ and $\Delta R \Sigma$, we want to prove that $\Gamma R \Sigma$. So let $A \to B \in \Gamma$ and $A \in \Sigma$. We have $(A \to B) \to (\top \to (A \to B)) \in \Pi$. So by definition of Π-theory $\top \to (A \to B) \in \Gamma$. However $\top \in \Delta$ and $\Gamma R \Delta$, so by definition of R, $A \to B \in \Delta$. Again by definition of R, $B \in \Sigma$, since $A \in \Sigma$ and $\Delta R \Sigma$. That is, R is transitive.

Now we will show that $\Vdash$ preserves truth in the Π-canonical model. Assume that $A \in \Gamma$ and $\Gamma R \Delta$. As Δ is nonempty, there is a $B \in \Delta$. The assumption gives $B \to A \in \Gamma$ (since $A \to (B \to A) \in \Pi$), and so $A \in \Delta$. □

Theorem 14 (Completeness Theorem for BPC). *$\Sigma \vdash_{\mathsf{BPC}} A \Leftrightarrow \Sigma \vDash_{\mathsf{BPC}} A$.*

Proof. Immediate by Lemma 7 and Theorem 13. □

4 Relation of F to Intuitionistic Propositional Logic

In this section, we prove conservativity results for IPC over F and over BPC. This clarifies what part of IPC these systems can prove.

Definition 15. *If L_1 and L_2 are logics with L_2 extending L_1, and Σ is a class of formulas, then L_2 is **conservative over** L_1 **with respect to** Σ if, all $S \in \Sigma$, if $L_2 \vdash S$, then $L_1 \vdash S$.*

4.1 Conservativity Results for IPC over F

We will provide two classes of formulas with respect to which IPC is conservative over F, the class of simple implications and the class of basic implications.

Definition 16. *Let us call a formula $A \to B$ with A and B containing only $\wedge$ and $\vee$ a* ***simple implication****, and a formula that is obtained by applying only $\wedge$ and $\vee$ to simple implications a* ***basic formula****. Finally a formula $A \to B$ with A and B basic formulas is a* ***basic implication****.*

Theorem 15. *If $\vdash_{\mathsf{F}} A \leftrightarrow B$, then $\vdash_{\mathsf{F}} E[A/p] \leftrightarrow E[B/p]$, where p is an atom.*

Proof. The proof is easy by induction on E. □

Theorem 16. *Let A be a formula such that A is constructed by applying only $\wedge$ and $\vee$ to formulas from a class Θ. Then there are formulas $A^{'}, A^{''}$ such that*

1. *$\vdash_{\mathsf{F}} A \leftrightarrow A^{'}$ and $A^{'}$ is a disjunction of conjunctions of formulas in Θ.*
2. *$\vdash_{\mathsf{F}} A \leftrightarrow A^{''}$ and $A^{''}$ is a conjunction of disjunctions of formulas in Θ.*

Proof. The proof is straightforward. □

We will apply Theorem 16 to Θ as the class of atoms, and as the class of simple implications. Now by the previous theorems, a simple implication $A \to B$ can be replaced by an F- and IPC-equivalent $A^{'} \to B^{'}$ such that $A^{'}$ is a disjunction of conjunctions and $B^{'}$ is a conjunction of disjunctions.

Lemma 8.

1. *For all formulas A, B, C, $\vdash_{\mathsf{F}} (A \vee B \to C) \leftrightarrow (A \to C) \wedge (B \to C)$,*
2. *For all formulas A, B, C, $\vdash_{\mathsf{F}} (A \to B \wedge C) \leftrightarrow (A \to B) \wedge (A \to C)$,*
3. *For all formulas $A_i, 1 \leq i \leq k$ and $B_i, 1 \leq j \leq m$,*

$$\vdash A_1 \vee ... \vee A_K \to B_1 \wedge ... \wedge B_m \text{ iff } \vdash A_i \to B_j \text{ for all } i,\ j,$$

where $\vdash$ can be read as $\vdash_{\mathsf{F}}$ as well as $\vdash_{\mathsf{IPC}}$.

Proof. Easy. □

Definition 17. *A formula $A \to B$ called a* ***very simple implication*** *if A is conjunction of atoms and B is disjunction of atoms. A formula $A \to B$ is called* ***very basic implication*** *if A is conjunction of very simple implications and B is disjunction of very simple implications.*

By Lemma 8 and Theorem 15 we can conclude that to show that IPC is conservative over F with respect to simple implications and basic implications it is sufficient to do so for very simple implications and very basic implications. We can do so now for very simple implications, and in fact even for CPC instead of IPC.

Theorem 17. *If* CPC *proves a very simple implication (and a fortiori if* IPC *does), then* F *proves it as well.*

Proof. Let $A \to B$ is a very simple implication, so $A = \bigwedge_i(p_i)$ and $B = \bigvee_j(q_j)$. Assume $\nvdash_{\mathsf{F}} A \to B$. Then by the completeness theorem there exists a rooted subintuitionistic model M and $w \in M$, such that $M, w \nVdash A \to B$. So there exists $v \in M$, such that $M, v \Vdash A$ and $M, v \nVdash B$. Now we select this point v from M and then we make the one point CPC model $M_{\mathsf{CPC}} = \langle v, (v, v), \vDash\rangle$ such that for all propositional variables p, $M_{\mathsf{CPC}}, v \vDash p$ if and only if $M, v \Vdash p$. Clearly

$$M_{\mathsf{CPC}}, v \vDash p_i, \ \ for\ all\ i$$

$$M_{\mathsf{CPC}}, v \nvDash q_j, \ \ for\ all\ j$$

That is $M_{\mathsf{CPC}}, v \nvDash A \to B$, so $\mathsf{CPC} \nvDash A \to B$. □

Up to now CPC (classical logic) did just as well as IPC, but to restrict the class of very basic implications further we need disjunction properties only available in (sub)intuitionistic logics. We need the slash $|$ for this purpose also under assumptions. The following both applies if $\vdash$ is read as $\vdash_{\mathsf{F}}$ and as $\vdash_{\mathsf{IPC}}$. Similarly for the $|$ defined in terms of $\vdash$.

Definition 18. *Let Γ be a set of formulas. We define the slash $\Gamma|A$ inductively on the structure of A as follows*

1. $\Gamma \mid p$ *iff* $\Gamma \vdash p$,
2. $\Gamma|A \wedge B$ *iff* $\Gamma|A$ *and* $\Gamma|B$,
3. $\Gamma|A \vee B$ *iff* $\Gamma|A$ *or* $\Gamma|B$,
4. $\Gamma|A \to B$ *iff* $\Gamma \vdash A \to B$ *and (if* $\Gamma|A$ *then* $\Gamma|B$*).*

Theorem 18. [5] *If $\Gamma|A$ for all $A \in \Gamma$, then $(\Gamma|B \Leftrightarrow \Gamma \vdash B)$.*

Proof. As in [5]. □

Theorem 19. [5] *If $\Gamma|C$ for all $C \in \Gamma$ and $\Gamma \vdash A \vee B$, then $\Gamma \vdash A$ or $\Gamma \vdash B$.*

Proof. Let $\Gamma \vdash A \vee B$. By Theorem 18, $\Gamma|A \vee B$. So $\Gamma|A$ or $\Gamma|B$. Again by Theorem 18, $\Gamma \vdash A$ or $\Gamma \vdash B$. □

Lemma 9. *If $A = A_1 \wedge ... \wedge A_k$, such that for all $1 \leq i \leq k$, A_i is a very simple implication, then $A|A$. Similarly if, for all $1 \leq i \leq k$, A_i is an atom.*

Proof. By assumption for all $1 \leq i \leq n$, $A_i = B_i \to C_i$. Clearly, $A_1 \wedge ... \wedge A_k \vdash B_i \to C_i$. We have $A_1 \wedge ... \wedge A_k \nvdash B_i$, because we can make a model M such that $M \Vdash A_1 \wedge ... \wedge A_k$ and $M \nVdash B_i$ (we make all atoms false). So $A_1 \wedge ... \wedge A_k \nmid B_i$ and therefore $A_1 \wedge ... \wedge A_k|B_i \to C_i$. So, $A_1 \wedge ... \wedge A_k|A_1 \wedge ... \wedge A_k$. □

Theorem 20. *For arbitrary A, D, if $A|A$ and $\vdash A \to D$, then $A|D$.*

Proof. Let $\Gamma = \{A\}$. Then by Theorem 18, $\Gamma \vdash A$, and by assumption $\vdash A \to D$. So, $\Gamma \vdash D$. Again by Theorem 18, $\Gamma|D$. That is $A|D$. □

Lemma 10. *If $A|A$, then $\vdash A \to E \vee C \ \Leftrightarrow\ \vdash A \to E$ or $\vdash A \to C$, for both* F *and* IPC.

Proof. By Theorem 20, we have $A|E \vee C$, so $A|E$ or $A|C$. By Theorem 18, $A \vdash E$ or $A \vdash C$. Therefore by the weak Deduction Theorem we conclude that $\vdash A \to E$ or $\vdash A \to C$. By using F rules and axioms the other direction is easy. □

Definition 19. *A very basic implication $A \to B$ is an* ***extremely basic implication*** *if B is a sole very simple implication.*

By Lemmas 8 and 10, we can conclude that to show that IPC is conservative over F with respect to basic implications it is sufficient to do so for extremely basic implications.

Theorem 21. IPC *is conservative over* F *with respect to basic implications.*

Proof. By the above, we can assume $\nvdash_{\mathsf{F}} A$ with A an extremely basic implication:

$$A = (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n) \to (C \to D).$$

Then there exist $M, w \in M$ such that

$$M, w \nVdash (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n) \to (C \to D)$$

So, there exists $v \in M$ with wRv and $M, v \Vdash A_i \to B_i$ for each $1 \leq i \leq n$ and

$$M, v \nVdash C \to D$$

So, there exists vRu with $M, u \Vdash C$ and $M, u \nVdash D$, and, if $M, u \Vdash A_i$ then $M, u \Vdash B_i$ for each i.

Now we select the point u from M and then we make the one point model $M_{\mathsf{IPC}} = \langle u, (u, u), \Vdash^I \rangle$, such that for all propositional variables p, $M_{\mathsf{IPC}}, u \Vdash^I p$ if and only if $M, u \Vdash p$. We will show that

$$M_{\mathsf{IPC}}, u \nVdash^I (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n) \to (C \to D).$$

It is easy to see that for any conjunction or disjunction E of atoms $M_{\mathsf{IPC}}, u \Vdash^I E$ iff $M, u \Vdash E$. This applies to each of the A_i, B_i, C and D. So, $M_{\mathsf{IPC}}, u \Vdash^I C$, $M_{\mathsf{IPC}}, u \nVdash^I D$, and if $M_{\mathsf{IPC}}, u \Vdash^I A_i$, then $M_{\mathsf{IPC}}, u \Vdash^I B_i$. So, $M_{\mathsf{IPC}}, u \nVdash^I (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n) \to (C \to D)$. Therefore $\mathsf{IPC} \nvdash A$. □

To see that the conservativity result does not apply to CPC just note that the formula $(p \to q \vee r) \to (p \to q) \vee (p \to r)$ is a very basic implication which is not provable in IPC or F, but is provable in CPC.

The conservativity result for IPC over F can be extended to conjunctions and disjunctions of simple implications and basic implications. In the case of simple implications CPC can no longer take the role of IPC however as the following example shows: $\vdash_{\mathsf{CPC}} (p \to q) \vee (q \to p)$, but $\nvdash_{\mathsf{IPC}} (p \to q) \vee (q \to p)$.

It is important to note that the result cannot be extended by mixing propositional variables and implications. If we define a *mixed* implication as a formula $A \to B$ in which the A and B are obtained by applying conjunctions and disjunctions to atoms and simple implications, then Theorems 17 and 21 do not extend to this wider class. Even such a simple IPC-theorem as $p \wedge (p \to q) \to q$ cannot be proved in F.

Finally, we can state the following corollary:

Corollary 4. *Assuming the rules of modus ponens, conjunction and a fortiori the system* F *is exactly the part of* IPC *axiomatized by its simple implications and basic implications.*

Proof. Just check that all the axioms of F are simple implications or basic implications. □

4.2 A Conservativity Result for IPC over BPC

Of course, the above conservativity results apply to stronger logics than F, but for BPC we can prove an additional theorem. In this subsection we give a formal definition of NNIL formulas [11] and we will prove that IPC is conservative over BPC with respect to NNIL formulas.

Definition 20. *The smallest class satisfying the following clauses is called* NNIL*.*

1. *All propositional variables are in* NNIL*,*
2. *if* $A, B \in \mathsf{NNIL}$ *then* $A \wedge B \in \mathsf{NNIL}$*,*
3. *if* $A, B \in \mathsf{NNIL}$ *then* $A \vee B \in \mathsf{NNIL}$*,*
4. *if* $A \in \mathsf{NNIL}$ *and* B *does not contain implications, then* $B \to A \in \mathsf{NNIL}$*.*

Definition 21. *The smallest class satisfying the following is the class of normal* NNIL *formulas,*

1. *All propositional variables are in* ***normal*** NNIL*,*
2. *if* A, B *is in* ***normal*** NNIL*, then* $A \wedge B$ *is in* ***normal*** NNIL*,*
3. *if* A, B *is in* ***normal*** NNIL*, then* $A \vee B$ *is in* ***normal*** NNIL*,*
4. *if* A *is in* ***normal*** NNIL *and* B *is conjunction of atoms then* $B \to A$ *is in* ***normal*** NNIL*.*

In F we have that $A \vee B \to C$ is equivalent to $(A \to C) \wedge (B \to C)$, so any NNIL formula is provably equivalent to a normal NNIL formula.

In IPC the normal form for NNIL-formulas is simpler, in the third clause one can take B to be an atom instead of a conjunction of atoms. This relies on the fact that $\vdash_{\mathsf{IPC}} (p \to (q \to r)) \leftrightarrow (p \wedge q \to r)$. Since $\nvdash_{\mathsf{BPC}} (p \to (q \to r)) \leftrightarrow (p \wedge q \to r)$ we cannot use this simplification here.[1]

Theorem 22. *If* $A \in \mathsf{NNIL}$ *and* $\vdash_{\mathsf{IPC}} A$*, then* $\vdash_{\mathsf{BPC}} A$*.*

Proof. Let $M = \langle W, R, V \rangle$ be a model for Basic Logic BPC. Then we define the intuitionistic model $\bar{M} = \langle W, \bar{R}, V \rangle$, by $\bar{R} = R \cup \{(w, w) \mid w \in W\}$. By induction on the complexity of $A \in \mathsf{NNIL}$, we will show that for all $w \in W$, if $M, w \nVdash_{\mathsf{BPC}} A$, then $\bar{M}, w \nVdash_{\mathsf{IPC}} A$. We only check implication cases, the other cases are easy.

[1] Neither is it the case that the class of NNIL-formulas is locally finite in BPC as it is in IPC, the fact that $\nvdash_{\mathsf{BPC}} (p \to (p \to q)) \leftrightarrow (p \to q)$ quickly leads to infinitely many non-equivalent NNIL-formulas in two variables.

Let $A = \wedge p_i \to C$ and $M, w \nVdash_B \wedge p_i \to C$, then there exist $v \in W$ such that, wRv and $M, v \Vdash_B \wedge p_i$, $M, v \nVdash_B C$. Then $\bar{M}, v \Vdash_{\mathsf{IPC}} \wedge p_i$ and by induction hypothesis $\bar{M}, v \nVdash_{\mathsf{IPC}} C$. We know $v\bar{R}v$, so $\bar{M}, v \nVdash_{\mathsf{IPC}} \wedge p_i \to C$. We can conclude that $\bar{M}, w \nVdash_{\mathsf{IPC}} \wedge p_i \to C$, since we have preservation and wRv.

Now, assume $\nvdash_{\mathsf{BPC}} A$. Then by the completeness theorem for BPC there exists a BPC-model M and $w \in M$, such that $M, w \nVdash_B A$, so $\bar{M}, w \nVdash_{\mathsf{IPC}} A$. Then, by soundness $\nvdash_{\mathsf{IPC}} A$. □

Returning to the example $\nvdash_{\mathsf{F}} p \wedge (p \to q) \to q$, BPC is still not able to prove this formula: even in the case of BPC we cannot mix implications and atoms in the conservativity result.

This conservation result is to a certain extent related to the bounded translation of IPC into BPC given by Aghaei and Ardeshir [1]. There is only room here for a quick sketch. In [1] it is proved that $\vdash_{\mathsf{IPC}} A$ iff $\vdash_{\mathsf{BPC}} [A]^n$ for some n where $[\,]^n$ is the n-th iterate of $[\,]^1$ (and a bound for the n is provided). By an extension of the method of the proof of Theorem 22 it can be shown that for NNIL-formulas A in particular, $\vdash_{\mathsf{BPC}} A$ iff $\vdash_{\mathsf{BPC}} [A]^1$ and hence also $\vdash_{\mathsf{BPC}} A$ iff $\vdash_{\mathsf{BPC}} [A]^n$. Theorem 22 then follows from the translation result.

5 Conclusion

We developed the subintuitionistic logics introduced by Corsi and Restall in a uniform manner. Proof systems for Corsi's basic system F are given for derivations without and with assumptions, and completeness theorems are proved, clarifying the role of the rules of modus ponens, conjunction and a fortiori.

Restall's notion of proof from a theory Π is then developed in the same manner. A related notion of proof corresponding to Restall's validity notion is extracted from Restall's paper leading to a transparent form of Restall's completeness theorem and an extension to a general completeness theorem that is related to but not the same as a completeness theorem proved by Celani and Jansana [2] thereby establishing a link between Restall's work and [2].

The more important results of this paper are the conservation results. First two classes of formulas are introduced, simple implications and basic implications, and the completeness results are then used to give a conservation theorem for intuitionistic logic IPC over Corsi's system F with respect to these two classes of formulas. In several ways this is shown to be a best result. For BPC an additional conservation result is given with respect to the class of NNIL formulas.

Our methods have been fruitfully used in investigations of a weaker logic WF characterized by neighborhood models. The system WF as developed by us is presumably the minimal one with a neighborhood semantics and is obtained by weakening the axioms 8, 9 and 11 to the corresponding rules [8]. The conservativity result applies in that case only to the simple implications, clearly showing the difference in strength of the logics F and WF.

Clearly, it is important now to further study BPC and the other known logics between F and IPC as well as the logics between F and WF with regard to conservativity properties. Also, the connection between Restall's and Celani and Jansana's consequence relations deserves further study.

Acknowledgement. We thank M. Ardeshir for important comments on the paper. We are grateful to two unknown referees who provided valuable remarks and corrections.

References

1. Aghaei, M., Ardeshir, M.: A bounded translation of intuitionistic propositional logic into basic propositional logic. Math. Logic Q. **46**(2), 195–206 (2000)
2. Celani, S., Jansana, R.: A closer look at some subintuitionistic logics. Notre Dame J. Formal Logic. **42**(4), 225–255 (2001)
3. Corsi, G.: Weak Logics with strict implication. Z. Math. Logik Grundlagen Math. **33**(5), 389–406 (1987)
4. Došen, K.: Modal translations in K and D. In: de Rijke, M. (ed.) Diamonds and Defaults. Synthese Library, vol. 229, pp. 103–127. Springer, Dordrecht (1994)
5. Kleene, S.C.: Disjunction and Existence under Implication in elementary Intuitionistic formalisms. J. Symbolic Logic **27**, 11–18 (1962)
6. Restall, G.: Subintuitionistic logics. Notre Dame J. Formal Logic **35**(1), 116–129 (1994)
7. Sano, K., Ma, M.: Alternative semantics for visser's propositional logics. In: Aher, M., Hole, D., Jeřábek, E., Kupke, C. (eds.) TbiLLC 2013. LNCS, vol. 8984, pp. 257–275. Springer, Heidelberg (2015). doi:10.1007/978-3-662-46906-4_15
8. Shirmohammadzadeh Maleki, F., de Jongh, D.: Weak Subintuitionistic Logics. Logic J. IGPL (2016). doi:10.1093/jigpal/jzw062
9. Suzuki, Y., Ono, H.: Hilbert-style proof system for BPL. Technical report, IS-RR-97-0040F. Japan Advanced Institute of Science and Technology (1997)
10. Visser, A.: A propositional logic with explicit fixed points. Studia Logica. **40**(2), 155–175 (1981)
11. Visser, A., van Benthem, J., de Jongh, D., Renardel de Lavalette, G.: NNIL, a study in intuitionistic propositional Logic. In: Ponse, A., de Rijke, M., Venema, Y. (eds.) Modal Logics and Process Algebra, a Bisimulation Perspective, pp. 289–326 (1995)

Erratum to: Equilibrium Semantics for IF Logic and Many-Valued Connectives

Christian G. Fermüller[1(✉)] and Ondrej Majer[2]

[1] Theory and Logic Group, Vienna University of Technology, Vienna, Austria
chrisf@logic.at

[2] Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic
majer@flu.cas.cz

Erratum to: Chapter "Equilibrium Semantics for IF Logic and Many-Valued Connectives" in: H.H. Hansen et al. (Eds.): Logic, Language, and Computation, LNCS, DOI: 10.1007/978-3-662-54332-0_16

The original version of this chapter contained two mistakes. The affiliation and grant number of the author Ondrej Majer were incorrect. The original chapter was corrected.

1. The affiliation of Ondrej Majer is given below.
   Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic
   majer@flu.cas.cz
2. The footnote on page 290 must read as follows:
   O. Majer—Supported by Czech Science Foundation grant GF15-34650L.

C.G. Fermüller—Supported by Austrian Science Foundation (FWF) grant I1897-N25 (MoVaQ-MFL).
O. Majer—Supported by Czech Science Foundation grant GF15-34650L.

The updated original online version for this chapter can be found at
DOI: 10.1007/978-3-662-54332-0_16

H.H. Hansen et al. (Eds.): TbiLLC 2015, LNCS 10148, p. E1, 2017.
DOI: 10.1007/978-3-662-54332-0_19

Author Index