Guram Bezhanishvili
Sebastian Löbner
Vincenzo Marra
Frank Richter (Eds.)

LNCS 7758

Logic, Language, and Computation

9th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2011
Kutaisi, Georgia, September 2011
Revised Selected Papers

Springer

Guram Bezhanishvili Sebastian Löbner
Vincenzo Marra Frank Richter (Eds.)

Logic, Language, and Computation

9th International Tbilisi Symposium
on Logic, Language, and Computation, TbiLLC 2011
Kutaisi, Georgia, September 26-30, 2011
Revised Selected Papers

Springer

Volume Editors

Guram Bezhanishvili
New Mexico State University
Department of Mathematical Sciences
Las Cruces, NM 88003, USA
E-mail: gbezhani@math.nmsu.edu

Sebastian Löbner
Heinrich Heine Universität
Department of Language and Information
Kruppstrasse 108, 40227 Düsseldorf
E-mail: loebner@phil.hhu.de

Vincenzo Marra
Università degli Studi di Milano
Dipartimento di Matematica "Federigo Enriques"
via Cesare Saldini, 50, 20133 Milan, Italy
E-mail: vincenzo.marra@unimi.it

Frank Richter
Universität Tübingen
Seminar für Sprachwissenschaft
Wilhelmstraße 19, 72074 Tübingen, Germany
E-mail: fr@sfs.uni-tuebingen.de

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-36975-9 e-ISBN 978-3-642-36976-6
DOI 10.1007/978-3-642-36976-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013932488

CR Subject Classification (1998): I.2.7, F.4.1, F.4.3

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

The Tbilisi Symposium on Logic, Language and Computation is an international biennial conference series organized jointly by the Center for Language, Logic and Speech (CLLS) of Tbilisi State University, the Georgian Academy of Sciences, and the Institute for Logic, Language, and Computation (ILLC) of the University of Amsterdam. The conference series is centered around the interaction between logic, language, and computation. The symposia are open to contributions from any of these three fields, and they aim at fostering interaction among them. The ninth edition of the symposium was held in Kutaisi, Georgia, September 26–30, 2011. This volume contains a selection of papers presented at the meeting.

The Tbilisi symposia are renowned not only for their high scientific standards, but also for their unforgettable social atmospheres. The ninth edition continued this tradition. It was held in Georgia's second largest city Kutaisi, the capital of the beautiful Imereti region. The scientific program consisted of tutorials, invited and contributed talks, and special sessions in the three major areas of the conference—logic, language, and computation.

The symposium offered three tutorials. One on language, by Daniel Hole (Humboldt University, Berlin), one on logic, by Vincenzo Marra (Università degli Studi di Milano), and one on computation, by Ulle Endriss (ILLC, Amsterdam). The tutorials were aimed at local students attending the conference as well as researchers from the remaining areas.

There were seven invited talks delivered at the symposium. One on language, by Peter beim Graben (Humboldt University, Berlin), three on logic, by Alexandru Baltag (ILLC, Amsterdam), Agi Kurucz (King's College London), and Sonja Smets (ILLC, Amsterdam), and three on computation, by Nikolaj Bjørner (Microsoft Research, Redmond), Christof Monz (Informatics Institute, Amsterdam), and Prakash Panangaden (McGill University, Montreal).

In addition, the symposium hosted two special sessions, one on frames, organized by Sebastian Löbner (Düsseldorf University), and one on logic, information, and agency, organized by Alexandru Baltag and Sonja Smets. The special sessions had their own invited speakers and contributed talks. Each session had three invited speakers. The one on frames had Nicola Guarino (National Research Council, Trento), Frank Richter (Tübingen University), and Manfred Sailer (Frankfurt University); and the one on logic, information, and agency had Dietmar Berwanger (CNRS, Cachan), Jan van Eijck (ILLC, Amsterdam), and Alessandra Palmigiano (ILLC, Amsterdam).

This proceedings volume contains the summaries of the tutorials, the contributed papers on language, and the contributed papers on logic and computation. Here we will provide a brief overview of the contributed papers.

The first group of papers consists of contributions on language. Rusudan Asatiani provides an analysis of passive constructions in Georgian. Asatiani argues that passive in Georgian can best be described in light of features derived from cognitive semantics rather than syntax. A broad array of data are considered, and the proposal is supported by production experiments conducted with Georgian informants.

Anton Benz and Fabienne Salfner investigate the dependencies that hold in discourse between implicatures and questions under discussion. In particular, they argue that not only do relevance implicatures depend on the question under discussion, but also scalar implicatures, which were previously thought to be derivable from logical forms alone.

The contribution by Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen addresses an extension of the framework of inquisitive semantics. Inquisitive semantics pursues the goal of making the notion of *information exchange* logically precise by adding the notion of *compliance* to capture the pertinence of a response in discourse. Their shift from inquisitive semantics to inquisitive witness semantics is designed to overcome shortcomings of previous formulations of *compliance*.

In their paper, Thomas Gamerschlag, Wiebke Petersen, and Liane Ströbel investigate three essential German posture verbs on the basis of frame representations. They demonstrate that their relationship to important cognitive modules makes them very suitable for taking frame semantic notions as the basis for careful decompositional analyses that can be argued to be cognitively plausible.

Bjørn Jespersen and Giuseppe Primiero observe that modal modifiers like *alleged* are standardly characterized in terms of failing inferences in comparison to other modifiers. They suggest two positive definitions, the first phrased in terms of Tichý's transparent intensional logic, and the second in terms of an extension of Martin-Löf's constructive type theory. The authors investigate some prominent consequences of their definitions.

Ralf Naumann's paper presents an extension of the theory of frames. Starting with a static theory of frames defined as Kripke models, Naumann suggests using operations on such models in order to capture their temporal development and refinements, and he shows how the resulting system can be applied in the analysis of the so-called dative alternation in English.

Umut Özge sets out to shed light on an open question of the semantic analysis of indefinites. In the framework of discourse representation theory, he investigates the semantics of Acc-indefinites in Turkish to determine their presuppositional behavior, thereby addressing the question of how the concepts of D-linking and existential import are related, both of which have been evoked in this context.

The second group of papers consists of contributions on logic and computation. Patrick Allo introduces a theory of defeasible inference, i.e., a non-monotonic reasoning by default that may be informally described as "tentative defeasible inference." This is motivated by an analysis of the classic logical omniscience problem in the context of default reasoning. Allo's proposal is formalized within the framework of adaptive logics.

The paper by Samuel Bucheli, Roman Kuznets, and Thomas Studer deals with justification logics, modal-like systems that explicitly include justifications for the agents' knowledge. Bucheli et al. introduce model-theoretic tools, such as filtrations and generated submodels, for the investigation of these logics. They thereby obtain uniform proofs of known decidability results, along with some new results as well.

Francien Dechesne and Mohammad R. Mousavi offer a new semantics for the formalism of process algebras. The latter were introduced as a means to specify the behavioral aspects of protocols in computer science. The semantics advanced by the authors is given in terms of interpreted systems, which have been used as models of multi-agent communication. The authors' paper thus provides a link between these two worlds.

Johannes Ebbing, Peter Lohmann, and Fan Yang investigate the computational complexity of the model-checking problem for an extension of Väänänen's modal dependence logic, obtained by adding to it the Boolean disjunction and the intuitionistic implication. The authors obtain results on several fragments of this logic—they show, for instance, that the model-checking problem for the full system is PSPACE-complete.

Tadeusz Litak, Dirk Pattinson, and Katsuhiko Sano provide a sequent calculus for the recently introduced coalgebraic predicate logic (CPL)—a formalism that has a wide range of applications, including probabilistic logic, coalition logic, and the logic of neighborhood frames. Among other things, they show that CPL is equipollent to the coalgebraic hybrid logic with the downarrow binder and universal modality.

Monica Dinculescu, Christopher Hundt, Prakash Panangaden, Joelle Pineau, and Doina Precup consider the problem of representing and reasoning about transition systems with hidden state. They show how to obtain a dual system by interchanging the notions of state and observation. For deterministic systems, the double dual gives a minimal representation of the behavior of the system. This is then extended to probabilistic transition systems and to partially observable Markov decision processes.

We would like to thank the contributing authors and reviewers alike, without whose hard work this volume would not have been possible. We are also very grateful to Matthias Baaz, Johan van Benthem, Ulle Endriss, Sebastian Löbner, and Peter van Ormondt for obtaining funding to support the conference.

We would like to conclude by remembering Dito Pataraia (1963–2011). Dito was an outstanding representative of the Georgian mathematical community. He was an indispensable source of inspiration, and shared generously his uniquely original insights on several important topics in category theory, mathematical logic, homological algebra, and algebraic topology. He had an amazing ability to effortlessly present complex and deep mathematical insights with incredible simplicity and clarity. He is best known for two important contributions: He found a constructive proof of the fixed point theorem for dcpo's, and had a major breakthrough in showing that each Heyting algebra occurs as the poset of all subobjects of some object in an elementary topos. He also contributed to the study of Hopf algebras, Hopf-algebraic knot theory, the fundamental group of a topos, and compact Hausdorff Boolean algebras.

Among his colleagues, Dito will be remembered for his extremely original ideas, unexpected intriguing constructions, and ingenious proofs; and among his friends, for his extraordinary kindness, unfailing humor and sharp, but open and unassuming mind. Unfortunately, TbiLLC 2011 turned out to be the last conference attended by Dito. We would like to dedicate these proceedings to Dito's memory.

December 2012

Guram Bezhanishvili
Sebastian Löbner
Vincenzo Marra
Frank Richter

Organization

The ninth Tbilisi Symposium on Language, Logic and Computation was held in Kutaisi, Georgia, September 26–30, 2011. The symposium was organized by the Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam in conjunction with the Centre for Language, Logic and Speech (CLLS) of Tbilisi State University, the Georgian Academy of Sciences, and the Akaki Tsereteli State University, Kutaisi.

Organizing Committee

Rusudan Asatiani	Tbilisi (Co-chair)
Paul Dekker	Amsterdam (Co-chair)
David Gabelaia	Tbilisi
Marina Ivanishvili	Tbilisi
Nino Javashvili	Tbilisi
Ali Karatay	Istanbul
Zaal Kikvidze	Kutaisi
Liana Lortkipanidze	Tbilisi
Peter van Ormondt	Amsterdam
Khimuri Rukhaia	Tbilisi
Nana Shengelaia	Tbilisi

Program Committee

Matthias Baaz	Vienna (Chair)
Guram Bezhanishvili	Las Cruces (Co-chair Logic)
Peter Bosch	Osnabrück (Co-chair Language)
Nani Chanishvili	Tbilisi (Language)
George Chikoidze	Tbilisi (Language)
Pauline Jacobson	Providence (Language)
Temur Kutsia	Linz (Co-chair Computation)
Jerome Lang	Paris (Co-chair Computation)
Sebastian Löbner	Düsseldorf (Language)
Luca Spada	Salerno (Co-chair Logic)
Henk Zeevat	Amsterdam (Co-chair Language)

Tutorials

Ulle Endriss	Amsterdam (Computation)
Daniel Hole	Berlin (Language)
Vincenzo Marra	Milan (Logic)

Invited Speakers

Alexandru Baltag	Amsterdam (Logic)
Nikolaj Bjørner	Redmond (Computation)
Peter beim Graben	Berlin (Language)
Agi Kurucz	London (Logic)
Christof Monz	Amsterdam (Computation)
Prakash Panangaden	Montreal (Computation)
Sonja Smets	Amsterdam (Logic)

Workshop on Frames

Wiebke Petersen	Düsseldorf
Nicola Guarino	Trento (Invited)
Rainer Osswald	Düsseldorf
Sebastian Löbner	Düsseldorf
Tanja Osswald	Düsseldorf
Wiebke Petersen	Düsseldorf
Frank Richter	Tübingen (Invited)
Manfred Sailer	Frankfurt (Invited)

Workshop on Logic, Information, and Agency

Patrick Allo	Brussels
Dietmar Berwanger	Cachan (Invited)
Francien Dechesne	Delft
Jan van Eijck	Amsterdam (Invited)
Prakash Panangaden	Montreal
Tadeusz Litak	Erlangen
Alessandra Palmigiano	Amsterdam (Invited)
Jonas De Vuyst	Brussels

Sponsoring Institutions

Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam
Centre for Language, Logic and Speech (CLLS) of Tbilisi State University
Georgian Academy of Sciences
Akaki Tsereteli State University, Kutaisi

Table of Contents

Part 1: Summaries of Tutorials

Part 2: Contributions in Language

Part 3: Contributions in Logic and Computation

Computational Social Choice (with a Special Emphasis on the Use of Logic)

Ulle Endriss

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam

Abstract. This is a summary of a tutorial on computational social choice, delivered at the *9th Tbilisi Symposium on Language, Logic and Computation* in Kutaisi, Georgia, in September 2011. The tutorial specifically focussed on the use of logic in the field.

1 Introduction

Social choice theory deals with questions regarding the design and analysis of methods for collective decision making [1]. In recent years there has been a growing interest in the computational aspects of collective decision making, giving rise to the field of *computational social choice* [5,4]. This tutorial provided an introduction to this field, highlighting in particular the role of logic. The first lecture was devoted to an exposition of the axiomatic method in social choice theory; the second lecture was about social choice in combinatorial domains; and the third lecture was an introduction to judgment aggregation.

Below we briefly introduce each of these three topics and provide references for further reading. Full details are available elsewhere [7].

2 The Axiomatic Method in Social Choice Theory

Much of the classical work in social choice theory is concerned with the formalisation of normative intuitions about the proper way of aggregating preferences by stating so-called *axioms* and with the investigation of the consequences of those axioms. The best-known example is Arrow's Theorem [2].

In Arrow's model, n individuals each express a preference over a set of alternatives $\mathcal{X}$ (where a preference is a linear order $\succ$ over $\mathcal{X}$). To aggregate this into a collective preference order that accurately reflects the views of the group, we are looking for a function F from n-tuples of linear orders on $\mathcal{X}$ to a single linear order on $\mathcal{X}$. Arrow argued that F should satisfy the following axioms:

- The *Pareto condition:* in case every individual agrees that $x \succ y$, then also the collective preference order should stipulate $x \succ y$.
- *Independence of irrelevant alternatives:* whether $x \succ y$ holds for the collective preference should only depend on the relative ranking of x and y in the individual orders (and not on, say, how many individuals agree with $x \succ z$).

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 1–3, 2013.

– *Absence of dictators:* it should not be the case that there exists one fixed individual such that the collective preference order is always identical to that individual's preference order, independently of what the others say.

Arrow's deeply surprising and thought-provoking theorem states that it is *impossible* to simultaneously satisfy these requirements—at least in case there are more than just two alternatives to rank [2]. A number of different proofs of Arrow's Theorem are known. What might be of particular interest to logicians is that some proofs make use of concepts from the theory of ultrafilters, and also that there have been several attempts to automate the process of proving this and similar results. One proof technique that is particularly helpful in understanding the issues raised by Arrow's result is based on the notion of a *decisive coalition*—it amounts to investigating the structure of the family of coalitions (sets of individuals) who together can determine the relative ranking of two given alternatives, independently of the preferences of the other individuals [7].

3 Social Choice in Combinatorial Domains

For many applications the set of alternatives the individuals are asked to express preferences over will have a combinatorial structure. For instance, if we are asked to elect a committee of size k by choosing from a field of n candidates, then there are $\binom{n}{k}$ possible committees; and in a referendum in which voters are asked to accept or reject n different propositions there are 2^n possible outcomes.

Because of this combinatorial explosion, we need to carefully choose the *language* used to represent the preferences of individuals. This is why work on social choice in combinatorial domains often borrows ideas from knowledge representation and reasoning [6]. In particular, several logic-based languages for preference modelling are used in the field [11]. Three approaches are particularly promising:

– *Distance-based voting:* The basic idea is to ask each individual for her most preferred combination and then make a collective choice that minimises the aggregated distance to the combinations—for a suitable notion of distance and, in some cases, restricting attention to a particular set of admissible outcomes. Examples include the *minimax rule* [3] and the *average-voter rule* [8].
– *Sequential voting:* Here the idea is to vote on each issue in turn, rather than to vote on all of them at the same time, which reduces the complexity of the problem as well as the potential for paradoxical outcomes [10,12].
– *Combinatorial vote:* The central idea in this approach is that individuals encode their preferences using a compact preference representation language and transmit an expression in the language of choice rather than an explicit preference order [11]. The main challenge is to develop efficient algorithms that are able to aggregate the compactly represented preferences.

4 Judgment Aggregation

Preferences are not the only structures we may wish to aggregate. *Judgment aggregation* studies the aggregation of judgments on mutually dependent propositions [13,9]. These dependencies are often modelled using propositional logic.

For instance, suppose Alice judges both p and q to be true, Bob judges p to be true and q to be false, and Carlo judges p to be false and q to be true. How should we aggregate this information? Going with the obvious choice of majority voting is problematic: there is a majority for p as well as a majority for q, but at the same time there is a majority *against* $p \wedge q$.

List and Pettit [13] were the first to treat this problem formally. They showed that it is in fact *impossible* to devise a method of aggregation that would treat all individuals as well as all propositions symmetrically (anonymity and neutrality), for which the collective judgment on a proposition solely depends on the individual judgments on that same proposition (independence), and for which the outcome is always a consistent set of formaulas including, for every formula φ under consideration, either φ or $\neg\varphi$. Subsequent work has investigated topics such as the connections to preference aggregation and Arrow's Theorem, ways of circumventing the impossibility, precise conditions on the set of formulas to be judged for impossibilities to arise, specific aggregation rules, and the computational complexity of working in the framework of judgment aggregation [9].

References

1. Arrow, K.J., Sen, A.K., Suzumura, K. (eds.): Handbook of Social Choice and Welfare. North-Holland (2002)
2. Arrow, K.J.: Social Choice and Individual Values, 2nd edn. John Wiley and Sons (1963)
3. Brams, S.J., Kilgour, D.M., Sanver, M.R.: A minimax procedure for electing committees. Public Choice 132(3-4), 401–420 (2007)
4. Brandt, F., Conitzer, V., Endriss, U.: Computational social choice. In: Weiss, G. (ed.) Multiagent Systems. MIT Press (2013) (in press)
5. Chevaleyre, Y., Endriss, U., Lang, J., Maudet, N.: A Short Introduction to Computational Social Choice. In: van Leeuwen, J., Italiano, G.F., van der Hoek, W., Meinel, C., Sack, H., Plášil, F. (eds.) SOFSEM 2007. LNCS, vol. 4362, pp. 51–69. Springer, Heidelberg (2007)
6. Chevaleyre, Y., Endriss, U., Lang, J., Maudet, N.: Preference handling in combinatorial domains: From AI to social choice. AI Magazine 29(4), 37–46 (2008)
7. Endriss, U.: Logic and social choice theory. In: Gupta, A., van Benthem, J. (eds.) Logic and Philosophy Today, vol. 2, pp. 333–377. College Publications (2011)
8. Grandi, U.: Binary Aggregation with Integrity Constraints. Ph.D. thesis, ILLC, University of Amsterdam (2012)
9. Grossi, D., Pigozzi, G.: Introduction to Judgment Aggregation. In: Bezhanishvili, N., Goranko, V. (eds.) ESSLLI 2010/2011, Lectures. LNCS, vol. 7388, pp. 160–209. Springer, Heidelberg (2012)
10. Lacy, D., Niou, E.M.S.: A problem with referendums. Journal of Theoretical Politics 12(1), 5–31 (2000)
11. Lang, J.: Logical preference representation and combinatorial vote. Annals of Mathematics and Artificial Intelligence 42(1-3), 37–71 (2004)
12. Lang, J., Xia, L.: Sequential composition of voting rules in multi-issue domains. Mathematical Social Sciences 57(3), 304–324 (2009)
13. List, C., Pettit, P.: Aggregating sets of judgments: An impossibility result. Economics and Philosophy 18(1), 89–110 (2002)

Binding – Data, Theory, Typology

Daniel Hole

Humboldt University, Berlin, Germany
holedan@googlemail.com

Abstract. The tutorial gave an overview of the treatment of variable binding in natural language semantics. A set of data was singled out, two families of approaches to deal with reflexivity were presented which yield a comparable data coverage, and the cross-linguistic variation of reflexivization strategies was reviewed. The modelling options map neatly onto the variation found in natural language.

Keywords: variable binding, reflexivity, pronoun, semantics, typology.

1 Introduction

The tutorial gave a three-fold overview of aspects pertaining to the issue of variable binding in natural language. In a first step, a set of phenomena was singled out, phenomena which constitute the core data in the domain under scrutiny. Each binding theory must be able to account for this set of data. Section 2 of the present article reviews some of these data patterns. In a second step, two families of analyses used to model reflexivity were introduced (section 3 of the present article). Reflexivity was used as a domain of illustration, because this phenomenon constitutes a widely discussed paradigm case within the larger domain of variable binding phenomena in natural language. The first kind of analysis centers around the reflexivization of verbs. The second kind of analysis leads to the reflexivization of larger constituents; it requires powerful composition tools that go well beyond functional application. The cross-linguistic overview of the third part of the tutorial aimed at showing that, in all likelihood, both families of theories are justified if a close form-function match is aimed at (section 4 of the present article).

2 Data and Descriptive Generalizations

2.1 Three Uses of Pronouns

(1) lists examples of different pronoun uses that occur side by side in many languages. Some of them have a long tradition of being analyzed as natural language counterparts of bound variables (Ross 1967).

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 4–11, 2013.

(1) a. **anaphoric**
A boy came in. He wore a red hat.
b. **deictic**
Look, the two over there!
[pointing:] *She*$_i$*'s my boss, and she*$_j$*'s my colleague.*
c. **bound**
i. *Paul*$_i$*/Everybody*$_i$ *likes himself*$_i$.
ii. *Mary*$_i$/[*None of the girls*]$_i$ *thinks she*$_i$*'s a genius.*

It depends on the individual grammar framework whether the pronouns in (1a/b) are analyzed as bound variables. The pronouns in (1c) will be analyzed as bound variables in the great majority of frameworks. What sets them apart from the examples in (1a/b) is that they have an overt antecedent in the same sentence. With a formal understanding of variable binding in mind, it is maybe not immediately clear why proper name antecedents as in (1c) should count as variable binders. It will not be possible to elucidate in this short survey how proper names can be expressions referring to individuals, and still be variable binders in a formal sense at the same time. Suffice it to say here that the most influential proposal in this domain manages to reconcile these two things (Heim and Kratzer 1998). With the quantified subject variants in (1c) (*everybody*, *none of the girls*), a variable binding analysis offers itself straightforwardly.

Another issue to comment on in connection with (1cii) is the fact that *she* (with antecedent *Mary*) has the form of a pronoun which may occur without sentence-internal antecedents. This may nourish suspicion about its bound status in (1cii); could one not say instead that *she* may refer to any salient discourse antecedent, and this antecedent just happens to be the same referent as who *Mary* refers to? Put differently, does one have to postulate a difference between *he* in (1a) and *she* in (1cii)? The following subsection will introduce the diagnostics to establish the fact that there is a difference.

2.2 Strict and Sloppy Identity: Some Classic Contrasts

The sentence in (2) has at least three different readings. Names for these readings are introduced in (i)-(iii) (Ross 1967).

(2) *Paul likes his teacher, and Peter does* [~~*like his teacher*~~]$_{\text{ELLIPSIS}}$, *too.*
i. 'Paul likes Paul's teacher, and Peter likes Peter's teacher.
sloppy identity (bound use of of *his*)
ii. 'Paul likes Paul's teacher, and Peter likes Paul's teacher.'
strict identity (anaphoric use of *his*; it co-refers with *Paul*)
(iii. 'Paul$_i$ likes c$_j$'s teacher, and Peter$_k$ likes c$_j$'s teacher.'
'third reading' (anaphoric use of *his* not co-referent with the subject; really a special case of the more general case to which (ii) belongs; Büring 2005)

The elliptical possessor in the second conjunct may either co-vary with the local antecedent (binding/sloppy identity), or be fixed to a single referent (the subject referent as with strict identity, or some discourse-given referent as with the 'third reading'). Sloppy identity phenomena invite analyses in terms of variable binding. Note in passing that, if the second conjunct is disregarded, strict and sloppy identity construals make no difference in the first conjunct. Which analysis should, then, be chosen for such sentences without second conjuncts? Büring (2005: 121) argues that natural languages generalize the binding construal.

There are classes of pronouns which force binding construals. Reflexive pronouns like English *x-self* are like this. This is illustrated in (3).

(3) *Mary pinched herself, and Paula did, too.*
 i. 'Mary pinched Mary, and Paula pinched Paula.' (sloppy identity)
 *ii. 'Mary pinched Mary, and Paula pinched Mary.' (*strict identity)
 *iii. 'Mary pinched Sue, and Paula pinched Sue.' (*3rd)

Apart from the lexical class of the pronoun at hand, there are certain syntactic restrictions which have to be fulfilled for a pronoun to receive a bound reading as in (2i) or (3) (command relations such as c-command or o-command, depending on the grammar framework chosen; Büring 2005). For lack of space, we will not go into the syntax of variable binding.

The next section will briefly sketch two ways of arriving at bound-variable construals in a compositional semantics, implemented for the empirical domain of reflexivity.

3 Bound Variables and Reflexivity in a Compositional Semantics

3.1 Verb-Centered Reflexivization

A reflexive clause with a referring expression as its subject as in (3) is characterized by the fact that the subject referent and a second participant of the event described by the verb are identical. Importantly, the identical reference of the two arguments is of the kind which produces sloppy-identity effects in diagnostic contexts such as (3). The reference of the clausal subject is biconditionally linked to a second event participant. In a compositional semantics, it is a natural move to implement this biconditional link at the level of verb meanings, because verb meanings allow simultaneous access to the argument positions of the subject and the object (this only holds *cum grano salis*; in an agent-severed event semantics as propagated by Kratzer (1996), verb meanings have no direct access to the agent argument position). The first of the two major modelling options for reflexivity centers around this property of verbs (Keenan 1987, Jacobson 1999).

(4a) gives a simplified lexical entry of the verb *pinch*, represented as a lambda-term.[1] (4b) is its reflexivized counterpart.

(4) a. $\lambda x_e . \lambda y_e$. y pinches x
'the (smallest) function which maps every x, x an individual of semantic type e, to the smallest function which maps every y, y an individual of semantic type e, to 1 if y pinches x, and to 0 otherwise'
[a function from individuals to [a function from individuals to truth-values]]

b. λx_e . x pinches x
'the (smallest) function which maps every x, x an individual of semantic type e, to 1 if x pinches x, and to 0 otherwise'
[a function from individuals to truth-values]

(5a) represents a function which takes transitive verb meanings as in (4a) as input and yields reflexivized verb meanings as in (4b) as output. (5b) applies this function to the denotation of *pinch*. This is one way to arrive at reflexivized verb meanings.

(5) a. $\lambda f_{\langle e,\langle e,t\rangle\rangle} . \lambda x_e . f(x)(x)=1$
'a function which reflexivizes transitive verbs'

b. $\lambda f_{\langle e,\langle e,t\rangle\rangle} . \lambda x_e . f(x)(x)=1[\lambda x_e . \lambda y_e$. y pinches x]
= λx_e . x pinches x

If a denotation as in the last line of (5b) is available as the term with which the subject argument combines by Functional Application, then sloppy identity readings may be derived; the object denotation will co-vary with whatever is the (local) subject.[2] We will return to the merits and deficits of this general account of reflexivization in section 4. In the following subsection we will turn to the second family of reflexivization theories.

3.2 Pronoun-and-VP-Centered Reflexivization

The second way of modelling reflexivity does not take the verb as its starting point, but the pronoun in object position, in conjunction with a mechanism which operates

---

1 Transitive verbs like *pinch* are represented as schönfinkeled, or curried, functions by many semanticists (cf. Heim and Kratzer 1998: 29-31). This means that a transitive predicate is not modelled as a function which takes a pair as its argument. Instead, transitive verbs take a single individual argument and yield another function as output. This second function then yields a truth-value as output (or whatever intermediate level the semantics assumes at the level where the highest argument has been filled in for the first time). With such a nested functional structure, denotations become available for each node in a syntactic tree with binary branching.

2 It is usually assumed that ellipsis as in the diagnostic contexts of (2) and (3) requires exact identity of structure and interpretation between the elided and the non-elided counterparts. If this is assumed, then (5b) is a denotation of the right kind.

across a distance. With a cross-linguistically common type of reflexive pronoun, a clause will only be grammatical if the reflexive pronoun co-refers with the subject of the clause. If one insists on implementing the idea that it is not the verb which establishes the 'identity link' between subject and object, but the object pronoun, then some other mechanism must kick in to yield the desired result. The mechanism which delivers this result in current semantic binding theories which do not rely on reflexivized verbs is predicate abstraction, or variants thereof (Heim and Kratzer 1998, Büring 2005, Hole 2008). Due to limitations of space we will not be able to delve into the details of any specific semantic binding theory. I will present a highly simplified narrative of what object-and-VP-centered implementations of reflexivization amount to. The interested reader is referred to the literature for further details.

The narrative goes like this. (i) Pronouns come with a birthmark, or index, that allows one to identify their referent in a given context. The birthmark is visible in the syntactic structure, and it gets interpreted as the individual which has it. As such, pronouns refer in a way that is not very different from proper names (if a Rigid Designator account of proper names is adopted; Kripke 1972), except that pronouns get recycled again and again (whereas there are as many fictitiously different names *Paul* as there are individuals called Paul). A proper name refers to exactly one indivdual, or group of individuals. A pronoun refers to exactly one individual, or group of individuals, per context. (ii) Some mechanisms introduce bare birthmarks, birthmarks that are not interpreted as individuals. If, in the larger linguistic structure, the pronoun birthmark and the second birthmark happen to be the same, and if the command relations alluded to at the end of 2.2 are fulfilled between this second occurrence of the birthmark and the first one, then a predicate is abstracted over the whole constituent minus the second, commanding, birthmark. (iii) Pronoun classes differ in their ability to undergo predicate abstraction. Some must undergo it (reflexive pronouns), some may (English possessive pronouns), some may not (English ordinary personal pronouns in object position with the binder-to-be in the local subject position). (6) illustrates major components of this narrative (○ symbolizes the birthmark (index)).

(6) [*Paul* [○ [*pinched* *himself*$_{○}$]]]

'Paul' — no linguistic meaning — '$\lambda x_e . \lambda y_e$. y pinched x' — whoever ○ stands for in a given context, say, Ed (*sic!*)

A: 'λy_e . y pinched whoever ○ stands for'

B: 'λy_e . y pinched y'

true iff Paul pinched Paul

The difficult step, the one which cannot be achieved by functional application, is the one from A to B. The birthmark argument must be 're-extracted', and the resulting argument slot must be identified with the subject argument slot. At this point the different proposals mentioned above all use slightly different tools. None of them makes do without some costly or inelegant mechanism. The last lambda-term in (6) is again of the right type to derive sloppy-identity/bound-variable construals.

In the last section I will discuss the empirical justification for assuming both types of reflexivization theories side by side.

4 A Typology of Reflexivization Strategies in Natural Language

The verb-centered implementation of reflexivization is simple, but it leaves no room for reflexive pronouns. Speakers have the intuition that a pronoun like *himself* in English refers, and this is not predicted by verb-centered theories. So we do need a pronoun-centered construal of reflexivity. In some other languages, no pronouns are required to construct a reflexive clause. An affix on the verb, or some other morphological mechanism, reflexivizes the predicate instead. This is precisely what one expects if one endorses a verb-centered approach to reflexivization. Examples from languages with clear verbal reflexivization strategies are provided in (7) (taken from Gast et al. 2007).

(7) a. Shona (Niger-Congo; Volta-Congo)
á-ká-zvi-rwádzísá
NOUNCLASS1.3SG-PAST-REFLEXIVE-suffer.CAUSATIVE
'He hurt himself.'

b. Abkhaz (North(west)-Caucasian; Abkhaz-Abasin)
sarà s-ʈʂə̀-s-š-we-yt'
ich POSSESSIVE.1.SG-REFLEXIVE-1.SG-kill-DYNAMIC-FINITE
'I kill myself.'

c. Classical Nahuatl (Uto-Aztecan; Aztecan)
mo-tlaʔsoʔtla
REFLEXIV.3-lieb
'He/She loves him-/herself.'/'They love themselves.'

Upon closer inspection, it is a typical feature of European languages to have two reflexivization strategies, one of them pronominal, and the other one similar to a verbal reflexivization strategy. (The case of English is not covered by this generalization.) Typically, one finds a reflexivization strategy with a reflexive pronoun which may be stressed, which can move in the sentence and which is canonically used with verbs describing typically other-directed actions like 'hating', 'criticizing', or 'attacking' (cf. König and Vezzosi 2004 for the notion of (non-)other-directedness). The (more) verbal reflexivization strategy of European languages makes use of a bleached pronominal element which is restricted to a position adjacent to the verb, which cannot be stressed and which is canonically used with typically self-directed actions (body-care, grooming; cf., again, König and Vezzosi 2004). The Russian pair *sebja* vs. *-sja* is a case in point, Italian *se (stesso)* vs. *si* another one. (8) provides examples from Dutch; capital letters indicate focal stress.

(8)	a.	*Jan*	*waste*	*zich.*
		Jan	washed	REFLEVIVE
		'Jan washed/got washed.' (as one does in the morning)		
	b.	*Jan*	*waste*	*zichself/*ZICH.*
		Jan	washed	himself/REFLEXIVE
		'Jan washed himSELF.' (as opposed to washing other people)		

The difference between the Dutch and the other European systems, on the one side, and the ones exemplified in (7), on the other, lies in the fact that Dutch *zich* (or Italian *si*) is an element with a clearly pronominal morphology. The verbal reflexive markers in (7) do not have this property. Still, a lot speaks in favor of treating *zich* as developing towards a non-referential reflexivizer, and the same can be said about the reflexive clitic pronouns of other European languages.[3]

The strategies to express reflexivity that were surveyed in this section were classified as pronominal, as verbal, or as somewhere in between. The verb-centered reflexivization mechanism as introduced in 3.1 matches well with affixal reflexive markers as in (7). The pronoun-and-VP-centered reflexivization mechanism matches well with English reflexive markers like *x-self*. And the clitic reflexive markers with a pronominal morphology of many European languages may be on their way from shifting from the pronoun-and-VP-centered mechanism to the verb-centered mechanism. (I will leave it open here whether there may be a semantic middleground in this domain corresponding to the morphological middleground that we are describing here.) In a nutshell, the constructional array of reflexivization patterns in the languages of the world justifies the assumption of the two major semantic mechanisms that have been proposed in the semantic literature.

5 Conclusions

The tutorial aimed at showing that major modelling options for variable binding, exemplified for the domain of reflexivity, find a neat counterpart in the major reflexivization stratregies that natural languages employ.

The slides of the tutorial can be accessed through the following link: http://www.illc.uva.nl/Tbilisi/Tbilisi2011/uploaded_files/mediaitem/kutaisi-hole.pdf

Acknowledgments. [The author]$_i$ would like to thank Sebastian Löbner for helpful comments. Mistakes remain his$_i$. The financial support from the *Deutsche Forschunsgemeinschaft* is gratefully acknowledged (Ho 2557/3-1).

[3] Cf. Reinhart and Reuland's (1993) influential account of the system of reflexization in Dutch and other European languages. Their analysis constitutes one way of spelling out the position of European reflexive pronouns as somewhere in between a verbal and a full pronominal reflexivization strategy. Another important reference for the typology of reflexive pronouns is Faltz (1985).

References

1. Büring, D.: Binding Theory. Cambridge University Press, Cambridge (2005)
2. Faltz, L.M.: Reflexivization: A Study in Universal Syntax. Garland, New York (1985)
3. Gast, V., Hole, D., Siemund, P., Töpper, S.: Typological Database of Intensifiers and Reflexives. Version 2.0 (2007), http://www.tdir.org
4. Heim, I., Kratzer, A.: Semantics in Generative Grammar. Blackwell, Oxford (1998)
5. Hole, D.: Dativ, Bindung und Diathese, habilitation thesis. Humboldt-Universität, Berlin (2008)
6. Jacobson, P.: Towards a variable-free semantics. Linguistics and Philosophy 22, 117–184 (1999)
7. Keenan, E.L.: On semantics and the Binding Theory. In: Hawkins, J. (ed.) Explaining Language Universals, pp. 105–144. Blackwell, Oxford (1988)
8. König, E., Vezzosi, L.: The role of predicate meaning in the development of reflexivity. In: Bisang, W., Himmelmann, N., Wiemer, B. (eds.) What Makes Grammaticalization, pp. 213–244. Benjamins, Amsterdam (2004)
9. Kratzer, A.: Severing the external argument from its verb. In: Rooryck, J., Zaring, L. (eds.) Phrase Structure and the Lexicon, pp. 109–137. Kluwer, Dordrecht (1996)
10. Kripke, S.A.: Naming and Necessity. In: Davidson, D., Harman, G. (eds.) Semantics of Natural Language, pp. 253–355. Reidel, Dordrecht (1972)
11. Reinhart, T., Reuland, E.: Reflexivity. Linguistic Inquiry 24, 657–720 (1993)
12. Ross, J.R.: Constraints on variables in syntax. PhD dissertation. MIT, Cambridge (1967)

Łukasiewicz Logic: An Introduction

Vincenzo Marra

Dipartimento di Matematica "*Federigo Enriques*"
Università degli Studi di Milano, Italy
vincenzo.marra@unimi.it

Abstract. This is a summary of the contents of a tutorial in logic offered at the conference. The tutorial provided a very gentle introduction to the elementary aspects of *Łukasiewicz* (*infinite-valued propositional*) *logic*.

1 Introduction

This tutorial offered an introduction to the elementary aspects of *Łukasiewicz* (*infinite-valued propositional*) *logic*, a non-classical system going back to the 1920's; *cf.* the early survey [5, §3], and its annotated English translation in [8, pp. 38–59]. Thanks to almost a century of hindsight, it is by now apparent that Łukasiewicz's terse formal system relates strongly to several fields of mathematics. While none of these connections were discussed in this basic introduction, I hope that the audience nonetheless caught a few glimpses of the mathematical treasures that await those willing to go deeper.

2 Syntax, Semantics, and Completeness

• Review of classical propositional logic.
• The syntax of Łukasiewicz logic $\mathscr{L}$. Propositional variables $\{X_1, X_2, \ldots\}$, and the set of well-formed formulæ Form. The basic connectives: $\to$ (implication), $\neg$ (negation), $\bot$ (*falsum*). Derived connectives. See Table 1.
• The semantics of Łukasiewicz logic. Assignments of truth values $w\colon \text{Form} \to [0,1] \subseteq \mathbb{R}$, or *evaluations*. See Table 2. *Tautologies* are defined as those formulæ that evaluate to 1 under every evaluation.
• The axiom schemata of Łukasiewicz logic. For each $\alpha, \beta \in \text{Form}$:

(A0) $\bot \to \alpha$ (*Ex falso quodlibet.*)
(A1) $\alpha \to (\beta \to \alpha)$ (*A fortiori.*)
(A2) $(\alpha \to \beta) \to ((\beta \to \gamma) \to (\alpha \to \gamma))$ (Implication is transitive.)
(A3) $((\alpha \to \beta) \to \beta) \to ((\beta \to \alpha) \to \alpha)$ (Disjunction is commutative.)
(A4) $(\neg\alpha \to \neg\beta) \to (\beta \to \alpha)$ (Contraposition.)

• *Provable formulæ* are defined as those formulæ that can be deduced from (A0–A4) via *modus ponens*, exactly as in classical logic. Recall that *modus ponens* is the inference rule:

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 12–16, 2013.

Table 1. Connectives in Łukasiewicz logic

Notation	**Definition**	**Name**
$\bot$	–	*Falsum*
$\top$	$\neg\bot$	*Verum*
$\neg\alpha$	–	Negation
$\alpha \to \beta$	–	Implication
$\alpha \vee \beta$	$(\alpha \to \beta) \to \beta$	(Lattice) Disjunction
$\alpha \wedge \beta$	$\neg(\neg\alpha \vee \neg\beta)$	(Lattice) Conjunction
$\alpha \leftrightarrow \beta$	$(\alpha \to \beta) \wedge (\beta \to \alpha)$	Biconditional
$\alpha \oplus \beta$	$\neg\alpha \to \beta$	Strong disjunction
$\alpha \odot \beta$	$\neg(\alpha \to \neg\beta)$	Strong conjunction
$\alpha \ominus \beta$	$\neg(\alpha \to \beta)$	But not, or Difference

Table 2. Formal semantics of connectives in Łukasiewicz logic

Notation	**Formal semantics**
$\bot$	$w(\bot) = 0$
$\top$	$w(\top) = 1$
$\neg\alpha$	$w(\neg\alpha) = 1 - w(\alpha)$
$\alpha \to \beta$	$w(\alpha \to \beta) = \min\{1, 1 - (w(\alpha) - w(\beta))\}$
$\alpha \vee \beta$	$w(\alpha \vee \beta) = \max\{w(\alpha), w(\beta)\}$
$\alpha \wedge \beta$	$w(\alpha \wedge \beta) = \min\{w(\alpha), w(\beta)\}$
$\alpha \leftrightarrow \beta$	$w(\alpha \leftrightarrow \beta) = 1 - \lvert w(\alpha) - w(\beta)\rvert$
$\alpha \oplus \beta$	$w(\alpha \oplus \beta) = \min\{1, w(\alpha) + w(\beta)\}$
$\alpha \odot \beta$	$w(\alpha \odot \beta) = \max\{0, w(\alpha) + w(\beta) - 1\}$
$\alpha \ominus \beta$	$w(\alpha \ominus \beta) = \max\{0, w(\alpha) - w(\beta)\}$

$$\frac{\alpha \quad \alpha \to \beta}{\beta}$$

for $\alpha, \beta \in$ FORM.

• **The soundness and completeness theorem.** *For every $\alpha \in$ FORM, α is provable if, and only if, it is a tautology.*

• **The computational complexity of $\mathscr{L}$.** *There is an algorithm that, on input $\alpha \in$ FORM, decides whether α is provable or not. Further, this decision problem is co-*NP*-complete.*

• By restricting all considerations to the set FORM_n of formulæ that only use n propositional variables $X_1, \ldots, X_n$, one obtains *Łukasiewicz logic over n variables.* We used this in some examples and theorems (see below).

3 Maximally Consistent Theories, and the Problem of Artificial Precision in Theories of Vagueness

• A formula α is *provable from* S, where $S \subseteq \text{Form}$, if it can be deduced from (A0–A4) and from S via *modus ponens*. In symbols, $S \vdash \alpha$. Thus, $\emptyset \vdash \alpha$ (or more simply $\vdash \alpha$) means that α is provable.
• A subset $S \subseteq \text{Form}$ is *consistent* if $S \nvdash \bot$, and *inconsistent* otherwise.
• A formula α is a *semantic consequence of* S, where $S \subseteq \text{Form}$, if every evaluation $w\colon \text{Form} \to [0,1]$ such that $w(S) = \{1\}$ satisfies $w(\alpha) = 1$. In symbols, $S \vDash \alpha$. Thus, $\emptyset \vDash \alpha$ (or more simply $\vDash \alpha$) means that α is a tautology.
• A *theory* in $\mathscr{L}$ is a set $\Theta \subseteq \text{Form}$ of formulæ that is *deductively closed*: whenever $\alpha \in \text{Form}$ is such that $\Theta \vdash \alpha$, then $\alpha \in \Theta$.
• A theory Θ is *maximally consistent* if it is consistent, and no proper superset $S \supsetneq \Theta$ of formalæ $S \subseteq \text{Form}$ is consistent.
• Classical logic is not just complete, but even *strongly complete*: for any subset S of formulæ, a formula is provable from S if, and only if, it is a semantic consequence of S. This theorem does not generalise to $\mathscr{L}$.
• **Failure of general strong completeness.** *There exist a formula* $\alpha \in \text{Form}$ *and a theory* $\Theta \subseteq \text{Form}$*, the latter not finitely axiomatisable, such that* $\Theta \vDash \alpha$ *but* $\Theta \nvdash \alpha$*.*
• **Strong completeness for finitely axiomatisable theories.** *For every* $\alpha \in \text{Form}$*, and every* finite $F \subseteq \text{Form}$*,* $F \vdash \alpha$ *if, and only if,* $F \vDash \alpha$*.*
• **Strong completeness for maximally consistent theories.** *For every* $\alpha \in \text{Form}$*, and every maximally consistent theory* $\Theta \subseteq \text{Form}$*,* $\Theta \vdash \alpha$ *if, and only if,* $\Theta \vDash \alpha$*.*
• Classical and intuitionistic logics both enjoy the *deduction theorem*: if α, β are any two formulæ, $\{\alpha\} \vdash \beta$ if, and only if, $\vdash \alpha \to \beta$. The deduction theorem links the notion of *provability* with the *semantics* of the implication connective. This state of affairs does not generalise to $\mathscr{L}$ in the most obvious manner. (And, incidentally, this circumstance ought to suggest caution to those who advance informal interpretations of the meaning of $\to$ in Łukasiewicz logic.)
• **Failure of the deduction theorem.** *There are formulæ* $\alpha, \beta \in \text{Form}$ *such that* $\{\alpha\} \vdash \beta$*, but* $\nvdash \alpha \to \beta$*.*
• **Weak deduction theorem for $\mathscr{L}$.** *For any formulæ* $\alpha, \beta \in \text{Form}$ *such that* $\{\alpha\} \vdash \beta$*, there is an integer* $n \geqslant 1$ *such that* $\vdash \underbrace{\alpha \odot \cdots \odot \alpha}_{n \text{ times}} \to \beta$*.*
• **The Łukasiewicz axioms characterise the real numbers.** *There is a natural bijection between maximally consistent theories* $\Theta \subseteq \text{Form}_1$*, and numbers in* $[0,1] \subseteq \mathbb{R}$*. Specifically, the bijection is given by* $r \in [0,1] \longmapsto \Theta_r$*, where* $\Theta_r := \{\alpha \in \text{Form}_1 \mid w_r(\alpha) = 1\}$*, and* $w_r\colon \text{Form}_1 \to [0,1]$ *is the unique evaluation of* Form_1 *such that* $w_r(X_1) = r$*.* (Here X_1 is a propositional variable, and Form_1 is the set of formulæ over the single variable X_1, as explained in Section 2.) We thereby see that the innocent-looking Łukasiewicz axioms (A0–A4) characterise the real numbers.
• **Artificial precision.** It has been argued by many in the philosophical literature on theories of vagueness that regarding $r \in [0,1]$ as a "degree of truth" of a

vague proposition — *e.g.* the proposition "x is tall", where x is some individual — replaces the vagueness of the monadic predicate $\mathsf{Tall}(\cdot)$ with the most implausible precision. For, what could determine that $\mathsf{Tall}(x)$ is to be assigned degree of truth $r := \frac{\sqrt{2}}{2}$, say, rather than $r' := \frac{\sqrt{2}}{2} + 10^{-1000}$? The preceding theorem shows exactly what it is that would determine such a difference: the distinct maximally consistent theories Θ_r and $\Theta_{r'}$, respectively, each of which encodes complete consistent knowledge about x's tallness. We have briefly discussed the import of this mathematical fact for theories of vagueness.

4 Functional Completeness

• A function $f\colon \{0,1\}^n \to \{0,1\}$ is called a *Boolean function.* Any such function is definable by a formula (over n variables) in classical logic. For example, $f\colon \{0,1\} \to \{0,1\}$ with $f(0) = 1$ and $f(1) = 0$ is definable by the formula $\alpha_f(X_1) := \neg X_1$. The fact that Boolean functions are precisely the definable functions is known as the *functional completeness theorem* for classical logic.
• In $\mathscr{L}$, a formula $\alpha(X_1, \ldots, X_n) \in \textsc{Form}_n$ defines a function $f_\alpha\colon [0,1]^n \to [0,1]$. Indeed, for $x := (x_1, \ldots, x_n) \in [0,1]^n$, one sets $f_\alpha(x) := w_x(\alpha)$, where $w_x\colon \textsc{Form}_n \to [0,1]$ is the unique evaluation of $\textsc{Form}_n$ such that $w(X_1) = x_1, \ldots, w(X_n) = x_n$. What functions are obtained in this manner?
• A function $f\colon [0,1]^n \to [0,1]$ is *piecewise linear* if it is continuous, and there is a finite set L of affine linear functions $\mathbb{R}^n \to \mathbb{R}$ such that, for each $x \in [0,1]^n$, $f(x) = l_x(x)$ for some $l_x \in L$. Further, such an f is a *$\mathbb{Z}$-map* (also called a *McNaughton function*) if each affine linear function in L can be chosen to have integer coefficients.
• **Functional completeness theorem for $\mathscr{L}$.** *A function $f\colon [0,1]^n \to [0,1]$ is definable by an n-variable formula of Łukasiewicz logic if, and only if, f is a $\mathbb{Z}$-map.*

5 Bibliography

Here I merely point to a few standard textbooks, which in turn provide many further references to specific topics.

• The standard introduction to Łukasiewicz logic is [1]. Advanced aspects are treated in [6].
• Łukasiewicz logic is part of a hierarchy of many-valued logics systematised by Petr Hájek, see [4]. A further useful reference is the recent handbook [2].
• Hájek's hierarchy provides instances of *substructural logics*, that is, logics that renounce one ore more of the *structural rules* — such as weakening or contraction — of Gentzen's sequent calculus LK for classical logic. For further information see [3].
• A few hints to theories of vagueness were given. This is a rather specialised branch of analytic philosophy. Interested readers can start from [9], [7].

References

1. Cignoli, R.L.O., D'Ottaviano, I.M.L., Mundici, D.: Algebraic foundations of many-valued reasoning. Trends in Logic—Studia Logica Library, vol. 7. Kluwer Academic Publishers, Dordrecht (2000)
2. Cintula, P., Hájek, P., Noguera, C. (eds.): Handbook of Mathematical Fuzzy Logic. Studies in Logic, vol. 1 & 2, 37 & 38. College Publications (2011)
3. Galatos, N., Jipsen, P., Kowalski, T., Ono, H.: Residuated lattices: an algebraic glimpse at substructural logics. Studies in Logic and the Foundations of Mathematics, vol. 151. Elsevier B.V., Amsterdam (2007)
4. Hájek, P.: Metamathematics of fuzzy logic. Trends in Logic—Studia Logica Library, vol. 4. Kluwer Academic Publishers, Dordrecht (1998)
5. Łukasiewicz, J., Tarski, A.: Untersuchngen über den Aussagenkalkül. C. R. Soc. Sc. Varsovie 23, 30–50 (1930)
6. Mundici, D.: Advanced Łukasiewicz Calculus and MV-algebras. Trends in Logic—Studia Logica Library, vol. 35. Springer, New York (2011)
7. Smith, N.J.J.: Vagueness and Degrees of Truth. OUP, Oxford (2008)
8. Tarski, A.: Logic, semantics, metamathematics. Papers from 1923 to 1938. Oxford at the Clarendon Press (1956); Translated by Woodger, J.H.
9. Williamson, T.: Vagueness. The Problems of Philosophy Series. Taylor & Francis Group (1996)

The Information Structure and Typological Peculiarities of the Georgian Passive Constructions*

Rusudan Asatiani**

Abstract. Functionally defined Passive Constructions are characterized as the conversive ones of corresponding active constructions, where a patient is promoted to the subject position, and an agent is demoted and transformed into a prepositional phrase. Georgian passive constructions do not always show such a conversion and actually express a variety of semantics: deponents, reflexives, reciprocals, potentials, etc. The peculiarities of Georgian passive define the restrictions of their usage in the processes of information structuring, where patient foregrounding implies certain morphosyntactic changes characteristic for conversive-passive constructions. The analysis of the Georgian sentence information structure provides a strong argument for interpreting Georgian passive as a grammatical category mostly governed by cognitive-semantic, and not simply by syntactic, features. This paper suggests a cognitive productive model and some semantic features that define the choice of either the passive or active formal models for grammatical representations of verbs showing so-called medial semantics.

Keywords: information structure, active-passive opposition, medial semantics, continuum of active-passive opposition, cognitive interpretations.

1 Introduction: Defining Objectives

Within the theory of functional grammar, passive constructions are considered conversive of corresponding active constructions, where a Patient is promoted to the subject position along the 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. Thus, the passive is considered a syntactic category. However, many languages present morphologically marked verb forms in such conversive (res. passive) constructions and, consequently, it is possible to speak about the morphosyntactic category of the passive voice.

In Georgian, there is an obvious formal opposition between the active – first and foremost, the transitive – and the passive – the active's conversive – verb forms represented by some morphological markers (see below 1-3) and syntactic features (see below 4), as follows.

* This paper is a product of the project "Typology of information structure," which is part of the SFB 632 "Information structure" at the University of Potsdam and Humboldt University Berlin, sponsored by the DFG.

** Tbilisi.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 17–34, 2013.

1. In the present tense forms the suffix representing S.3.SG in active verbs is *-s*, while in the passive ones it is *-a.*

2. In the present and derivationally connected tense forms – so-called I series forms – passive verbs have only one suffix *-eb-*, the so-called thematic marker, while active ones can have various thematic markers: *-eb-, -ob-, -op-, -av-, -am-, -i.*

3. As opposed to the active verb forms, the passive ones can be represented by the vowel prefixes *i-, e-* or the suffix *-d*;

4. The subject in passive constructions always stands in Nominative, while in active constructions – most importantly, in the constructions built by transitive verbs – it can be represented either by the Nominative (in Present) or Ergative (in Aorist) or Dative (in Perfect) case.

The following are Present tense examples:

(1-a)	*k'ac-i*	*xat'-av-s*	*surat-s*	
	man-NOM	paint-THM-ACT.S.3.SG	picture-DAT	
	'The man paints the picture.'			
(1-b)	*surat-i*	*i-xat'-eb-a*	*k'ac-is*	*mier*
	picture-NOM	PASS-paint-THM-PASS.S.3.SG	man-GEN	by
	'The picture is painted by the man.'			

(2-a)	*k'ac-i*	*a-šen-eb-s*	*saxl-s*	
	man-NOM	NV-build-THM-ACT.S.3.S	house-DAT	
	'The man builds the house for himself.'			
(2-b)	*saxl-i*	*šen-d-eb-a*	*k'ac-is*	*mier*
	house-NOM	buld-PASS-THM-PASS.S.3.SG	man-GEN	by
	'The house is built by the man for himself.'			

(3-a)	*k'ac-i*	*u-gzavn-i-s*	*c'eril-s*	*kal-s*	
	man-NOM	OV-send-THM-ACT.S.3.S	letter-DAT	woman-DAT	
	'The man sends the letter to the woman.'				
(3-b)	*c'eril-i*	*e-gzavn-eb-a*	*kal-s*	*k'ac-is*	*mier*
	picture-NOM	PASS-send-THM-PASS.S.3.SG	woman-DAT	man-GEN	by
	'The letter is sent by the man to the woman.'				

The following are Past tense examples:

(4-a)	*k'ac-ma*	*da-xat'-a*	*surat-i*	
	man-ERG	PV-paint-AOR.S.3.S	picture-NOM	
	'The man painted the picture.'			
(4-b)	*surat-i*	*da-i-xat'-a*	*k'ac-is*	*mier*
	picture-NOM	PV-PASS-paint-AOR.S.3.SG	man-GEN	by
	'The picture was painted by the man.'			

(5-a)	*k'ac-ma*	*a-a-šen-a*	*saxl-i*	
	man-ERG	PV-NV-build-AOR.S.3.S	house-NOM	
	'The man built the house for himself.'			
(5-b)	*saxl-i*	*a-šen-d-a*	*k'ac-is*	*mier*
	house-NOM	PV-build-PASS-AOR.S.3.SG	man-GEN	by
	'The house was built by the man for himself.'			

(6-a) *k'ac-ma ga-u-gzavn-a c'eril-i kal-s*
man-ERG PV-OV-send-AOR.S.3.S letter-DAT woman-DAT
'The man sends the letter to the woman.'

(6-b) *c'eril-i ga-e-gzavn-a kal-s k'ac-is mier*
picture-NOM PV-PASS-send-AOR.S.3.SG woman-DAT man-GEN by
'The letter is sent by the man to the woman.'

The following are Present Perfect tense examples:

(7-a) *k'ac-s da-u-xat'-av-s surat-i*
man-DAT(PV-SINV.3.CV-paint-THM):PRF-OINV.3.S[SINV.3.SG]picture-NOM
'The man has painted the picture.'

(7-b) *surat-i da-xat'-ul-a k'ac-is mier*
picture-NOM PV-paint-PRT-be.PRS.S.3.SG man-GEN by
'The picture has been painted by the man.'

(8-a) *k'ac-s a-u-šen-eb-i-a saxl-i*
man-DAT (PV-SINV.3.CV-build-THM-0):PRF-OINV.3.S[SINV.3.SG] house-NOM
'The man has built the house for himself.'

(8-b) *saxl-i a-šen-eb-ul-a k'ac-is mier*
house-NOM PV-build-THM-PRT-be.PRS.S.3.SG man-GEN by
'The house has been built by the man for himself.'

(9-a) *k'ac-s ga-u-gzavn-i-a c'eril-i kal-is-tvis*
man-DAT (PV-SINV.3.CV-send-0):PRF-OINV.3.S[SINV.3.SG] letter-DAT woman-GEN-for
'The man has sent the letter to the woman.'

(9-b) *c'eril-i ga-h-gzavn-od-a kal-s k'ac-is mier*
picture-NOM (PV-IO.3-send-IMP):PRF-S.3.SG woman-DAT man-GEN by
'The letter has been sent by the man to the woman.'

All the features taken together clearly distinguish an opposition between active and passive verb forms, although none of them can be regarded as a simple marker for the passive voice as far as they do not exist only in passive constructions. The following examples illustrate this.

1. The main function of *-s, -a* suffixes is to mark S.3.SG. Once this function is identified, examples of it can be found in various cases. For example, *-s* expresses S.3.SG in a subjunctive mood of passive verb forms and also in some static verbs. For instance, *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'), *zi*(sit.PRS)-*s*(S.3.SG) ('S/he is sitting'), etc., while *-a* can be a marker of active verb's S.3.SG in past tenses (see examples (4-a), (5-a), (6-a), (8-a), (9-a) above).

2. The main function of *-eb-* is to mark out dynamic verb forms. In expressing this function, *-eb-* also occurs with some active verbs (see example (2-a), (8-a), etc.).

3. The vowel prefixes are also polyfunctional: in general, they represent derivational changes of verb valency – either the increase or decrease of verb arguments syntactically linked with a verb. For instance, *-i-* expresses such categories

as the subjunctive version (see examples (2-a), (5-a)), the reflexive version (e.g., *i*(SV)-*ban*(wash)-*s*(S.3.SG) ('S/he washes her/himself.'), *i*(SV)-*p'ars*(shave)-*av*(THM)-*s*(S.3.SG) ('He shaves himself.')), potentials (see examples (12) below), deponents (see examples (10) below) and also has an additional function to form the future tense of some intransitive, active verbs (e.g.: [*i*(CV)-*cxovr*(live)-*eb*(THM)]:FUT-*s*(S.3.SG) ('S/he will live'), [*i*(CV)-*mɣer*(sing)-*eb*(THM)]:FUT-*s*(S.3.SG) ('S/he will sing')).[1]

4. The Nominative is the case that is characteristic for the subjects of some intransitive, non-active, non-covnersive passive verbs, and also static verbs in past forms (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'); *is*(S/he.NOM.SG) *gd*(lay strewn)-*i*(PRS)-*a*(S.3.SG) : *is*(S/he.NOM.SG) *e*(CV)-*gd*(lay strewn)-*o*(PST.S.3.SG)). The Ergative (or the Dative) case can also be the subject marker for 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.NOM.SG) *i*(CV)-*pikr*(think)-*a*(AOR.S.3.SG) ('S/he thought'); *mas* [*u*(SINV.3.CV)-*cek'*(dance)-*v*(THM)-*i*(PST)]:PRF-*a*(OINV.3[S.INV.3.SG]) ('S/he has danced'); [*u*(CV)-*muš*(work)-*av*(THM)-*i*(PST)]:PRF-*a*(OINV.3[S.INV.3.SG]) ('S/he has worked')).

Georgian morphosyntactically distinguished passive constructions do not always show the conversion of corresponding active ones, and they can in fact express a variety of semantics. The following examples illustrate this point.

(10) Active semantics: *e*(PASS)-*kač*(tug)-*eb*(THM)-*a*(S.3.SG) 'S/he tugs hard at smth./smb.', *a*(CV)-*c'v*(push)-*eb*(THM)-*a*(S.3.SG) 'S/he pushes smth./smb.' *e*(PASS)-*laparak'*(speak)-eb(THM)-*a*(S.3.SG).

(11) Dynamic actions: *dg*(stand)-*eb*(THM)-*a*(S.3.SG) 'S/he is standing up', *tvr*(get drunk)-*eb*(THM)-*a*(S.3.SG) 'S/he gets drunk', *šr*(dry)-*eb*(THM)-*a*(S.3.SG) 'S/he dries', *tb*(warm)-*eb*(THM)-*a*(S.3.SG) 'S/he gets warm'.

(12) Potentials: *i*(CV)-*č'm*(eat)-*eb*(THM)-*a*(S.3.SG) 'It is edible', *i*(CV)-*sm*(drink)-*eb*(THM)-*a*(S.3.SG) 'It is drinkable', *i*(CV)-*k'itx*(read)-*eb*(THM)-*a*(S.3.SG) 'It can be read'.

(13) Reciprocals: *e*(CV)-*tamaš*(play)-*eb*(THM)-*a*(S.3.SG) 'S/he plays with smb.=They play together', *e*(CV)-*cek'v*(dance)-*eb*(THM)-*a*(S.3.SG) 'S/he dances with smb.=They dance together'.

These verbs do not have an active counterpart and can produce corresponding active semantics by special derivational models. The initial forms for them are semantically non-active forms, while for the conversive-passive forms, on the contrary, the initial forms are the active ones.[2]

Thus, we have two different formal models defined by some morphological and syntactic features. Let us define them as active-passive formal models as follows.

[1] For the polyfunctionality of *i*- prefix see ASATIANI 2001.

[2] For the structural models of the so-called passive forms and their semantic interpretations see IVANISHVILI & SOSELIA 1999.

Table 1.

	‘Active Model’	‘Passive Model’
S.3.SG suffix (in present)	*-s*	*-a*
Thematic marker (in I-series tense forms)	*-eb-*	*-eb-, -ob-, -op-, -av-, -am-, -i-, -Ø-*;
Special markers	–	*i-, e-, -d, -Ø-*
Subject case	NOM (in present)/ ERG (in Aorist)/ DAT (in Perfect)	NOM

Primarily, the active model serves as a formal representation of transitive verbs with an affected object – let us call them prototypical actives –, while the passive model serves as formal representation of functional, conversive-passive, which we can call prototypical passives. As far as these models can also be used for the verbs that do not represent a prototypical semantics – let us call such verbs Medial ones – (see examples (10), (11), (12), (13) above), it is obvious that the question of what is the real function of the morphosyntactically differentiated models still needs to be answered. These formal models can not be interpreted simply, and their semantic and/or functional analysis requires further investigation.

2 Theoretical Approaches: Information Structures of a Sentence

According to one of the theoretical approaches implemented in contemporary linguistics, one of the devices for active-passive functional differences can be explained by a variety of information structures.

In general, linguistic structuring of information – its packaging – proceeds through the oppositions, where one part of the information stands out against the background of the other part. From the communicational, pragmatic point of view, this information is highlighted, important, and represents the foregrounding of a certain part of this information. Any kind of *foregrounding* – including *highlighting*, *logical emphasis*, *promotion*, *standing out as primary, important* – can be regarded as a *single, common phenomenon*, which represents the main strategy of structuring linguistic expressions. From this perspective, Topic, Focus, Subject, Theme, Point of View, and others are the same to the extent that they represent various forms of foregrounding. Foregrounding, according to such a broad interpretation, can be realized on various linguistic levels, and in this case we can distinguish Conceptual, Functional, Discourse and Pragmatic devices (ASATIANI 2007).

2.1 Conceptual Foregrounding

In the process of the linguistic structuring of extra-linguistic situations some languages conventionally conceptualize either an Agent or Patient as the central part of the information. As a result, one of two constructions arises, the Nominative – which shows the Agent's foregrounding – or the Ergative – which shows the Patient's foregrounding. The first construction formally foregrounds *who is acting*, while the second foregrounds *what is done*. From the grammatical point of view, conceptual foregrounding is represented by the unmarked, Nominative case: In the nominative languages the Agent always stands in nominative, while in the ergative languages, it is the Patient – not the Agent – who appears in nominative.

Georgian shows split ergativity (DIXON 1979; HARRIS 1981). Transitive verbs in Present Tense form build the Nominative constructions, in which conceptual foregrounding means to put an Agent in the central position. The following example illustrate this.

(10)	***monadire***	*k'l-av-s*	*irem-s.*
	hunter.NOM	kill-THM-PRS.ACT.S.3.SG	deer-DAT
	'The hunter kills the deer.'		

The Aorist forms build the Ergative construction, in which conceptual foregrounding puts a Patient in the central position. For example:

(11)	*monadire-m*	*mo-k'l-a*	***irem-i.***
	hunter-ERG	PV-kill-AOR.S.3.SG	**deer-NOM**
	'The hunter killed the deer.'		

The Perfect forms build the Dative – so-called inversive – construction, in which a Patient is foregrounded to a central position as well (SHANIDZE 1973). For example:

(12)	*monadire-s*	*mo-u-k'l-av-s*	***irem-i.***
	hunter-DAT	(PV-CV-kill-THM):PRF-S.3.SG	**deer-NOM**
	'The hunter has killed the deer.'		

2.2 Functional Foregrounding

The Patient's foregrounding in the nominative languages, where the Agent is a conceptually highlighted element, can further be achieved by the changes of functional roles, and, as a result, conversive-passive constructions arise. In the passive construction, the Patient is functionally promoted, and it is defined as a Subject. The term *Subject* actually denotes the foregrounding of a certain part of information – *whom* or *what* the information concerns. The Active construction shows the Agent's foregrounding – in other words, the Agent is the Subject and, consequently, stands in nominative – while the Passive construction shows the Patient's foregrounding – the Patient is the Subject and, consequently, stands in nominative. Thus, a syntactic opposition between active and passive constructions can be triggered by different models of information structuring. In discourse, this is one of the main sources for the production of passive constructions.

In Georgian, Passivization is a regular way for the Patient's foregrounding in active-transitive verbs in Present Tense forms. For example:

(13) Active: *monadire* *k'l-av-s* *irem-s.*
hunter.NOM kill-THM-S.3 deer-DAT
'The hunter kills the deer.'

(14) Passive: *irem-i* *i-k'vl-eb-a* *monadir-is mier.*
deer-NOM PV-PASS-kill-THM-PRS.PASS.S.3.SG hunter-GEN by
'The deer is killed by the hunter.'

The Passive construction is not always clearly distinguishable in a formal sense by the verb forms in Aorist. This occurs when the verb forms showing subjective version of active transitive verb represented by *i- prefix* and morphologically represented *i*-passive forms can not be distinguishable in aorist. This is true for all verbs having *i*-conversive-passive. Only syntactic features – alignment of arguments – make it possible to differentiate an active-passive opposition. The following examples illustrate this point.

(15) Active: *monadire-m* *mo-i-k'l-a* *irem-i* *tav-is-tvis.*
hunter-ERG PV-SV-kill-AOR.S.3.SG **deer-NOM** self-GEN-for
'The hunter killed the deer for himself.'

(16) Passive: *irem-i* *mo-i-k'l-a* *monadir-is* *mier.*
deer-NOM PV-PASS-kill-AOR.S.3.SG hunter-GEN by
'The deer was killed by the hunter.'

Conceptually, this fact makes sense. In ergative constructions the Patient is already defined as conceptually foregrounded. From the informational point of view, its further functional foregrounding seems to be redundant.

Finally, in the Perfect Tense Forms, the formal opposition between the Passive and the Active constructions yields an absolutely different picture. The active constructions trigger the Dative – so-called inversive – constructions, while passive ones preserve the Nominative model. Therefore, the opposition is expressed mainly by syntactic features, namely, by verb arguments in various case patterns.

Thus, it can be supposed that in Georgian, due to split-ergativity and the restrictions of passivization for some medial-active verbs, more complex processes define the choice of either active or passive models in the course of information structuring.

Another device defining either appearance or disappearance of PC in Georgian is a relatively free word order. This makes it possible to put the focused Patient in a marked – mostly sentence-initial and pre-verbal – position without any kind of functional promotion and/or demotion – respectively, passivization. As well, sentence intonation contour plays an important role.

Thus, a passive construction is not the only means to express the functional foregrounding of Patient. Consequently, the role of passive construction in the process of Patient functional foregrounding in Georgian also needs further investigations.

3 Methodology

Sentences raised in natural conversation are the most valuable for the task of identifying the main formal models of information structures. It is possible to stimulate such situations of natural converstations by means of experimental tasks specifically designed for this purpose. Our empirical data is collected on the basis of *Questionnaire on Information Structure* (henceforth QUIS), that is being developed within the Sonderforschungsbereich 632 *Information Structure* at the University of Potsdam and the Humboldt University Berlin (SKOPETEAS et al. 2006). QUIS comprises of a set of translation tasks and production experiments for primary data collection. The so-called "production experiments" contain a range of experimental settings, that introduce spontaneous expressions, for instance, picture descriptions, map tasks, some games, etc. For our purposes, the following experiments were particularly interesting.

3.1 Description of the Experimental Setup

This particular experiment explores the interrelation between a patient's animacy and an agent's visibility. The relevant constructions show patient foregrounding. With respect to animacy, it is assumed that foregrounding, in general, is more probable with the animate and less probable with the inanimate. With respect to agent visibility, it is assumed that patient foregrounding is more probable, if the agent is not identifiable, and less probable, if the agent is identifiable.

In the following experiments the assumptions are implemented as follows. All settings are equal in that the target picture shows the patient. It is introduced by a previous picture. In the target picture, four different cases are presented:

Figure 1: the patient is animate; the agent is non-identifiable;
Figure 2: the patient is animate; the agent is identifiable;
Figure 3: the patient is inanimate; the agent is identifiable;
Figure 4: the patient is inanimate; the agent is non-identifiable.

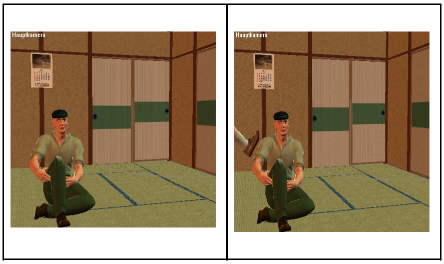


Fig. 1.

Fig. 2.

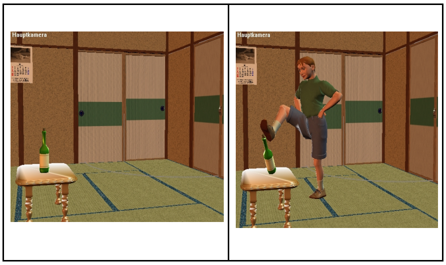


Fig. 3.

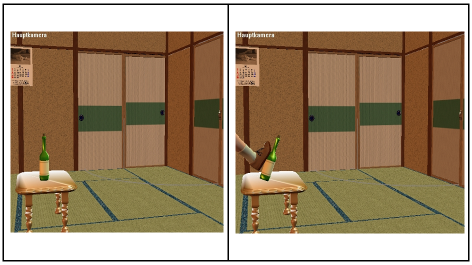


Fig. 4.

Procedure

The instructor explains to the informant:

> *You will be shown two scenes that belong together. They belong to the same story.* ***Imagine that the first scene takes place first, and the second scene some time later, say, after five minutes.*** *Please provide a short description of what is going on in each scene.*

The instructor shows the first picture to the informant and asks:

> *What is going on in this scene?*

The description may be free, and as long as the informant wishes. Most importantly, the informant has to understand the setting in the picture. The instructor shows the second picture and asks:

> *What is going on in this scene?*

As the result of such procedures, an audio-recorded database was created, consisting of 192 sentences. This semi-spontaneous data was collected during eight field sessions (16 informants, mostly Georgian students) using 12 pairs of stimuli pictures from the QUIS experiment tasks (SKOPETEAS et al. 2006).

3.2 Data Analysis

On the basis of the audio-recorded data, it becomes clear that Georgian informants prefer to produce active constructions, in which a foregrounded patient occupies an initial position in the sentence. The subject is either uncertain – represented in a verb form by S.3.PL suffixes (see example (17) below) – or the subject is represented by the indefinite pronoun *viɣac,* 'somebody' (see example (18) below) and the word order changes. An example of the above follows.

(17) *[botl-s]$_{Topic}$* *k'r-av-en* *pex-s*
bottle-DAT push-THM-ACT.PRS.S.3.PLfoot-DAT
'(They) are kicking the bottle.'

(18) *[ma-s]$_{Topic}$* *viɣac* *pex-s* *u-rt'q'-am-s.*
3.SG-DAT somebody foot-DAT [IO.3]OV-hit-THM-ACT.PRS.S.3.SG
'Somebody is hitting him with foot.'

Only one example is found in our data, where the invisible agent is represented by the active construction with the concrete noun *kali,* ‘a woman’:

(19) *[t’apa-s]$_{Topic}$* *i-ɣ-eb-s* *kal-i.*
frying-pan-DAT SV-take-THM-ACT.PRS.S.3.SG woman-NOM
‘A woman is picking up the frying-pan.’

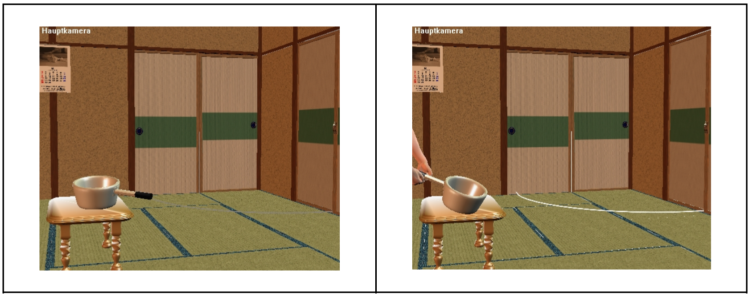


Fig. 5.

This case can be explained with reference to a general idea that women are thought of as *cooks*. According to such a cultural background – social and gender differences – in this context the invisible subject becomes identifiable.

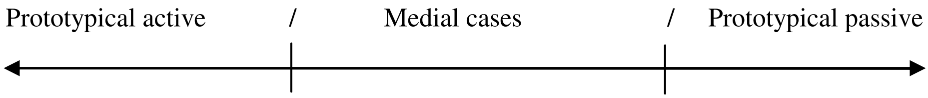
Thus, based on the conducted data analysis, we can conclude, that passive constructions in Georgian are not defined by the invisible agents and/or by the animate patients. These semantic features, which supposedly must evoke conversive-passive constructions, do not provoke the syntactic changes of functions. In the process of information structuring, when in a given situation an invisible Agent, together with the animate Patient is presented, and the passive constructions would be most appropriate, 100% of Georgian informants prefer to use active constructions with the uncertain subject represented in verb forms by either S.3.PL suffixes (48%), or by indefinite pronouns *viɣac/raɣac* ‘somebody/something’ (52%). Therefore, patient foregrounding is not expressed by passive constructions. There is no functional foregrounding, and the patient and/or agent do not change their functional characteristics. This is an additional argument to interpret Georgian passive as a grammatical category primarily governed by verb semantics – more widely, it is governed by the cognitive processes of information structuring – and not by the noun semantics defining functional changes, which are realized as a syntactic category. Consequently, it is necessary to find other features, that define the formal opposition between the active and the passive morphological distinction and different models of formal representation. We will show, that these features will be semantic and not simply syntactic.

4 Cognitive-Semantic Interpretation of the Active-Passive Morphosyntactic Oppositions

In many languages, as in Georgian, the active-passive constructions do not always express syntactically defined conversive forms. The passive formal model is also used to mark other related constructions. In general, there are languages in which the passive formal model marks reflexives and reciprocals – for instance, Russian – deponents – for instance, Latin – or middle verbs – for instance, Greek. In some languages, this model goes further and also expresses other grammatical relations. In Japanese, for example, it is the formal representation for polite constructions and, more importantly, plural forms. Consequently, attempts at new theoretical approaches have been undertaken to explain such cases. One such approach is Shibatani's interpretation (SHIBATANI 1985).

4.1 A Continuum of Active-Passive Opposition

SHIBATANI considers the active-passive opposition as a continuum, where polar dimensions fit in with the prototypical active and passive constructions, while non-polar, inter-medial cases share only some semantic-categorical features of the categories characteristic of the prototypical ones.



Languages apply various strategies for formal representations of such non-polar cases. We call them *Medial* cases. Languages either create new formal models, or choose from the existing ones a model that conventionally is regarded as the most appropriate and proximate according to certain semantic-categorical features. In such cases, simple taxonomic functional interpretations – the change of syntactic functions – or semantic interpretations – the definition of active-passive semantics – concerning the formal models are much more difficult to make, and sometimes even impossible.

4.2 Georgian Active-Passive Continuum

As demonstrated, the active-passive opposition in Georgian can not always be defined by distinct information structures. The patient's functional foregrounding is not always the source for a passive formal representation model (see 3.2). As for the semantic interpretation, it is absolutely clear, that there are no simple one-to-one correspondences between the active-passive semantic oppositions and the active-passive formal models. As shown in section 1, a case of active semantics does not always express itself in the active model. For instance, the passive morphosyntactic model sometimes actually expresses active semantics: *dg-eb-a* 'S/he is standing up', *ekač-eb-a* 'S/he tugs hard at smth.', *ac'v-eb-a* 'S/he pushes smth./smb.' A passive semantics is also not always expressed by the passive model. For instance, some static verbs that actually have non-active, passive semantics, have the same suffix *-s-* in present, expressing S.3.SG as the active ones: *dga-s* 'stand-S.3.SG', *c'ev-s* ' (smb.) lie-S.3.SG ', *zis* 'sit-S.3.SG', *dev-s'* (smth.) lie-S.3.SG', *ɣir-s* 'cost-S.3.SG', *c'ux-s* 'worry-S.3.SG' (see 1).

Georgian active-passive opposition could be interpreted as a continuum, in which the prototypical active corresponds to the transitive active constructions represented by the active model. As well, the prototypical passive defined by patient foregrounding corresponds to the active construction conversive form, in turn, represented by the passive model. In this case, the process of the grammaticalization of the medial forms – not prototypically active or passive verbs – can be explained by the following general cognitive tendency:

> *In the process of the formal representation of the medial forms, Georgian applies either the active or the passive formal model. The strategy of choice is defined by the specific conventionally accepted linguistic so-called 'decision' about 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 medial forms:

> *If a medial prototypically non-active and/or non-passive verb semantics tends toward an end – in other words, it is semantically* ***telic*** *– then a verb selects the passive formal model of representation. If a medial verb semantics does not tend toward an end – in other words, it is semantically* ***atelic*** *– then a verb chooses the active formal representation model.*

A formal interpretation of this fact is fairly simple:

> *If a verb with medial semantics can take just one preverbal prefix or preverbs*[3] *showing some direction of action – sometimes also creating new semantics of a verb – then the verb has 'passive form.'*

Compare, for instance, the first set of examples to the second one.

First Set:

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'

ča(PV:FUT)-*dg-eb*(THM)-*a*(S.3.SG) 'S/he will stand in'

e(CV)-*mal*(hide)-*eb*(THM)-*a*(S.3.SG) 'S/he is hiding from smth. or smb.'

---

[3] In Georgian, the so-called Preverbs are preverbal affixes that show a direction of an action and additionally form the future tense for transitive and conversive-passive verb forms as well as the perfective-imperfective aspect (see SHANIDZE 1973).

da(PV:FUT)-*e*(CV)-*mal-eb*(THM)-*a*(S.3.SG) 'S/he will hide from smth. or 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. or smb.

da(PV:FUT)-*a*(CV)-*c'v-eb*(THM)-*a*(S.3.SG) 'S/he will lie down on smth. or smb.'

Second Set:

cxovr(live)-*ob*(THM)-*s*(S.3.SG) 'S/he lives'

pikr(live)-*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 shivers'

gor(roll)-*av*(THM)-s(S.3.SG) 'S/he/it rolls'

suntk(breath)-*av*(THM)-*s*(S.3.SG) 'S/he breathes'

bč'(discuss)-*ob*(THM)-*s*(S.3.SG) 'S/he discusses'

brial(sparkle)-*eb-s*(S.3.SG) 'It sparkles'

In second set, the last medial verbs having the form similar to active-transitive ones do not attach preverbs.

4.3 Expanded Continuum

Cognitive-semantic interpretations of the active-passive continuum can be broadened to encompass all spectrums of the medial verb forms, including the verbs expressing static states.[4] If such medial verb forms are taken into account, the process of information structuring can be reinterpreted as a hierarchically organized process. In this process, another opposition of categories – namely, "dynamic/static" – takes a distinct role. Hence the following rule is operating.

If medial verb forms express static events, then the verb in present tense has an auxiliary conjugation.

In other words, the Georgian language creates a new model of formal representation with auxiliary conjugation in present tense only for static verbs. This model is different from either active or passive. The following examples illustrate this point.

(20) *me*(1.SG) *v*(S.1)-*dga*(stand)-*v*(S.1)-*ar*(be.SG)
šen(2.SG) (S.2)*dga*(stand)-*x*(S.2)-*ar*(be.SG)
is(3.SG) *dga*(stand)-*s*(S.3.SG)
('I am/You are/ He is standing.')

[4] According to Georgian grammatical tradition, these are referred to as static passive and medio-passive states (SHANIDZE 1973).

(21) *me*(1.SG) *v*(S.1)-*c'ev*(lie)-*v*(S.1)-*ar*(be.SG);
šen(2.SG) (S.2)*c'ev*(lie)-*x*(S.2)-*ar*(be.SG);
is(3.SG) *c'ev*(lie)-*s*(S.3.SG)
('I am/You are/ He is lying.')

In some instances, such medial verbs correspond to the (20) type conjugation. In the Georgian grammatical tradition, they are reffered to as static passives (SHANIDZE 1973). For example:

> *gdi-a* 'lie. strewn/thrown about-S.3.SG', *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'.

In other instances, such medial verbs trigger the (21) type conjugation, reffered to as medio-passives (SHANIDZE 1973). For example:

> *dga-s* 'stand-S.3.SG', *c'ev-s* '(smb).lie'-S.3.SG', *zi-s* 'sit-S.3.SG', *dev-s*' (smth).lie-S.3.SG', *ɣir-s* 'cost-S.3.SG', *c'ux-s* 'worry-S.3.SG'.

According to these examples, the static verbs having auxiliary conjugation fall into two subgroups. One group has the S.3.SG suffix *-s*, also characteristic for the prototypically active verbs in present. The second group shows S.3.SG ending *-a*, which is the same as is characteristic for the prototypically passive verbs in present. The functional differences are more refined, hence the discovery of specific semantic nuances implying the opposition needs a more careful analysis. We suggest the following formal testing expression.

> *If a verb generates a correct phrase with the adverb* ***tavad*** ***'him/her/itself, personally'****, then it chooses the active model. For instance, the expressions* ***tavad dga-s*** ('S/he stands herself, personally'), ***tavad c'ev-s*** ('S/he lies herself, personally'), ***tavad c'ux-s*** ('S/he is worried herself, personally') *are correct. If such phrases are not correct, then the passive model of representation is chosen. For instance, the expressions* ****tavad gdi-a*** ('It is lying strewn about (itself)') ****tavad kidi-a*** ('It is hanging (itself)') ****tavad c'eri-a*** ('It is written (itself)') *are unnatural or improper.*

The testing phrase {tavad, (S), V} leads us to distinguish the following valuable semantics: *A state is provoked and/ or controlled by a subject's will, a subject 'acts' him/her/itself, personally*. We denote this feature by the term *Agentivity* and define the following rule.

> *A verb expressing states that can be controlled or triggered by the subject itself – a subject that is conventionally close to agentivity – selects the active model, while verbs expressing a state not controlled or triggered by the subject itself select the passive representation model.*

Therefore, summarizing our argument, we suggest a dynamic model for the choice of the voice in language production. This model reflects the cognitive-semantic grounds for formal-grammatical representations of an active-passive opposition and the medial forms.

5 The Hierarchically Organized Dynamic Model

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 models. The so-called *decision* of which model will be most appropriate for the concrete medial verb semantics is taken step by step conventionally based on the most optimal cognitive interpretations originating from some crucial semantic features.

Step 1: Prototypically active and prototypically passive relations are represented by the main formal models. These are the active – transitive verbs with an affected object showing agent foregrounding – and the passive – conversive forms of active relations showing patient foregrounding – constructions. (For the morphosyntactic features of the constructions, see section 1.)

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

> **Strategy 1.** The new model is created (NM)
> **Strategy 2.** Either active or passive models 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 decisive role – verbs expressing *Static* states are marked according to strategy 1, and the new model 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 model of representation (strategy 2).

Step 3: For the *Dynamic* subgroup, further choices are defined by the semantic feature *Telicity*:

> *Telic medial verbs choose the passive formal model of representation, while atelic medial verbs – the active model of representation*[5].

For the *Static* subgroup, further choices are defined by the semantic feature *Agentivity*:

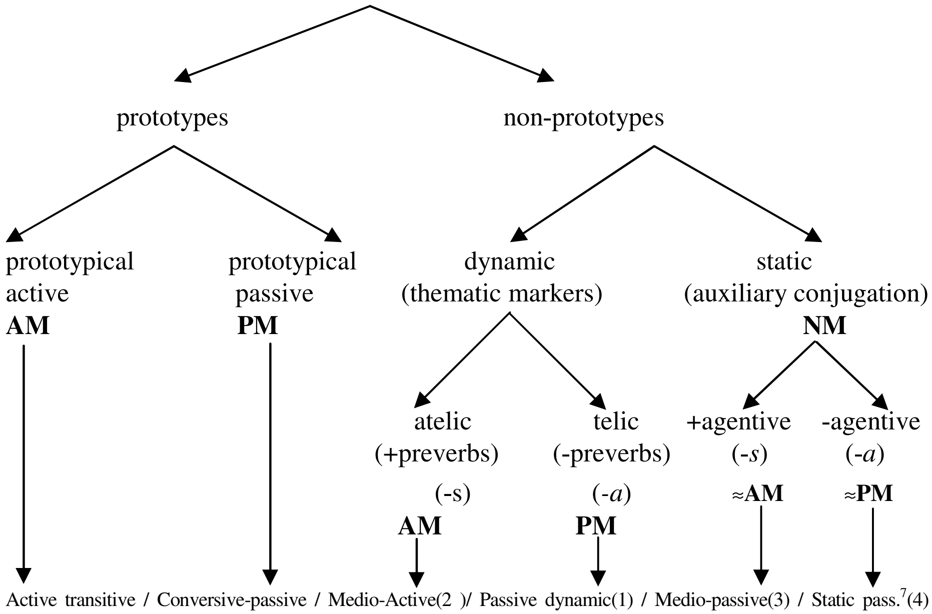
> *Verbs denoting static states, which are more or less controlled by the subject itself, have the same S.3.SG ending in present tense as the active verbs, while all other verbs choose the same S.3.SG suffix as the passive ones.*[6]

---

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

[6] In other words, the morphosyntactic feature are partially identical with the morphosyntactic features characteristic for the transitive and/or conversive-passive verb forms – only S.3.SG endings are chosen.

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



Examples:

(1)-type medial verbs: *dg-eb-a* 'S.3.SG is standing up', *šr-eb-a* 'S.3.SG becomes dry', *xm-eb-a* 'S.3.SG dries out', *tetr-d-eb-a* 'S.3.SG turns white', *k'ac-d-eb-a* 'S.3.SG becomes man', *i-q'ep-eb-a* 'S.3.SG barks', *i-gin-eb-a* 'S.3.SG is sworn at', *c'v-eb-a* 'S.3.SG lies down', *tvr-eb-a* 'S.3.SG gets drunk'.

(2)-type medial verbs: *cxovr-ob-s* 'S.3.SG lives', *pikr-ob-s* 'S.3.SG thinks', *arseb-ob-s* 'S.3.SG exists', *k'ank'al-eb-s* 'S.3.SG shivers', *gor-av-s* 'S.3.SG rolls', *suntk-av-s* 'S.3.SG breathes', *bč'-ob-s* 'S.3.SG discusses', *brial-eb-s* 'S.3.SG sparkles'.

(3)-type medial verbs: *dga-s* 'S.3.SG stand-s', *c'evs* 'S.3.SG (smb.) lies', *zi-s* 'S.3.SG sits', *dev-s* 'S.3.SG (smth.) lies', *ɣir-s* 'S.3.SG costs', *c'ux-s* 'S.3.SG worries'.

(4)-type medial verbs: *gdi-a* 'S.3.SG lies strewn/thrown about', *q'r-i-a* 'S.3.SG lie scattered/strewn a lot of smth./smb.', *pen-i-a* 'S.3.SG is spread out', *k'id-i-a* 'S.3.SG is hanging on', *c'eri-a* 'S.3.SG is written', *xat'-i-a* 'S.3.SG is drawn', *ab-i-a* 'S.3.SG is tied'.

---

7 The names for verb-classes are given according to the Georgian grammatical tradition (SHANIDZE 1973). The term *Medio-active* denotes intransitive, active, atelic verbs with ergative subject and no accusative object. On the other hand, *Medio-passive* denotes intransitive, semantically passive verbs having no active counterpart and representing static events. A subject of these verbs is always in nominative.

6 Conclusions and Notes

We have shown that the representation of the continuum of active-passive opposition together with the dynamic hierarchically organized productive model explain the complex processes that define the choice of either the active or passive formal models of representation for the non-prototypical, so-called medial forms in Georgian. The efficacy of such an approach confirms that Georgian morphological passive does not always represent the syntactic changes mostly implied by the information structuring, namely by the patient's foregrounding.

It must be mentioned that because of these peculiarities morphologically represented passive verb forms create an opposition with the syntactic passive, that is formed by the periphrastic constructions: {Passive Participle + auxiliary verb *q'opna* 'to be'}. The following examples illustrate this point.

> *da*(PV)-*c'er*(write)-*il*(PRT)-*i*(NOM)=*a*(be.PRS.S.3.SG)
> 'It is written.'
> *da*(PV)-*c'er*(write)-*il*(PRT) *i*(CV)-*kn*(be.PST)-*a*(AOR.S.3.SG)
> 'It was written.'
> *da*(PV)-*c'er*(write)-*il*(PRT) [*i*(CV)-*kn*(be.PST)-*eb*(THM)]:FUT-*a*(S.3.SG)
> 'It will be written.'

The main function of the opposition is to formalize the functional differences between the syntactically and semantically defined passive constructions. Periphrastic, analytical passive represents functional changes – patient's functional foregrounding – of semantic roles: Patient => Subject, Agent => Prepositional phrase. On the other hand, synthetic, morphological passive can represent semantically passive – non-active, yet dynamic – verbs. Even in cases when an active verb does not have the morphologically opposed conversive-passive, it still has periphrastically opposed conversive form. For example:

> *i-k'vl-ev-s* '(S)he researches smth.' : *gamo-k'vle(v)-ul-i-a* 'Smth. is researched' (yet, **i-k'vle(v)-eb-a*); *c'armo-a-dgen-s* '(S)he presents smth.' : *c'armo-dgen-il-i-a* 'Smth. is presented' (yet, **c'armo-i-dgin-eb-a*); *a-rčev-s* '(S)he chooses smth./smb.' : *a-rče(v)-ul-i-a* 'Smth./smb. is chosen' (yet, **i-rčev-eb-a*).

It can be concluded that Georgian analytical, periphrastic passive corresponds to the syntactically defined conversive-passives, while synthetic, morphological passive has different functional loading and represents mostly semantically defined peculiar verb forms.

7 Glossary

0: zero, 3: 3rd person, ACT: Active, AOR: aorist, CV: characteristic vowel, DAT: dative, ERG: ergative, FUT: future, GEN: genitive, IMP: imperfect, 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, SG: singular, SINV: inverted subject, SUBJ: subjunctive, SV: subjective version, THM: thematic suffix.

References

1. Asatiani, R.: Mart'ivi c'inadadebis t'ip'ologiuri analizi. tanamedrove kartuli salit'erat'uro enis masalaze (Typology of simple sentence. An analysis based on the data of modern literary Georgian). Mecniereba, Tbilisi (1982)
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 Main Devices of Foregrounding in the Information Structure of Georgian Sentences. In: ten Cate, B.D., Zeevat, H.W. (eds.) TbiLLC 2005. LNCS (LNAI), vol. 4363, pp. 21–30. Springer, Heidelberg (2007)
4. Chafe, W.L.: Meaning and the structure of language. Chicago Un. Press, Chicago (1971)
5. Chikobava, A.: Mart'ivi c'inadadebis p'roblema kartulshi (The Problem of the Simple Sentence in Georgian). I. Mecniereba, Tbilisi (1968)
6. Davitiani, A.: Kartuli enis sint'aksi. I. Mart'ivi c'inadadeba (Syntax of the Georgian Language. I. Syntax). Ganatleba, Tbilisi (1973)
7. Dixon, R.M.W.: Ergativity. Cambridge Un. Press, Cambridge (1979)
8. Enukidze, L.: C'inadadebis akt'ualuri danac'evreba da misi mimarteba sint'aksuri da semant'ik'uri analizis tanamedrove metodebtan (Actual Parsing of a Sentence and Its Relation to Contemporary Methods of Syntactic and Semantic Analysis). In: Tanamedrove Zogadi Enatmecnierebis Sak'itxebi VI, pp. 24–35. Enatmecnierebis inst'it'ut'i, Tbilisi (1981)
9. Harris, A.: Georgian Syntax: A Study in Relational Grammar. Cambridge University Press, Cambridge (1981); Word Order Harmonies and Word Order Change In Georgian. In: Sornicola, R., Poppe, E., Haley, A. (eds) Stability, Variation and Change of Word-Order Patterns over Time, pp. 133–163. Benjamins, Amsterdam (2000)
10. Hewitt, G.: Georgian: A Structural Reference Grammar. Benjamins, Amsterdam (1995)
11. 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, Language, Logic and Computation, Batumi, Georgia, September 12-16 (1999); September 23-28, Borjomi, Georgia (2001)
12. Kvachadze, L.: Tanamedrove kartuli enis sint'aksi (The Syntax of the Modern Georgian languge). Rubik'oni, Tbilisi (1996)
13. Kibrik, A.: Beyond Subject and Object: Toward a Comprehensive Relational Typology. In: Linguistic Typology, pp. 279–346. I. Mouton de Gruyter, Berlin (1997)
14. Shanidze, A.: Kartuli enis gramat'ik'a II. Sint'aksi (Grammar of the Georgian Language II. Syntax). Tsu gamomcemloba, Tbilisi (1948); Kartuli enis gramat'ik'is sapudzvlebi (Fundamentals of the Grammar of the Georgian Language Grammar). TSU gamomcemloba, Tbilisi
15. Shibatani, M.: Passives and Related constructions: A prototype Analysis. Language 61(4) (1985)
16. Skopeteas, S., et al.: Questionnaire on Information Structure (QUIS). In: Interdisciplinary Studies on Information Structure 4. Working Papers of the SFB 632. Universitätsverlag Potsdam, Potsdam (2006)
17. Tuite, K.: Kartvelian morphosyntax. Lincom Europa, Munich (1998); Deponent verbs in Georgian. In: Bublitz, W., Boeder, W., von Roncador, M., Vater, H. (eds.) Philologie, Typologie und Sprachstruktur: Festschrift Für Winfried Boeder. Peter Lang (2002)

Discourse Structuring Questions and Scalar Implicatures

Anton Benz and Fabienne Salfner

Centre for General Linguistics (ZAS) Berlin

Abstract. In this paper we discuss the interdependence of scalar implicatures and discourse structuring questions. We show that even prototypical cases of scalar implicatures can depend on an explicitly or implicitly given Question under Discussion. Particularly, we argue against the idea that scalar implicatures are automatically generated by the logical form of an utterance. We distinguish between three types of discourse questions each having different effects on implicatures.

1 Introduction

In this paper we discuss the interdependence of scalar implicatures and discourse structuring questions. We show that even prototypical cases of scalar implicatures can depend on an explicitly or implicitly given Question under Discussion. Particularly, we argue against the idea that scalar implicatures are automatically generated by the logical form of an utterance. This in accordance with approaches in the literature that try to show that implicatures generally have to be analysed as discourse phenomena (cf. Asher and Lascarides (2003), Geurts (2007), van Kuppevelt (1996)).

Already Grice (1975) mentions that implicatures can be discourse dependent without going further into detail. Grice considers discourse structure as one parameter among others, in particular for context dependent particularized implicatures.

In the Neo-Gricean approaches (cf. Levinson (1983, 2000)) discourse structure plays no role, but implicatures depend on the logical form of an utterance there. Chierchia (2004) even considers implicatures being part of the semantics. The discussion about Chierchia's theory led to a an increase of interest in context and discourse dependency of generalized implicatures, see Geurts (2007).

The example in (1) illustrates that even prototypical quantity implicatures depend on the kind of the preceding question, which may remain implicit:

(1) a. A: Wer hat die Bilder gemalt?
who has the pictures painted

B: Einige Bilder hat Hans gemalt/ Hans hat einige Bilder
some pictures has Hans painted Hans has some pictures
gemalt.
painted

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 35–50, 2013.

b. A: Was hat Hans gestern gemacht?
what has Hans yesterday did

B: Hans hat einige Bilder gemalt.
Hans has some picture painted

c. A: Wie lief das Geschäft gestern?
how run the business yesterday

B: Am Morgen haben einige Leute Frühstück bestellt. Dann
in.the morning have some people breakfast ordered the
war es leer, später wurde es besser.
was EXPL empty later became EXPL better

Only in (1a) *einige* 'some' triggers its common scalar implicature 'not all'. Neither in (1b) nor in (1c), where *einige* 'some' is part of an elaborating sentence, the expected implicatures are generated. Instead, in (1b) and (1c) *einige* 'some' has the implicature 'no great number' and furthermore in (1c) *einige* 'some' has the additional implicature that 'all clients who came in the morning had breakfast'. The standard theory (Levinson, 1983) would need to sidestep to a kind of cancellation or to a scale selection mechanism to explain the effects in the examples. However, example (1c) would still remain puzzling.

The fact that implicatures have to be considered as discourse phenomena is not surprising with respect to relevance implicatures, but is unexpected for scalar implicatures, as the latter are regarded as prototypical examples of generalized context independent implicatures. Therefore, in this paper we focus on scalar implicatures.

We show that implicatures are discourse phenomena and that they depend on an explicitly or implicitly given Question under Discussion. In general, the logical form of an utterance is not sufficient to derive them. That holds for scalar as well as for relevance implicatures.

At first, in Section 2 we sketch the standard theory of Levinson and discuss examples whose logical form is not sufficient for explaining their implicatures and we briefly present the recent work on the discourse dependency of scalar implicatures by Geurts (2007). In Section 3, we discuss an existing question based information structural approach (van Kuppevelt, 1996) and show that there are effects with implicatures that cannot be explained purely on the basis of his theory, but that we have to take into account the source of the discourse question: in the hearer, in the speaker's mental database or whether they are triggered by discourse relations. In Section 4, we distinguish these three types of discourse questions and discuss how they can influence implicatures.

2 Implicatures and Standard Theory

In this section we discuss the standard theory of implicatures with respect to discourse dependency of implicatures. Grice (1989) distinguishes between what is said by an utterance and what is implicated by an utterance. In (2) it is said that at least two boys came and it is implicated that 'not all boys came to the party'.

(2) Some of the boys came to the party.

Grice assumes that conversation is cooperative. Each utterance is subordinated to a common discourse goal. This is considered by the hearer while interpreting an utterance. In the classic out-of-petrol-example (Grice, 1989, p. 32) the common goal of B's utterance is to solve A's problem of finding petrol for his car:

(3) *A stands in front of his obviously immobilised car.*
A: I am out of petrol.
B: There is a garage round the corner.

On the basis of the fact that the utterance is a contribution to the task of A getting fuel for his car and on the basis that A can count on B being cooperative, A can conclude that the garage is open and that it sells fuel as far as A knows.

In addition to the cooperativity principle, Grice assumes that speakers adhere to a number of so–called *conversational maxims.* Somewhat simplifying matters, the maxims ask the speaker to be truthful (Quality), and to be as informative as possible (Quantity) as long as it is relevant (Relevance). In addition, the form of the utterance should be orderly, concise, and clear (Manner).

Grice further differentiates between particularized implicatures, which depend on the particular utterance context, and generalized implicatures, which do not depend on the utterance context. To the latter belong so–called *scalar implicatures.* These had been investigated and systematized especially by so–called Neo–Griceans (Horn, 1972, 1989; Gazdar, 1979; Levinson, 1983).

The general schema for calculating scalar implicatures is as follows: Let $A(x)$ be a sentential frame with free position x that can be filled with expressions that are ordered in a *scale* $\langle E_1, \ldots, E_n \rangle$. In (2), $A(x)$ is '*x of the boys came to the party,*' and the relevant scale is ⟨all, some⟩. In order for $\langle E_1, \ldots, E_n \rangle$ to count as a scale, it must hold that $A(E_n)$ entails $A(E_{n+1})$ but that $A(E_{n+1})$ does not entail $\mathrm{A}(\mathrm{E}_n)$. Then, according to the standard theory, an utterance of $A(E_{n+1})$ *implicates* that it is not the case that $A(E_n)$. In (2), $A(all)$ implies $A(some)$ but not vice versa; hence, an utterance of $A(some)$ implicates that $\neg A(all)$. This explains (2).

It follows from this account, that scalar implicatures are calculated on sentence level. They are generated by the logical form of an utterance and are independent of the discourse context.[1] In Levinson (2000, Sec. 1.5.2) generalized implicatures are explicitly considered as non–monotonic entailments of the utterance meaning, with the exception of manner implicatures which depend on the form of an utterance. In Neo–Gricean approaches, discourse context only comes into play when implicatures are *cancelled* by context. As they are considered non–monotonic inferences, other sources can block them. For example,

[1] "Some quite detailed arguments can be given to show that all but the Manner implicatures must be read from the semantic representation, including some specification of logical form."(Levinson, 1983, pg. 122f)

if it is known that the implicature is false, irrelevant, or that the speaker does not know whether it is true, the implicature is removed. For example, if Peter says that '*I don't know by whom the pictures were painted. Hans painted some of them. It may be he painted all,*' then the implicature of the second sentence from *some* to 'not all' is not valid. In Levinson's (2000) account, the implicature is first generated on sentence level and then cancelled by the context information.

The problems of a too simplistic view of the relation between implicatures and discourse can be illustrated by the introductory example (1), here repeated as (4)[2]:

(4) a. A: Who painted the pictures?
B: Hans painted some pictures.
$\rightsquigarrow$Hans did not paint all the pictures.

b. A: What did Hans do yesterday?
B: Hans painted some pictures.
$\not\rightsquigarrow$Hans did not paint all the pictures.

If one assumes that in (4) the logical forms of B's utterances are identical, then both occurrences should give rise to the same implicatures. There is no reason for thinking that in (4b) B does not know how many pictures Hans painted, or that this information is irrelevant. It seems that the implicature is not generated at all, rather than being generated and then cancelled by contextual information. In fact, this has led some semanticists to argue that the logical forms in (4) are fundamentally different. Diesing (1992), for example, assumes two kinds of indefinite DPs, one with a presuppositional reading and one with a non-presuppositional reading. That means that in (4a) the existence of some pictures is presupposed, whilst in (4b) the existence of some pictures is merely asserted. The logical form of the sentences then would differ in whether the internal argument of the DP is presupposed or not. Other linguists argued against Diesing's semantic ambiguity approach by proposing that the effects are rather a pragmatic phenomenon connected to the topic/focus structure, e.g. Reinhart (1995); Büring (1996). Whichever stance one takes on the semantic issues, it is not decidable on the basis of the answer alone what the speaker intended to communicate. The implicature effects arise in a question-answer structure. In the semantic ambiguity account this question–answer structure has to be taken into account for disambiguation, which is necessary for determining the logical form of the answer, and it has to be taken into account for determining the scales which are activated by the logical form. In a pragmatic account, disambiguation has to be replaced by other mechanisms which explain why the quantificational domain of *some* is restricted in (4a), and unrestricted in (4b). Such pragmatic mechanisms could be saturation and free enrichment (Carston, 2004), or rhetorical relations as studied by Asher and Lascarides (2003). In both cases, whether one follows the semantic or the pragmatic approach, discourse context plays an essential role for determining the implicatures or non–implicatures of B's answers.

[2] Note: "implicates" is marked by "$\rightsquigarrow$" and "does not implicate" is marked by "$\not\rightsquigarrow$"

So–called localist approaches tried to integrate scalar implicatures in compositional semantics (Chierchia, 2004; Levinson, 2000). Implicatures are generated at sentential or even sub–sentential level, and triggered by lexical items. These approaches would also have to assume a semantic ambiguity for avoiding undesired implicatures in Example (4b).

In the context of the debate about localist approaches, Geurts (2007) argues that implicatures are discourse phenomena rather than sentence level phenomena. He discusses three types of examples which cannot be explained without considering the wider discourse context. First, there are implicatures which can only be derived from the over-all discourse consisting of several sentences (5a). Second, there are examples in which the hearer has to consider the discourse status of certain referents (5b). Third, there are examples that show that also presupposed material can trigger implicatures (5c).

(5) a. When Jill opened the box, it contained five oranges. She took one out.
b. A cousin of mine read some of Derrida's books.
c. Jill knows that Jack took some of the apples.

In (5a) one would answer the question for the number of the remaining oranges surely by *four*. Although it could have happen that someone other than Jill has taken oranges out from the box, too. But if that had been the case, the speaker would have said it. That means that he would have chosen a more informative discourse, instead of a more informative proposition.

If the indefinite noun *a cousin* in (5b) is assumed to introduce an existentially quantified discourse referent, then a Gricean theory would predict that none of the speaker's cousins have read all books of Derrida.[3] But there are certain circumstances in which that does not have to be the case, for example, when the speaker uses this sentence to introduce a report about one particular cousin. Then, we can only derive that this particular cousin did not read all books of Derrida.

In (5c) the factive verb *to know* triggers the presupposition that the embedded utterance *Jack took some apples* is true. Thus, the implicature that 'Jack did not take all apples' is generated by a presupposition.

In this paper, we are interested in the interplay of discourse structuring questions and implicatures. Discourse structuring questions have received some attention in recent years, in particular, in connection with information structure (Roberts, 1996; Büring, 2003). It is interesting to see how Bürings's (2003) theory can explain the implicature in (4a). He proposes a question–based discourse structure from which he can derive intonation based implicatures. He assumes a D(iscourse)–Tree in which a superior question is divided into sub-questions, whose answers together provide an answer of the superior question (cf. Büring (2003, pg. 516). The contrastive topic intonation then signals that the speaker gives a partial answer only. That yields the implicature that for other topics

[3] This follows from $\exists x(cousin(x) \wedge Some(\text{D-book})(\lambda y.read(x,y))) \leadsto \neg\exists x(cousin(x) \wedge All(D\text{-}book)(\lambda y.read(x,y)))$.

another proposition might hold. See for illustration Example (6), in which *Fred* is the Contrastive Topic (CT) and *beans* is the Focus (F):

(6) A: What about Fred? What did he eat?
B: FRED_{CT} ate the BEANS_{F}

The intonation of B's answer triggers the implicature that others might have eaten other things.

Analogously, we can assume that in Example (4a) B only gives a partial answer to the question, indicated by using *some*. The question is then divided into the sub-questions '*Who painted some pictures?*' and '*Who painted the other pictures?*'. B merely answers the first one, and thus implicates that 'Hans did not paint all the pictures'.

The relation between rhetorical structure of discourse and implicatures has received little attention so far. Recently, Asher (2009) discussed a number of examples in the context of Segmented Discourse Representation Theory (Asher and Lascarides, 2003). For example, Chierchia (2004) predicts that the standard scalar implicatures occur only in upward entailing contexts. In downward entailing contexts, they vanish. Here, scales may become reverted. An example is (7a) in which it clearly is the case that the speaker is still happy if more than one person reads his book. In addition, the example gives rise to the implicature that the speaker will not be happy if no person reads his book. The rhetorical relation between antecedent and consequent is one of causation.

(7) a. If one person reads my book, I'll be happy.
b. If you take cheese or dessert, you pay 20$; but if you take both there is a surcharge.

However, in (7b), the contrast relation between the two conditionals requires the implicature from '*cheese or dessert*' to 'only cheese or only dessert'. Hence, it seems that the difference between the causation and the contrast relation is responsible for the implicature to arise or not.

The relevant discourse relation in Example (4), here repeated as (8), is that of QUESTION–ANSWER–PAIR.

(8) a. A: Who painted the pictures?
B: Hans painted some pictures.
$\leadsto$Hans did not paint all the pictures.
b. A: What did Hans do yesterday?
B: Hans painted some pictures.
$\not\leadsto$Hans did not paint all the pictures.

In (8a), the activated set of alternatives is {some, all}, whereas in (8b) the activated set of alternatives is {painting pictures, planting flowers, ...}. According to Asher in (8a) the implicature is a result of the speaker's giving an *overanswer*. *Overanswer* roughly means that when we assume that a question induces a partition on the information state (Groenendijk and Stockhof, 1984) a complete answer picks out one cell in the partition, and overanswers require additional premises to infer a complete answer, (cf. Asher (2009, pg. 21)). For Example (8a) this means that B actually answers the question '*Did Hans painted all the pictures?*', which can be considered as a sub-question of the question '*Who painted the pictures?*'. According to Asher, the answer covers the *Yes*-partition completely, and in addition a subset of the *No*-partition. This results in the implicature that 'Hans painted not all pictures'. In contrast, in (8b) the given question does not allow the accommodation of the sub-question '*Did Hans painted all pictures?*', therefore the implicature 'not all' does not arise. The differences in the set of alternatives are therefore a result of the difference in the structure of the questions.

Beside approaches that are based on discourse and/or information structure, there are several approaches that arise from the study of questions, e.g. Zeevat (1994, 2007) or Schulz and van Rooij (2006). These accounts explain scalar implicatures by assuming an exhaustivity operator although differing in detail.[4] For reason of space we confine our discussion to some brief remarks on the approaches of Zeevat (2007) and Schulz and van Rooij (2006). Zeevat applies an exhaustivity operator to the expression in an answer that corresponds to the wh-element in the question. He assumes that the exhaustivity operator is sufficient to explain scalar implicatures. So he only draws an indirect connection between questions and implicatures. His account works well for scalar expressions in NPs corresponding to the wh-element, but it makes the wrong predictions regarding the implicatures of *einige* 'some' in our minimal pair (1a) vs (1b). In (1a) the exhaustivity effect would merely be that 'nobody else beside Hans painted some pictures', whilst in (1b) we merely would arrive at the implicature that 'Hans didn't do anything beside painting pictures'. The account of Schulz and van Rooj is similar to Zeevat at least in assuming that the exhaustivity operator works on the expression that answers the question and, therefore, yielding the same predictions regarding our example (1a-b).

In the next section, we will discuss a theory on implicatures which takes advantage of the structure of discourse questions and their interplay with information structure.

3 Van Kuppevelt's Information Structural Account

To date, the most detailed account of implicatures in terms of discourse structuring questions is the information structural account of van Kuppevelt (1996). The basis of this approach is similarly to Büring (2003) the assumption that discourse is structured by a hierarchy of explicit or implicit Questions under Discussion (QUD). The questions define the discourse and the sentence topic. Implicatures are semantically inferred from information structure; more precisely, from the

[4] An exhaustivity operator was first introduced by Szabolcsi (1981) in her grammatical approach on focus, followed by Groenendijk and Stockhof (1984), whose account builds the starting point for the approaches of Zeevat (2007) and Schulz and van Rooij (2006).

topic–comment structure. If the background question is '*How many children does Nigel have?*', then the set of possible answers defines a semantic topic. The semantic comment is the alternative which is specified in the answer. Hence, if the answer is *Nigel has 14 children* then '*Nigel has* __ *children*' is the topic phrase, and '*14*' the comment phrase.

Van Kuppevelt claims that implicatures can only be triggered in the comment of an utterance but not in the topic. He provides the following examples for illustration:

(9) a. How many children does Nigel have?
Nigel has fourteen$_{comment}$ children.
$\not\rightsquigarrow$Nigel, and nobody else, has at least fourteen children.
$\rightsquigarrow$Nigel does not have more than fourteen children.

b. Who has fourteen children?
Nigel$_{comment}$ has fourteen children.
$\rightsquigarrow$Nigel, and nobody else, has at least fourteen children.
$\not\rightsquigarrow$Nigel does not have more than fourteen children.

The answer in (9a) does not implicate that Nigel, and nobody else, has at least fourteen children. But it implicates that 'Nigel does not have more than fourteen children'. The explanation according to van Kuppevelt is that *fourteen children* is the comment, and therefore triggers an implicature, whilst *Nigel* is part of the topic, hence it does not produce implicatures.

The answer in (9b) implicates that 'Nigel, and nobody else, has at least fourteen children'. But it does not implicate that Nigel does not have more than fourteen children. Since *fourteen children* is part of the topic, it does not produce implicatures, and since *Nigel* is the comment it triggers an implicature.

The following examples show some of the more intricate problems:

(10) a. A: How many books did Harry buy?
B: Harry bought $four_{comment}$ books, $if\,not\,five_{comment}$.

b. [Someone of my group bought no less than four books]
Who bought four books?
B: *Harry bought $four_{comment}$ books, if not five.

c. [I would like to know who bought how many books]
Who bought $four_{comment}$ books?
B: $Harry_{comment}$ bought $four_{comment}$ books, $if\,not\,five_{comment}$.

In Example (10a), the comment–phrase is, according to van Kuppevelt, divided into two parts, and only together do they provide an answer. Hence, the implicature from *four* to 'not more than four' can not be drawn. In (10b), the first part of the answer is already sufficient. Hence, '*if not five*' is not part of the answer, and hence not part of the comment. The utterance as a whole becomes infelicitous. In (10c), we find the same answer with two comment–phrases, with the second one '*four ... if not five*' divided into two parts as in (10a). We arrive at the implicature that 'only Harry bought four, if not five books'.

There are two aspects of van Kuppevelt's theory which we think are unsatisfactory. Both aspects are closely related to each other. The first point concerns the discourse structuring questions. Van Kuppevelt seems to assume that the questions can, in principle, be asked by the addressee of an answer. In extended dialogue contributions, these questions remain implicit, and have to be reconstructed for the analysis. When dialogues turns become more complex, it may not be obvious how to do this. Often a whole series of questions needs to be assumed for reconstructing an assertion as answer to a background question. This can be demonstrated with an example which van Kuppevelt discusses in van Kuppevelt (1995, ex. 8).

(11) F_1 A: Yesterday evening a bomb exploded near the Houses of Parliament.

Q_1 B: Who claimed the attack?

A_1 A: A well-known foreign pressure group which changed its tactic claimed the attack.

According to van Kuppevelt, in (11) F_1 is the so-called FEEDER (an utterance that does not constitute an answer to a topic-forming question) that induces a question that needs to be answered by the following discourse.

If we have a closer look at this example, we see that the question-answer structure is not so straightforward as van Kuppevelt suggests. Moreover a whole series of (implicit) questions arises to reach A_1 as an appropriate answer, i.e. '*Is there any claim of responsibility?*', '*Who claimed the attack?*', '*Is this group already known?*' and '*Is it expected that this group commits bomb attacks?*'. A_1 then answers the whole series of questions at one go.

The question what parameters can be used to derive which questions belong to such a series must be postponed to future work. This particular example suggests that on the one hand we can use information from the linguistic form ('*Who claimed the attack?*'). And at other hand there are questions that are triggered by non-at issue content (e.g. *well-known* → '*Is this group already known?*' or the appositive relative clause *which changed its tactic* → '*Is it expected that this group commits bomb attacks?*').

As might become obvious, the discussed series of questions is not derivable straightforwardly from the example. This leads to our second concern. Van Kuppevelt does not distinguish between the questions with respect to their source, i.e. whether they arises in the speaker or in the hearer or elsewhere. That this matters will become clear from the following examples.

The examples in (12) seem to contradict van Kuppevelts claim. However, they can be explained in his theory, but effort in form of additional assumptions is needed, as we will see.

(12) a. A: Who made the pictures?
B: $\text{John}_{comment}$ painted some pictures.
$\leadsto$John did not paint all the pictures.

b. What did John do yesterday?
B: John painted some $\text{pictures}_{comment}$.
$\not\leadsto$John did not paint all the pictures.

Since in (12a) *some* is part of the topic, van Kuppevelts theory would predict that *some* does not generate a scalar implicature, but in fact *some* implicates that 'John did not paint all the pictures'. And vice versa in (12b). Since *some* is part of the comment we should expect that a scalar implicature arises. But against our expectations *some* does not implicate that 'John did not paint all the pictures'.

To explain why example (12a) is still in accordance with his theory van Kuppevelt has to assume the accommodation of a more complex question in the background, the question then would be '*I would like to know who painted which pictures?*' instead of '*Who painted the pictures?*' yielding the same mechanism as for Example (10c). As a result, we get several comment phrases and therefore, speaker B gives a partial answer here. But this partition into two sub-questions cannot originate in the hearer, since he just asked '*Who painted the pictures?*'. It rather originates in the speaker.

Also Example (12b) can be explained in accordance with van Kuppevelt. We need to assume that the set of alternatives is not the set of the pictures but a set of various activities that John might have done yesterday, e.g. {painting pictures, planting flowers, visiting grandma...}. Therefore, *some* is actually not the comment and thus it cannot generate an implicature.

Now, let us turn to van Kuppevelt's own examples:

(13) a. A: Who has four children?
B: Nigel$_{comment}$ has four children.
$\rightsquigarrow$Nigel, and nobody else, has at least four children.
$\not\rightsquigarrow$Nigel does not have more than four children.
b. A: Peter has two children, and John has five children. Who has four children?
B: Nigel$_{comment}$ has four children.
$\rightsquigarrow$Nigel, and nobody else, has exactly four children.

As we have seen in (13a) above, the fact that *Nigel* is comment entails that it implicates that 'Nigel, and nobody else, has at least four children', but it does not implicate that 'Nigel does not have more than four children'. That is in accordance with van Kuppevelt. But if we extend the context in a way as in (13b), asking the very same question leads to the implicature that 'Nigel, and nobody else, has exactly four children' although *Nigel* is comment. This can be explained by the assumption that the utterance of speaker A *Peter has two children, and John has five children.* indicates an implicit, more complex question in the background, namely '*Who has how many children?*'. And, the utterance *Peter has two children, and John has five children.* answers the sub-questions '*Who has four children?*' and '*Who has five children?*'.

But as the discussion about (14b) will reveal, sometimes such explanations are not sufficient.

(14) a. A: Parents with at least four children get free entry. Who has (at least) four children?
B: Nigel$_{comment}$ has four children.

⇝Nigel, and nobody else, has at least four children.
⇝̸Nigel does not have more than four children.

b. A: Parents with at least four children get free entry. Who has (at least) four children?
B: Nigel$_{comment}$ has seven children.
⇝̸Nigel, and nobody else, has at least four children.
⇝Nigel does not have more than seven children.

The answer in (14a) implicates that 'Nigel, and nobody else, has at least four children'. And it does not implicate that 'Nigel does not have more than four children'. Thus, the implicatures are generated as predicted by van Kuppevelts theory.

Interestingly, (14b) does not implicate that 'Nigel, and nobody else, has at least four children'. But it implicates that 'Nigel does not have more than seven children.' This implicature is generated, since B gives an over-informative answer. The corresponding question to this answer was obviously not given by the hearer. But it could have arisen in the speaker himself as a query of his mental database. This can also serve as explanation for the question why the implicature that 'Nigel and nobody else, has at least four children' does not occur. The speaker then names the first person who comes into his mind for whom the question applies, that means he gives a mention-some answer. In van Kuppevelt's terms the speaker gives a partial answer.

However, van Kuppevelt's theory does not yield this kind of derivations straightforwardly. We can conclude so far that information structure alone is not sufficient to explain implicatures. Moreover, it seems to be important to differentiate where the questions arise.

In the next section, we discuss in what way implicatures can depend on discourse questions by the example of German *einige* 'some'. We distinguish between questions that arise in the hearer, questions that are raised by discourse relations and questions that can be considered as queries in the speaker's mental database.

4 Roles of Question under Discussion

Before we turn to the discussion of the relevant data we need to mention that English *some* and German *einige* 'some' obviously differ in their behavior regarding implicatures, as illustrated in (15).

(15) A: What did Hans do yesterday?
B: Hans painted *some* pictures.
⇝̸ Hans did not paint all the pictures.
⇝̸ Hans painted a small number of pictures.
B': Hans hat *einige* Bilder gemalt.
⇝̸ Hans did not paint all the pictures.
⇝ Hans paint a small number of pictures.

As the example illustrates, the meaning of English *some* and German *einige* is identical. But the implicature of German *einige* is more complicated. English *some* triggers a scale $\langle some, all \rangle$, but German *einige* can trigger two different scales: $\langle einige, alle \rangle$ or $\langle einige, viele \rangle$.

From the discussion above we have seen that the question which kind of scale is activated is an issue in both the theory of Asher and the theory of van Kuppevelt. But which scale is activated can depend on the kind of question, as the following example with a definite-indefinite contrast shows.

(16) a. A: Who painted the pictures?
B: John painted *einige* pictures.
b. A: Who painted pictures?
B: John painted *einige* pictures.
c. A: What did John do yesterday?
B: John painted *einige* pictures.

Example (16) illustrates the dependency of the implicature from the definite-indefinite contrast in the question. In (16a) with a definite article in the question, *einige* 'some' implicates 'not all'. But in (16b) with a bare plural in the question, there is no implicature of *einige* 'some', neither 'not all' nor 'not a large number' (as *einige* implicates in (16c)). Interestingly, the neutral alternative to the examples in (16) is not 'John painted all pictures'. It seems that the contrast between 'John painted *einige* pictures' and 'John painted *the* pictures.' is sufficient to explain the implicature. This is supported by the observation that we can replace the non-restrictive definite DP by a restrictive version still getting a scalar implicature, see (17):

(17) John painted the blue pictures.
$\rightsquigarrow$John painted not all pictures.

This suggests that the basis for the implicature here is neither the scale ⟨all, some⟩ nor the scale ⟨many, some⟩, but rather a part-of relation introduced by contrast to the definite DP. The example is in accordance to the observations of Hirschberg (1991) (see also Levinson (2000, Sec. 2.2.4)), who introduces scales based on dominance relations to explain generalized scalar implicatures.

These kind of examples demonstrates again that the logical structure alone is obviously not sufficient to explain implicatures.

The source from which Questions under Discussion originate and how they influence potential implicatures can differ. We distinguish between a) implicit or explicit questions which originate in the hearer, b) questions raised by discourse relations and c) questions arising as queries to the speaker's mental database.

4.1 Implicit or Explicit Questions Which Originate in the Hearer

In this section we discuss a couple of examples including *einige* 'some' with various questions that originate in the hearer.

(18) a. A: Who painted the pictures?
B: John painted *einige* pictures.
⇝John did not paint all the pictures.

b. A: What did John do yesterday?
B: John painted *einige* pictures.
⇝John painted a small number of pictures.

As we discussed above the logical forms of B's utterances are identical, but the utterances activate different scales. This yield different implicatures.

Another example is (19).

(19) a. A: How was business going yesterday?
B: In the morning, there were *einige* people who ordered breakfast. Then it was very quiet, later it became better.
⇝Not many people ordered breakfast.
⇝In the morning all people ordered breakfast.

In (19), the question asks for the course of business, rather than the amount of people who had breakfast. As the amount is not salient *einige* 'some' just has its semantic meaning as indefinite yielding in the implicature 'not many'.

A likewise quite interesting example is (20b).

(20) a. A: Who came to the party?
B: *Einige* students came to the party.
⇝Not all students came to the party.

b. B': *Einige* proof theorists from Humboldt-university came to the party.
$\not\rightsquigarrow$Not all proof theorists from Humboldt university came to the party.

Example (20a) is one of the typical examples for an utterance with *einige* 'some' generating the scalar implicature 'not all'. But if we replace *einige Studenten* 'some students' by *einige Beweistheoretiker von der Humboldt-Universität* 'some proof theorists from Humboldt-university' as in (20b) the implicature vanishes immediately. If the implicature really would be generated by the logical form we would expect no differences between (20a) and (20b) with regard to the implicature generating behaviour. In fact, it seems plausible that in (20a) the implicature that 'not all students came' follows from world knowledge (i.e. that at Humboldt-university there are a lot of students), and therefore the implicature is not triggered by *einige* 'some'. And, to go one step further it is plausible to assume that in fact there is no implicature at all. We only get a real implicature if we replace *einige* 'some' by *einige der* 'some of the' indicating a part-of relation. Actually, in (20b) the composition of the guest's group is at issue and not the proportion of students partying. Therefore no implicature arises.

4.2 Questions Raised by Discourse Relations

Beside the questions that anticipate questions that might arise in the hearer there are questions that are triggered by discourse relations, as illustrated in (21).

(21) a. A: This vehicle is very secure.
B: Really? Last year there had been some/*einige* accidents with it.
⇝The number of accidents was high.

b. A: Do you think John can drive?
B: He had *einige* glasses of beer.
⇝He had probably too many glasses of beer.

In (21a), we see an utterance in which *einige* 'some' generates the implicature 'many', i.e. interestingly the stronger expression on the scale is implicated. That can be explained by assuming that the most likely discourse relation between the utterance that contains *einige* and the preceding statement of speaker A is Counterevidence, triggering an implicit question '*What is B's counterevidence?*'. Thus the implicature is generated by the fact that the stronger interpretation is necessary to maintain the expected discourse relation. Explanations in this vein were suggested by Asher and Lascarides (2003). The example is furthermore important, since it shows that a theory that merely based on scales as well as a theory purely based on discourse questions have problems to find an explanation.

(21b) is another example in which a discourse relations are responsible for the implicatures, namely Question-answer relation and additionally an Evidence relation. In this example *einige* implicates 'too many'.

4.3 Questions Arising as Queries in the Speaker's Mental Database

A third type of discourse questions are questions that function as queries in the speaker's mental database. Let us consider the Nigel-example again.

(22) A: Parents with at least four children get free entry. Does Nigel have four children?
B: Nigel has seven children.
⇝Nigel does not have more than seven children.

The answer in (22) implicates that 'Nigel does not have more than seven children'. B seems to answer the question: '*How many children does Nigel have?*', although this question is not asked by the hearer. This suggests that the question originates in the speaker himself as a query in his mental database. The example shows that in spite of the obvious irrelevance of the question '*How many children the parents have?*' an over-informative answer yields an implicature. Therefore, it would be a misunderstanding to believe that the question structure entails straightforwardly that scalar expressions in particular sentence positions do not generate implicatures.

5 Conclusion

In this paper, we have illustrated that implicatures are discourse phenomena and that they depend on an explicitly or implicitly given Question under Discussion.

In general, the logical form of an utterance is not sufficient to derive them. That holds even for the most prototypical example of scalar implicatures: those that are triggered by *some/einige* and numerals.

Our starting point was the standard theory which explains implicatures using scales of expressions $(E_1, \ldots, E_n)$ such that the use of a scale element E_i in a sentence frame $A(E_i)$ automatically activates a scale and triggers an implicature. The examples we discussed have shown that in general the scales connected to an element are not unique and that the choice of the scale depends on the discourse context.

In Section 3 we presented an approach that derives the scale selection and activation from discourse structuring background questions in connection with information structure. We have seen that this theory alone is also not sufficient to explain the various phenomena.

All things considered, it has turned out that it is useful to differentiate between three types of discourse structuring questions: those, which arise in the hearer, those, which are a kind of queries on the speaker's mental database and those being triggered by discourse relations. The next challenge we have to take up is the question how a theory that predict in what way a whole series of discourse question is composed could look like.

References

Asher, N.: Implicatures in discourse (2009), handout

Asher, N., Lascarides, A.: Logics of Conversation. Cambridge University Press, Cambridge (2003)

Büring, D.: A weak theory of strong readings. In: Galloway, T., Spence, J. (eds.) Proceedings of Semantics and Linguistic Theory (SALT) 6, pp. 17–34. CLC Publications, Ithaca (1996)

Büring, D.: On D-trees, beans, and B-accents. Linguistics & Philosophy 26(5), 511–545 (2003)

Carston, R.: Relevance theory and the saying/implicating distinction. In: Horn, L., Warden, G. (eds.) The Handbook of Pragmatics, pp. 632–656. Blackwell Publishing, Oxford (2004)

Chierchia, G.: Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: Belletti, A. (ed.) Structures and Beyond, pp. 39–103. Oxford University Press, Oxford (2004)

Diesing, M.: Indefinites. MIT Press, Cambridge (1992)

Gazdar, G.: Pragmatics: Implicature, Presupposition, and Logical Form. Academic Press, New York (1979)

Geurts, B.: Implicatures as discourse phenomenon. In: Puig-Waldmüller, E. (ed.) Proceedings of the Sinn und Bedeutung 11, pp. 261–275 (2007)

Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and Semantics, vol. 3, pp. 41–58. Academic Press, New York (1975)

Grice, H.P.: Studies in the Way of Words. Harvard University Press, Cambridge (1989)

Groenendijk, J., Stockhof, M.: Studies in the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, University of Amsterdam (1984)

Hirschberg, J.: A Theory of Scalar Implicature. Garland Publishing, New York (1991)

Horn, L.R.: On the Semantic Properties of the Logical Operators in English. Ph.D. thesis, Indiana University (1972)
Horn, L.R.: A Natural History of Negation. University of Chicago Press, Chicago (1989)
van Kuppevelt, J.: Discourse structure, topicality, and questioning. Journal of Linguistics 31, 109–147 (1995)
van Kuppevelt, J.: Inferring from topics: Scalar implicatures as topic-dependent inferences. Linguistics and Philosophy 19, 393–443 (1996)
Levinson, S.C.: Pragmatics. Cambridge University Press, Cambridge (1983)
Levinson, S.C.: Presumptive Meanings - The Theory of Generalized Conversational Implicatures. MIT Press, Cambridge (2000)
Reinhart, T.: Interface Strategies. OTS Working Papers in Linguistics (1995)
Roberts, C.: Information structure in discourse: Toward an integrated formal theory of pragmatics. In: Yoon, J.H., Kathol, A. (eds.) OSU Working Papers in Linguistics, vol. 49, pp. 91–136. The Ohio State University, Department of Linguistics, Ohio (1996)
Schulz, K., van Rooij, R.: Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. Linguistics and Philosophy 29, 205–250 (2006)
Szabolcsi, A.: The semantics of the topic-focus articulation. In: Formal Methods in the Study of Language, Part 2. Mathematical Centre Tracts, Amsterdam (1981)
Zeevat, H.: Exhaustivity, questions and plurals in update semantics. In: Questions in Dynamic Semantics, Current Research in the Semantics/Pragmatics Interface, vol. 17, pp. 161–194. Elsevier (2007), edited reprint of Zeevat 1994

Towards a Logic of Information Exchange*
An Inquisitive Witness Semantics

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen

ILLC/Department of Philosophy, University of Amsterdam

1 Introduction

Traditionally, the meaning of a sentence is identified with its *truth conditions*. This approach is driven by the age-old attention that philosophy has devoted to the study of argumentation. In terms of truth conditions one defines *entailment*, the crucial notion that rules the soundness of an argument: a sentence φ is said to entail another sentence ψ in case the truth conditions for φ are at least as stringent as the truth conditions for ψ.

But argumentation is neither the sole, nor the primary function of language. One task that language more widely and ordinarily fulfills, among others, is to enable the *exchange of information* between individuals. If we embark on the enterprise of studying this particular use of language, the objects of our study are no longer arguments, or proofs, but rather conversations, or dialogues. Just like in logic we traditionally pursue a formal characterization of well-formed proofs, we would then like to obtain a characterization of well-formed dialogues. Thus, as advocated in [17,18], instead of the notion of entailment, which judges whether a sentence can be inferred from a given set of premises, we need to consider another logical notion, which judges whether a sentence forms a pertinent response to the foregoing discourse. This notion, which we will call *compliance*, may be regarded as the crucial logical relation in the study of information exchange. The aim of this paper is to contribute to a better formal characterization of compliance.

In order to pursue our goal, the first thing we need is a shift in perspective on meaning. For, the static notion of meaning as truth conditions is not particularly suited to understand the dynamics of information exchange, or at the very least, not in its usual form. Such a shift in perspective has been initiated by Stalnaker [33], who gave the notion of meaning a dynamic and conversational twist. He proposed to take the meaning of a sentence to consist in its potential to bring about an information change, enhancing the so-called *common ground* of a conversation, and with it the information states of the conversational participants.

* We are very grateful to Morgan Mameni and Matthijs Westera for helpful discussions of the issues addresses in this paper and many closely related topics. Earlier versions of the paper were presented at the *Prague Workshop on the Logic of Questions*, October 26, 2011; and at the workshop *In Search of Answers: The Guiding Role of Questions in Discourse and Epistemology*, Center for Formal Epistemology, Carnegie Mellon, Pittsburg, November 5, 2011. The research reported here was made possible by financial support from the Netherlands Organization of Scientific Research (NWO), which is gratefully acknowledged.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 51–72, 2013.

However, if we take seriously the goal to study information exchange through conversation, this change of perspective is not sufficient. For, Stalnaker's proposal is limited to *assertions*, whose meaning can be identified with their informative potential. In conversation, information exchange takes place through a complex interplay of requests and provisions of information, in a process in which issues are raised, addressed, and sometimes resolved. It is the requests for information that drive the development of the conversation, setting the momentary goal of the exchange and thus eliciting certain reactions from other participants. If we really want to understand how information exchange works, our notion of meaning should thus encompass *inquisitive potential* (the potential to request information) alongside informative potential (the potential to provide information). This simple observation forms the cornerstone of *inquisitive semantics* [6,11,20,32]. This framework is intended to provide new logical foundations for the analysis of discourse, especially the type of discourse that is aimed at the exchange of information. For instance, Farkas and Roelofsen [15] show that inquisitive semantics makes it possible to both simpify and enrich the discourse theory of Farkas and Bruce [14], which in turn builds on much previous work on discourse [23,33,4,12,16,31,22,1,3]. A comparison of inquisitive semantics with classical theories of questions [24,28,21], treatments of questions in dynamic semantics [27,26,17], and an earlier version of inquisitive semantics [18,30], shown to be defective in [6,11], is provided in [9,19].

The most basic implementation of inquisitive semantics, the system Inq_B, will be summarized in section 2.1 below. Its notion of meaning encompasses both informative and inquisitive potential. The question, then, is whether this framework allows us to construe a suitable formal notion of compliance. If we limit our attention to a propositional language, the answer seems to be positive: in section 2.2, we present a natural candidate, which was first defined and discussed in [20] and [8]. Unfortunately, however, examples from [6] and [7] show that the same strategy does not yield satisfactory results in the case of a first-order language. These examples are discussed in section 2.3. We will argue that, in fact, no fully satisfactory notion of compliance can be given based on the notion of meaning used in Inq_B, since sentences with intuitively distinct compliant responses are assigned the same semantic value. These considerations will lead us to devise a more fine-grained semantics, described in section 3, that we will call *inquisitive witness semantics*, Inq_W. While this enriched system coincides with Inq_B on the treatment of informative and inquisitive content, it allows for the formulation of a notion of compliance that does justice to the formerly problematic cases, assigning them the intended set of compliant responses. However, in the conclusion we will look at an example that shows that our solution is not yet quite general, and needs to be further refined.

2 Background

We start with a brief recapitulation of Inq_B. We will first consider the language of propositional logic, and then move on to the first-order setting. More elaborate expositions of Inq_B can be found in [6,11,20,32].

2.1 Propositional InqB

In this section we consider a language $\mathcal{L}_\mathcal{P}$, whose formulas are built up from $\bot$ and a set $\mathcal{P}$ of proposition letters, by means of the binary connectives $\wedge, \vee$ and $\rightarrow$. We use $\neg\varphi$ as an abbreviation of $\varphi \rightarrow \bot$, $!\varphi$ as an abbreviation of $\neg\neg\varphi$, and $?\varphi$ as an abbreviation of $\varphi \vee \neg\varphi$. We refer to $!\varphi$ and $?\varphi$ as the non-inquisitive and the non-informative projection of φ, respectively.

The basic ingredients for the semantics are *worlds* and *states*.

Definition 1 (Worlds)
A world is a function from $\mathcal{P}$ to $\{0,1\}$. We denote by W the set of all worlds.

Definition 2 (States)
A state is a set of worlds. We denote by $\mathcal{S}$ the set of all states.

The meaning of a sentence will be defined in terms of the notion of *support* (just as, in a classical setting, the meaning of a sentence is usually defined in terms of truth). Support is a relation between states and formulas. We write $s \models \varphi$ for 's supports φ'. Intuitively, the support relation captures the conditions under which a formula φ is *redundant* in a state s, meaning that φ does not provide any information that is not already available in s and does not raise any issues that are not already resolved in s. This intuition will be made more precise momentarily, when formal notions of informativeness and inquisitiveness have been introduced, see in particular fact 6 below.[1]

Definition 3 (Support)

$s \models p$	*iff*	$\forall w \in s : w(p) = 1$
$s \models \bot$	*iff*	$s = \emptyset$
$s \models \varphi \wedge \psi$	*iff*	$s \models \varphi$ *and* $s \models \psi$
$s \models \varphi \vee \psi$	*iff*	$s \models \varphi$ *or* $s \models \psi$
$s \models \varphi \rightarrow \psi$	*iff*	$\forall t \subseteq s :$ *if* $t \models \varphi$ *then* $t \models \psi$

It follows from the above definition that the empty state supports any sentence φ. Thus, we may think of $\emptyset$ as the *absurd* state. The following two facts bring out two basic properties of the support relation.

Fact 1 (Persistence). *If $s \models \varphi$ then for every $t \subseteq s$: $t \models \varphi$*

Fact 2 (Singleton States Behave Classically). *For any w and φ:*

$$\{w\} \models \varphi \quad \Longleftrightarrow \quad w \models \varphi \ \text{ in classical propositional logic}$$

[1] For further discussion of the notion of support we refer to [11,19]. Inq_B can also be presented in such a way that support is not the basic notion [10,32]. Rather, this alternative mode of presentation starts with a direct recursive definition of the propositions expressed by the formulas in the language. The proposition expressed by a formula φ then determines in which states φ is informative and/or inquisitive, and in which states φ is supported.

It can be derived from definition 3 that the support-conditions for $\neg\varphi$, $!\varphi$, and $?\varphi$ are as follows.

Fact 3 (Support for Negation and the Projection Operators)

1. $s \models \neg\varphi$ *iff* $\forall w \in s : w \not\models \varphi$
2. $s \models !\varphi$ *iff* $\forall w \in s : w \models \varphi$
3. $s \models ?\varphi$ *iff* $s \models \varphi$ *or* $s \models \neg\varphi$

In terms of support, we define the *proposition* expressed by a sentence.

Definition 4 (Propositions, Entailment, and Equivalence)

- $[\varphi] := \{s \in \mathcal{S} \mid s \models \varphi\}$
- $\varphi \models \psi$ *iff* *for all s: if* $s \models \varphi$, *then* $s \models \psi$
- $\varphi \equiv \psi$ *iff* $\varphi \models \psi$ *and* $\psi \models \varphi$

We will refer to the maximal elements of $[\varphi]$ as the *alternatives* for φ.

Definition 5 (Alternatives). *Let* φ *be a sentence.*

1. *Every maximal element of* $[\varphi]$ *is called an* alternative *for* φ.
2. *The* alternative set *of* φ, $[\![\varphi]\!]$, *is the set of alternatives for* φ.

The following result guarantees that the alternative set of a sentence completely determines the proposition that the sentence expresses, and vice versa.

Fact 4 (Propositions and Alternatives). *For any state s and sentence* φ*:*

$$s \in [\varphi] \iff s \text{ is contained in some } \alpha \in [\![\varphi]\!]$$

Example 1 (Disjunction). Inquisitive semantics differs from classical semantics in its treatment of disjunction. To see this, consider figures 1(a) and 1(b). In these figures, it is assumed that $\mathcal{P} = \{p, q\}$; world 11 makes both p and q true, world 10 makes p true and q false, etcetera. Figure 1(a) depicts the classical meaning of $p \vee q$: the set of all worlds that make at least one of p and q true. Figure 1(b) depicts the alternative set of $p \vee q$ in Inq_B. It consists of two alternatives. One alternative is made up of all worlds that make p true, and the other of all worlds that make q true.

We think of a sentence φ as expressing a proposal to update the common ground of a conversation—formally conceived of as a set of possible worlds—in such a way that the new common ground supports φ. In other words, given fact 4, a sentence proposes to update the common ground in such a way that the resulting common ground is contained in one of the alternatives for φ.

Worlds that are not contained in any state supporting φ will not survive any of the updates proposed by φ. In other words, if any of the updates proposed by φ is executed, all worlds that are not contained in $\bigcup[\varphi]$ will be eliminated. Therefore, we refer to $\bigcup[\varphi]$ as the *informative content* of φ.

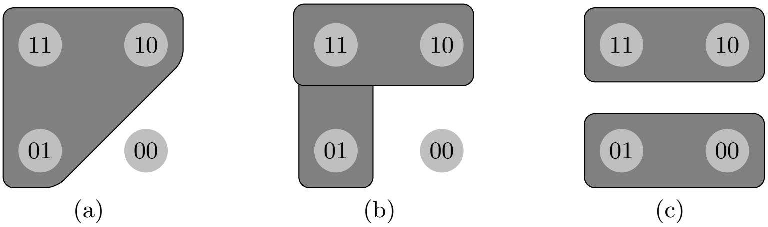


Fig. 1. (a) classical picture of $p \vee q$, (b) inquisitive picture of $p \vee q$, and (c) polar question $?p$

Definition 6 (Informative Content). $\mathit{info}(\varphi) := \bigcup[\varphi]$

Classically, the informative content of φ is captured by the set of all worlds in which φ is classically true. We refer to this set of worlds as the *truth set* of φ.

Definition 7 (Truth Sets)
The truth set of φ, $|\varphi|$, *is the set of all worlds where* φ *is classically true.*

The following result says that, as far as informative content goes, Inq_B does not diverge from classical propositional logic. In this sense, Inq_B is a conservative extension of classical propositional logic.

Fact 5 (Informative Content is Classical). *For any* φ*:* $\mathit{info}(\varphi) = |\varphi|$

A sentence φ is informative in a state s iff it proposes to eliminate at least one world in s, i.e., iff $s \cap \mathsf{info}(\varphi) \neq s$. On the other hand, φ is inquisitive in s iff in order to reach a state $s' \subseteq s$ that supports φ it is not enough to incorporate the informative content of φ itself into s, i.e., $s \cap \mathsf{info}(\varphi) \not\models \varphi$, which means that φ requests a response from other participants that provides additional information.

Definition 8 (Inquisitiveness and Informativeness in a State)

– φ *is* informative *in* s *iff* $s \cap \mathit{info}(\varphi) \neq s$
– φ *is* inquisitive *in* s *iff* $s \cap \mathit{info}(\varphi) \not\models \varphi$

As mentioned above, the support relation intuitively captures when a formula φ is redundant in a state s. The following fact establishes that this intuition is reflected by the system in a very precise way: φ is supported by s just in case it is neither informative nor inquisitive in s.

Fact 6 (Support as Redundancy)

– $s \models \varphi$ *iff* φ *is neither informative nor inquisitive in* s*.*

Besides notions of informativeness and inquisitiveness *relative to a state* we may also define absolute notions of informativeness and inquisitiveness.

Definition 9 (Absolute Inquisitiveness and Informativeness)

– φ *is* informative *iff it is informative in at least one state.*
– φ *is* inquisitive *iff it is inquisitive in at least one state.*

By persistence (fact 1), an alternative characterization of informativeness and inquisitiveness can be given in terms of informative content, as follows.

Fact 7 (Inquisitiveness, Informativeness, and Informative Content)

- φ *is* informative *iff* $\mathsf{info}(\varphi) \neq W$
- φ *is* inquisitive *iff* $\mathsf{info}(\varphi) \not\models \varphi$

Finally, by fact 4, inquisitiveness can also be characterized in terms of the alternative set for a formula.

Fact 8 (Inquisitiveness and Alternatives)

- φ *is* inquisitive *iff* $[\![\varphi]\!]$ *contains at least two alternatives.*

Example 2 (Disjunction Continued) As in the classical setting, $p \vee q$ is *informative*, in that it proposes to eliminate worlds where both p and q are false. But it is also *inquisitive*, in that it proposes to move to a state that supports p or to a state that supports q, while merely eliminating worlds where both p and q are false is not sufficient to reach such a state. Thus, $p \vee q$ requests a response that provides additional information. This inquisitive aspect of meaning is not captured in the classical setting.

The fact that disjunction is hybrid in the formal system, does not as such embody an *empirical* claim that indicative disjunctions in natural language are both informative and inquisitive. As is shown in [19], the formal system is equally applicable in case it can be argued that, e.g., in a language like English, the semantic properties of informativeness and inquisitiveness are strictly divided over the two syntactic sentential categories of indicative and interrogative sentences, and that hybrid sentences do not exist. Inquisitive semantics offers a general logical framework that can be used to formulate and compare different linguistic theories that are concerned with informative and inquisitive aspects of meaning, but it is not a linguistic theory in itself.

Definition 10 (Questions, Assertions, and Hybrids)

- φ *is a* question *iff it is not informative;*
- φ *is an* assertion *iff it is not inquisitive;*
- φ *is a* hybrid *iff it is both informative and inquisitive.*

Example 3 (Questions, Assertions, and Hybrids) We saw above that $p \vee q$ is both informative and inquisitive, i.e., hybrid. Figure 1(a) depicts the alternative set of $!(p \vee q)$, which consists of exactly one alternative. So $!(p \vee q)$ is an assertion. Figure 1(c) depicts the alternative set of $?p$. Together the alternatives for $?p$ cover the entire logical space, so $?p$ does not propose to eliminate any world. That is, $?p$ is a question.

The framework of propositional basic inquisitive semantics makes it possible to express a wide range of different types of questions. Next to simple polar questions like $?p$, it can also deal with conditional questions like $p \to ?q$, alternative questions like $?(p \vee q)$, and choice questions like $?p \vee ?q$.

2.2 Compliance

We now move on to consider a particular notion of compliant responses. In order to motivate this notion, consider the question in (1) and the responses in (1a-d).

(1) Is Mary going to the party?
  a. Yes, she is going.
  b. John is going.
  c. Yes, she is going, and John is going with her.
  d. Yes, she is going; are you going as well?

We would like to have a notion of compliance under which (1a) is a basic compliant response to (1), but (1b-d) are not. (1b) should not count as a basic compliant response because it does not resolve the issue raised by (1), (1c) should not count as a basic compliant response because it provides more information than is necessary to resolve the issue raised by (1), and (1d) should not count as a basic compliant response because, besides providing exactly enough information to resolve the issue raised by (1), it also raises a new issue. Thus, basic compliant responses are those responses that provide exactly enough information to resolve the given issue and do not raise any new issues.

Definition 11 (Basic Compliant Responses)
ψ is a basic compliant response to φ just in case:

1. *ψ is an assertion*
2. *$\psi \models \varphi$*
3. *There is no assertion ξ such that $\psi \models \xi$, $\psi \not\equiv \xi$, and $\xi \models \varphi$*

Equivalently, a basic compliant response to a formula φ may be characterized as an assertion whose informative content coincides with one of the alternatives for φ. For, a response to φ is issue-resolving just in case it provides enough information to establish a state that supports φ. A basic compliant response is defined as a *minimally informative* issue-resolving assertion, that is, one that establishes a *minimally informed* state that supports φ. By definition, these states are precisely the alternatives for φ.

Fact 9 (Basic Compliant Responses)
ψ is a basic compliant response to φ iff $[\![\psi]\!] = \{\alpha\}$ for some $\alpha \in [\![\varphi]\!]$.

If φ is a question, then the basic compliant responses may be viewed as the basic answers to the question, corresponding to those answers that are usually called *direct* or *principal* answers in various erotetic frameworks [2,29,25,34]. Thus, a theory of compliance is also automatically a theory of answerhood for questions.

It should be emphasized that basic compliant responses are not supposed to be the *only* responses to a given sentence that are predicted to be compliant. In terms of basic compliant responses, a more general notion of compliant responses can be defined (see [20]). In the case of questions, among the non-basic compliant responses we find partial answers, as well as sub-questions. For our present purposes, however, considering the general notion of compliance is of little interest.

For, the problem we will focus on concerns essentially the determination of the set of *basic* compliant responses to a sentence.

As long as we restrict ourselves to the language of propositional logic, the notion of basic compliant responses, and the more general notion of compliance that it gives rise to, seem to give satisfactory results (again, see[20]). However, we will see right below that this is no longer generally the case if we move to the first-order setting.

2.3 First-Order InqB

Let $\mathcal{L}$ be a first-order language. The worlds that make up a *state* will now be first-order models for $\mathcal{L}$. We will assume that all worlds in a state share the same domain and the same interpretation of individual constants and function symbols. This assumption is enacted using the notion of a discourse model.

Definition 12 (Discourse Models)
A discourse model $\mathbb{D}$ for $\mathcal{L}$ is a pair $\langle D, I\rangle$, where D is a domain and I an interpretation of all individual constants and function symbols in $\mathcal{L}$.

Definition 13 ($\mathbb{D}$-worlds and $\mathbb{D}$-states)
Let $\mathbb{D} = \langle D, I\rangle$ be a discourse model for $\mathcal{L}$. Then:

- *A $\mathbb{D}$-world w is a model $\langle D_w, I_w\rangle$ such that $D_w = D$ and I_w coincides with I as far as individual constants and function symbols are concerned. The set of all $\mathbb{D}$-worlds is denoted by $W_{\mathbb{D}}$.*
- *A $\mathbb{D}$-state is a set of $\mathbb{D}$-worlds. The set of all $\mathbb{D}$-states is denoted by $\mathcal{S}_{\mathbb{D}}$.*

Thus, a $\mathbb{D}$-state s is a set of first-order models for $\mathcal{L}$ that are all based on the same discourse model $\mathbb{D}$. This means that all the models in s share the same domain, and assign the same interpretation to individual constants and function symbols. The interpretation of *predicate symbols* is not fixed by $\mathbb{D}$, and may therefore differ from model to model in s. This amounts to the assumption that the domain of discourse and the interpretation of individual constants and function symbols are common knowledge among the participants, and that the exchange of information only concerns the denotation of the predicate symbols.[2]

The definitions below assume a fixed discourse-model $\mathbb{D} = \langle D, I\rangle$ for $\mathcal{L}$. Moreover, for any assignment g, we denote by $|\varphi|_g$ the *truth set* of φ relative to g, i.e., the set of worlds w such that $w \models_g \varphi$ in classical first-order logic.

Definition 14 (Support in First-Order Inq_B)
Let s be a $\mathbb{D}$-state, g an assignment, and φ a formula in $\mathcal{L}$.

$$
\begin{array}{lll}
s \models_g \varphi & \textit{iff} \quad s \subseteq |\varphi|_g & \textit{for atomic } \varphi \\
s \models_g \bot & \textit{iff} \quad s = \emptyset &
\end{array}
$$

[2] This simplifying assumption is made here for ease of exposition. A similar semantics without this assumption is also conceivable, but the extra complexity involved would not be relevant to the issue we shall be concerned with.

$$
\begin{array}{lll}
s \models_g \varphi \wedge \psi & \textit{iff} & s \models_g \varphi \textit{ and } s \models_g \psi \\
s \models_g \varphi \vee \psi & \textit{iff} & s \models_g \varphi \textit{ or } s \models_g \psi \\
s \models_g \varphi \to \psi & \textit{iff} & \forall t \subseteq s : \textit{if } t \models_g \varphi \textit{ then } t \models_g \psi \\
s \models_g \forall x.\varphi & \textit{iff} & s \models_{g[x/d]} \varphi \textit{ for all } d \in D \\
s \models_g \exists x.\varphi & \textit{iff} & s \models_{g[x/d]} \varphi \textit{ for some } d \in D
\end{array}
$$

Definition 15 (Propositions, Entailment, and Equivalence)

- $[\varphi]_g \quad := \quad \{s \in \mathcal{S}_{\mathbb{D}} \mid s \models_g \varphi\}$
- $\varphi \models \psi$ *iff for all* $\mathbb{D}$, s *and* g: *if* $s \models_g \varphi$, *then* $s \models_g \psi$
- $\varphi \equiv \psi$ *iff* $\varphi \models \psi$ *and* $\psi \models \varphi$

All the basic logical notions defined in the propositional setting, like informativeness, inquisitiveness, questions, assertions, and hybrids, carry over immediately to the first order setting. As is to be expected, given the inquisitive nature of disjunction, the existential quantifier is inquisitive as well. A state s may well embody the information that at least one object in the domain has the property P, without supporting the existential $\exists x.Px$. For, the former merely requires that in every world $w \in s$ there be some object $d \in D$ such that $d \in I_w(P)$. In order to support $\exists x.Px$, on the other hand, there must be some object $d \in D$ such that in every $w \in s$: $d \in I_w(P)$. In other words, what is required to support the existential $\exists x.Px$ is that there be a specific object which is known in s to have the property P.

The inquisitive nature of existential quantification makes it possible to express mention-some questions in the logical language. But, witnessing the status of inquisitive semantics as a general logical framework, it is equally possible to express mention-all questions: $\forall x.?Px$ is only supported in a state s if the full denotation of the predicate P is known, i.e., if all worlds in s agree on the denotation of P.

In effect, this means that two of the main rival theories of questions in natural language semantics, the Hamblin analysis [24] and the partition analysis [21], may both be formulated and compared within the logical framework of inquisitive semantics. At the same time, the framework provides the means to express certain types of questions, such as conditional questions, that are not, or at least not obviously, within the reach of either of these two theories.

2.4 The Boundedness Problem

All the basic properties of the propositional system having to do with informative and inquisitive content still hold in the first order setting. For instance, the classical treatment of informative content is still preserved (fact 2).

However, one feature of the system is not preserved: the proposition expressed by a sentence is no longer fully determined by the alternative set of that sentence (fact 4). In other words, it is no longer the case that every state supporting φ is contained in a maximal state supporting φ. In fact, as shown by Ciardelli [6,7], there are first-order formulas that do not have any maximal supporting states.

Example 4 (The boundedness formula). Consider a language which has a unary predicate symbol P, a binary function symbol $+$, and the set $\mathbb{N}$ of natural numbers as its individual constants. Consider the discourse-model $\mathbb{D} = \langle D, I \rangle$, where $D = \mathbb{N}$, I maps every $n \in \mathbb{N}$ to itself, and $+$ is interpreted as addition. Let $x \leq y$ abbreviate $\exists z(x + z = y)$, let $B(x)$ abbreviate $\forall y(P(y) \to y \leq x)$, and for every $n \in \mathbb{N}$, let $B(n)$ abbreviate $\forall y(P(y) \to y \leq n)$. Intuitively, $B(n)$ says that n is greater than or equal to any number in P. In other words, $B(n)$ says that n is an *upper bound* for P.

A $\mathbb{D}$-state s supports a formula $B(n)$, for some $n \in \mathbb{N}$, if and only if $B(n)$ is true in every world in s, that is, if and only if n is an upper bound for P in every w in s. Now consider the formula $\exists x.B(x)$, which intuitively says that there is an upper bound for P. This formula, which Ciardelli refers to as the *boundedness formula*, does not have a maximal supporting state. To see this, let s be an arbitrary state supporting $\exists x.B(x)$. Then there must be a number $n \in \mathbb{N}$ such that s supports $B(n)$, i.e., $B(n)$ must be true in all worlds in s. Now let w^* be the $\mathbb{D}$-world in which P denotes the singleton set $\{n + 1\}$. Then w^* cannot be in s, because it does not make $B(n)$ true. Thus, the state s^* which is obtained from s by adding w^* to it is a proper superset of s itself. However, s^* clearly supports $B(n + 1)$, and therefore also still supports $\exists x.B(x)$. This shows that any state supporting $\exists x.B(x)$ can be extended to a larger state which still supports $\exists x.B(x)$, and therefore no state supporting $\exists x.B(x)$ can be maximal.

This example shows that our notion of basic compliant responses, which makes crucial reference to maximal supporting states, does not always yield satisfactory results in the first-order setting. At first sight, it is tempting to conclude from this observation that there must be something wrong with the given notion of basic compliant responses. However, the problem is deeper than that. Namely, the following example, again from [6,7], shows that the very notion of meaning assumed in Inq_B is not fine-grained enough to serve as a basis for a suitable notion of compliance in the first-order setting.

Example 5 (The positive boundedness formula). Consider the following variant of the boundedness formula: $\exists x(x \neq 0 \wedge B(x))$. This formula says that there is a *positive* upper bound for P. Intuitively, it differs from the ordinary boundedness formula in that it does not license $B(0)$ as a compliant response. However, in terms of support, $\exists x(x \neq 0 \wedge B(x))$ and $\exists x.B(x)$ are equivalent. Thus, support is not fine-grained enough to capture the fact that these formulas intuitively do not have the same range of compliant responses.

3 An Inquisitive Witness Semantics

In this section we will develop a first-order inquisitive witness semantics, Inq_W, which explicitly reflects the idea that an existentially quantified sentence like $\exists x.Px$ is supported in a state if and only if there is a specific witness in that state which is known to have the property P.

This idea is not entirely new. For instance, when informally describing the clause for existential quantification in Inq_B, Ciardelli [7] writes that "an existential will only be supported in those states where a specific witness for the existential is known." However, in Inq_B, states merely encode a certain body of information. To know a witness for a certain property is simply to know that the property holds of a specific individual. To say that a sentence *introduces a witness*, then, is just to say that the sentence provides the information that a certain individual has a certain property. But notice that, on this notion of witnesses, (i) a sentence may introduce infinitely many witnesses and (ii) these witnesses need not even be mentioned explicitly by the sentence. To deal with compliance, we need a stricter notion of witnesses: only individuals that are explicitly mentioned in the conversation should count as such. So, we need to devise a system which keeps track of the mentioned individuals, alongside the information that has been provided about them.[3]

3.1 Witnesses, States, and Support

In developing such a system, the first question to ask is what our formal notion of witnesses should be. The simplest answer would be that witnesses are objects in the domain D. This is indeed sufficient for the simplest cases of existential quantification. For instance, it would be reasonable to think of a state s as supporting a sentence $\exists x.Px$ just in case there is a specific object $d \in D$ which is known in s to have the property P. However, this notion of witnesses as objects in D is not general enough. In particular, it becomes problematic when we consider formulas where an existential quantifier is embedded under a universal quantifier. For instance, it would not be appropriate to think of a state s as supporting a sentence $\forall x.\exists y.Rxy$ just in case there is a specific object $d \in D$ which is known in s to stand in the relation R with all other objects in D. Intuitively, this is not what $\forall x.\exists y.Rxy$ requires.

To avoid problems of this sort, we will take witnesses to be *functions* from D^n to D, where $n \geq 0$. Notice that some of these functions are 0-place functions into D, which can simply be identified with objects in D. So witnesses *can* still be objects in D. But they can be other things as well.

In the definitions below, we will assume a fixed first-order language $\mathcal{L}$ and a fixed discourse-model $\mathbb{D} = \langle D, I \rangle$ for $\mathcal{L}$.

Definition 16 (Witnesses)

- *For any $n \in \mathbb{N}$, let $D^\star_n$ be the set of functions $\delta\colon D^n \to D$.*
- *Then $D^\star = \bigcup_{n \geq 0} D^\star_n$ is the set of all witnesses based on D.*

[3] In the concluding section of the paper, we will see that to completely solve the problem under investigation, it is not sufficient to keep track of the mentioned individuals and the information provided about them, but that we should also keep track of which part of the information provided is about which individual. The system to be introduced below achieves the former, but not the latter.

The next step is to reconsider our notion of a state. Before, states were sets of worlds, reflecting a certain body of information. Now states will not only reflect a certain body of information, but also contain a set of witnesses. We will assume that the set of witnesses available in a state always includes the identity function. The rationale behind this assumption will become clear in a moment, when we define how witnesses are put to use in the semantics.

Definition 17 (States with Witnesses)

- *A $\mathbb{D}$-state is a pair $\langle V, \Delta \rangle$, where V is a set of $\mathbb{D}$-worlds and Δ is a finite set of witnesses based on D, which contains the identity function $id : D \to D$.*
- *The set of all $\mathbb{D}$-states is denoted by $\mathcal{S}_{\mathbb{D}}$.*
- *If $s = \langle V, \Delta \rangle$ is a $\mathbb{D}$-state, then $\mathsf{worlds}(s) := V$ and $\mathsf{witn}(s) := \Delta$.*

We will often drop reference to $\mathbb{D}$, and simply refer to $\mathbb{D}$-states as states. The set of all states is partially ordered by the following *extension* relation.

Definition 18 (Extension). *Let s and t be two states. Then we say that s is an extension of t, $s \geq t$, iff $\mathsf{worlds}(s) \subseteq \mathsf{worlds}(t)$ and $\mathsf{witn}(s) \supseteq \mathsf{witn}(t)$.*

Notice that there is a minimal state, namely $\mathsf{top} := (W, \{id\})$, of which any other state is an extension. The extension relation will be used in the support definition, in particular in the clause for implication: a state s supports an implication iff every extension of s that supports the antecedent, supports the consequent as well.

Before turning to the definition of support, however, we introduce two more auxiliary notions. The first is the notion of a *witness feed*. The role of these witness feeds will be similar to that of assignments: they will be used to store certain information in evaluating whether or not a certain formula is supported by a certain state. In particular, they play a role in evaluating existentially quantified formulas in the scope of one or more universal quantifiers. This will be further explained once we have specified the support relation.

Definition 19 (Witness Feeds). *A witness feed ε is a finite subset of D.*

Finally, we assume that the interpretation I of individual constants and function symbols in our discourse model $\mathbb{D}$ is extended in the following natural way to an interpretation of all terms $t \in \mathcal{L}$: if the free variables occurring in t are, orderly, $x_1, \ldots, x_n$, then $I(t)$ is the function $D^n \to D$ which maps a tuple $(d_1, \ldots, d_n) \in D^n$ to the element $d \in D$ denoted by the term t in $\mathbb{D}$ when x_i is interpreted as d_i for all $i = 1, \ldots, n$.

We now have all the necessary ingredients to state the support relation.

Definition 20 (Support in $\mathbf{Inq}_W$)
Let s be a $\mathbb{D}$-state, g an assignment, ε a witness feed, and φ a formula in $\mathcal{L}$.

$$s \models_{g,\varepsilon} R(t_1, \ldots, t_n) \quad \text{iff} \quad \begin{array}{l} \text{(i) } \mathsf{worlds}(s) \subseteq |R(t_1, \ldots, t_n)|_g \\ \text{(ii) } I(t_i) \in \mathsf{witn}(s) \text{ for } i = 1, \ldots, n \end{array}$$

$$
\begin{array}{llll}
s \models_{g,\varepsilon} & \bot & \textit{iff} & \mathsf{worlds}(s) = \emptyset \\
s \models_{g,\varepsilon} & \varphi \wedge \psi & \textit{iff} & s \models_{g,\varepsilon} \varphi \textit{ and } s \models_{g,\varepsilon} \psi \\
s \models_{g,\varepsilon} & \varphi \vee \psi & \textit{iff} & s \models_{g,\varepsilon} \varphi \textit{ or } s \models_{g,\varepsilon} \psi \\
s \models_{g,\varepsilon} & \varphi \to \psi & \textit{iff} & \forall t \geq s : \textit{if } t \models_{g,\varepsilon} \varphi \textit{ then } t \models_{g,\varepsilon} \psi
\end{array}
$$

$s \models_{g,\varepsilon} \forall x.\varphi$ *iff* $s \models_{g[x/d],\varepsilon\cup\{d\}} \varphi$ *for all* $d \in D$

$s \models_{g,\varepsilon} \exists x.\varphi$ *iff* $s \models_{g[x/\delta(e_1,\ldots,e_n)],\varepsilon} \varphi$ *for some* $\delta \in \mathsf{witn}(s)$ *and* $e_1, \ldots, e_n \in \varepsilon$

We will use $s \models_g \varphi$ as an abbreviation of $s \models_{g,\emptyset} \varphi$. The clauses that have changed w.r.t. Inq_B are those for atomic formulas, implication, universal quantification, and existential quantification. Let us look at these four clauses in some detail.

Atoms. For a state s to support an atomic sentence $R(t_1, \ldots, t_n)$, the sentence has to be true in all worlds in $\mathsf{worlds}(s)$, as before, but moreover, for every term t_i, the function $I(t_i)$ that it denotes must be available as a witness in $\mathsf{witn}(s)$. To illustrate this, consider the formula $R(a, f(b))$ where a and b are individual constants and f is a unary function symbol. Suppose $I(a) = d_1$ and $I(f(b)) = d_2$: then a state s supports the sentence $R(a, f(b))$ if and only if (i) for every $w \in \mathsf{worlds}(s)$ we have that $\langle d_1, d_2\rangle \in I_w(R)$, and (ii) d_1 and d_2 are available as witnesses in $\mathsf{witn}(s)$.

Recall that in uttering a sentence, a speaker proposes to update the common ground of the conversation in such a way that it comes to support the sentence. Thus, in particular, in uttering $R(a, f(b))$, a speaker proposes to add d_1 and d_2 to the witness set of the common ground. In this sense, we can think of atomic sentences like $R(a, f(b))$ as introducing new witnesses. We will see that other sentences, in particular existentials, may request a response that introduces new witnesses.

Implication. In order to determine whether a state s supports an implication $\varphi \to \psi$ we have to consider all extensions t of s that support φ. An extension t of s is a state such that $\mathsf{worlds}(t) \subseteq \mathsf{worlds}(s)$ and $\mathsf{witn}(t) \supseteq \mathsf{witn}(s)$. Thus, it may be that all the extensions of s that support φ contain certain witnesses that are not contained in s itself. This means that if ψ requires certain witnesses, as long as we need to introduce them to support φ, it is not necessary for s as such to already contain them for the implication to be supported in s.

To illustrate this, let us show that $\mathsf{top} \models_{gepsilon} Pa \to \exists x.Px$. Given the atomic clause, every $t \geq \mathsf{top}$ that supports Pa must be such that $I(a) \in \mathsf{witn}(t)$. In other words, every $t \geq \mathsf{top}$ that supports Pa contains a witness, namely $I(a)$, which is known to have the property P. It follows that $t \models_{gepsilon} \exists x.P(x)$, which in turn means that $\mathsf{top} \models_{gepsilon} Pa \to \exists x.Px$, even though top itself does not contain any witnesses besides the identity function.

Universal Quantification. The clause for universal quantification is very much like the clause we had in Inq_B. Only now the witness feed plays a role as well. In determining whether a state s supports a formula $\forall x.\varphi$ we do not only set the current assignment g to $g[x/d]$, but we simultaneously augment the current witness feed ε with the same object d. Then we check whether φ is supported by s relative to the adapted assignment and the augmented witness feed. As we will see below, the augmented witness feed is put to use when φ contains an existential quantifier.

Existential Quantification. In checking whether $s \models_{g,\varepsilon} \exists x.\varphi$ holds, we have to check whether $s \models_{g[x/d],\varepsilon} \varphi$ holds for some object $d \in D$ which is obtained by applying some witness $\delta \in \mathsf{witn}(s)$ to objects $e_1, \ldots, e_n$ in the witness feed. We call this element d a witness *for* the existential. This means that in uttering $\exists x.Px$, a speaker requests a response that introduces a suitable witness and establishes of this witness that it has the property P. The fact that the set of witnesses always contains the identity function id ensures that any element e of the witness feed can always be used *itself* as a witness for an existential, since e can be obtained by applying id to e. The invariable presence of the identity function in the witness set, required by our definition of states, is designed precisely to make the elements of the current witness feed directly available as witnesses for an existential.

Example 6 (Interaction between existentials and universals). Consider the sentence $\forall x.\exists y.Rxy$. In order to determine whether $s \models_g \forall x.\exists y.Rxy$, we have to check whether $s \models_{g[x/d],\{d\}} \exists y.Rxy$ for all $d \in D$. And this means that we have to verify whether for every $d \in D$, there is a witness $f \in \mathsf{witn}(s)$ such that $s \models_{g[x/d][y/f(d,\ldots,d)],\{d\}} Rxy$. This witness f may be an element of the domain, a unary function, or a function of higher arity. It may also be the identity function, which, as we saw, means that the element d itself may serve as a witness for the existential. This, then, is how universal and existential quantifiers interact: universal quantifiers add objects to the current witness feed, and these objects then may serve as input for functional witnesses that may be needed for existentials in the scope of a universal. In this way, the witness that is required for the embedded existential in $\forall x.\exists y.Rxy$ may functionally depend on the value of x under the current assignment. We will return to this example in section 3.5, where we illustrate the relevance of Inq_W for natural language semantics.

As in Inq_B, support is *persistent*. That is, if a state s supports a formula φ relative to a certain assignment g and a certain witness feed ε, then any extension of s also supports φ relative to g and ε.

Fact 10 (Persistence). *If* $s \models_{g,\varepsilon} \varphi$ *and* $t \geq s$, *then* $t \models_{g,\varepsilon} \varphi$

Also as in Inq_B, we take $\neg\varphi$ to be an abbreviation of $\varphi \to \bot$, and $!\varphi$ an abbreviation of $\neg\neg\varphi$. The derived clauses for $\neg\varphi$ and $!\varphi$ read as follows.

Fact 11 (Support for Negation)

- $s \models_{g,\varepsilon} \neg\varphi$ *iff for all* $w \in \mathsf{worlds}(s)$: $w \not\models_g \varphi$ *classically*
- $s \models_{g,\varepsilon} !\varphi$ *iff for all* $w \in \mathsf{worlds}(s)$: $w \models_g \varphi$ *classically*

3.2 Propositions, Entailment, and Equivalence

Based on the notion of support, we define the proposition expressed by a formula, and the notions of entailment and equivalence, just as in Inq_B. Recall that our definitions assume a fixed first-order language $\mathcal{L}$ and a fixed discourse-model $\mathbb{D} = \langle D, I \rangle$ for $\mathcal{L}$.

Definition 21 (Propositions, Entailment, and Equivalence)

1. $[\varphi]_g \;:=\; \{s \in \mathcal{S}_{\mathbb{D}} \mid s \models_g \varphi\}$
2. $\varphi \models \psi$ *iff for all* $\mathbb{D}$, s *and* g*: if* $s \models_g \varphi$, *then* $s \models_g \psi$
3. $\varphi \equiv \psi$ *iff* $\varphi \models \psi$ *and* $\psi \models \varphi$

In Inq_B, states were sets of possible worlds, ordered by inclusion, and we referred to *maximal* states supporting φ as *alternatives* for φ, where maximality was determined by the inclusion-order. Thus, alternatives for φ in Inq_B were *minimally informed* states supporting φ. In Inq_W, states are ordered by the extension relation, $\geq$, and alternatives for φ will be defined as $\geq$-minimal states supporting φ. Thus, in Inq_W alternatives for φ are states that support φ with a minimum amount of information and a minimal set of witnesses.[4]

Definition 22 (Alternatives). *Let* φ *be a formula and* g *an assignment.*

1. *Every* $\geq$*-minimal element of* $[\varphi]_g$ *is called an* alternative *for* φ *relative to* g*.*
2. *The* alternative set *of* φ *relative to* g, $[\![\varphi]\!]_g$, *is the set of alternatives for* φ *relative to* g*.*

We also introduce notions of *factive* support, entailment, and equivalence, which ignore witness issues.

Definition 23 (Factive Support, Entailment, and Equivalence)

1. $V \models_g^\star \varphi$ *iff there is a state* s *with* $\mathsf{worlds}(s) = V$ *such that* $s \models_g \varphi$
2. $\varphi \models^\star \psi$ *iff for all* V, g*: if* $V \models_g^\star \varphi$, *then* $V \models_g^\star \psi$
3. $\varphi \equiv^\star \psi$ *iff* $\varphi \models^\star \psi$ *and* $\psi \models^\star \varphi$

Witness sensitivity was introduced in order to be able to discriminate sentences that differ in the responses they license. Factive notions disable this sensitivity, thus taking into account only inquisitive and informative content of sentences. Not surprisingly, then, the system that we obtain by disregarding witness issues in Inq_W is precisely our good old Inq_B.

Fact 12 (Factive Support and Support in Inq_B)

$$V \models_g^\star \varphi \text{ in } \mathsf{Inq}_W \iff V \models_g \varphi \text{ in } \mathsf{Inq}_B$$

[4] We do not know at this point whether the equivalent of fact 4 holds for Inq_W, that is, whether any state that supports a formula φ relative to an assignment g must be an extension of some alternative for φ relative to g.

Clearly, this also means that factive entailment and equivalence in Inq_W amount to entailment and equivalence in Inq_B. We say that a formula is witness-insensitive in case it is supported by a state as soon as it is factively supported by the information available in that state.

Definition 24 (Witness Insensitivity)

φ is witness insensitive iff for all s, g: if $\mathit{worlds}(s) \models_g^\star \varphi$, then $s \models_g \varphi$

Fact 13 (Partial Characterization of Witness Insensitivity)

1. *An atomic formula is witness insensitive iff it does not contain any individual constant or a function symbol;*
2. *$\bot$ is witness insensitive;*
3. *If φ and ψ are witness insensitive, so are $\varphi \vee \psi$ and $\varphi \wedge \psi$;*
4. *If ψ is witness insensitive, so is $\varphi \to \psi$;*
5. *$\exists x.\varphi$ is not witness insensitive for any φ;*
6. *$\forall x.\varphi$ is witness insensitive iff φ is witness insensitive.*

Given that negation $\neg\varphi$ is defined as $\varphi \to \bot$, and non-inquisitive projection $!\varphi$ as $\neg\neg\varphi$, item 2 and 4 above guarantee that negation and non-inquisitive projection block witness sensitivity of their complement.

3.3 Informativeness and Inquisitiveness

As before, we define the informative content of a sentence φ relative to an assignment g as the set of worlds that are contained in at least one state that supports φ relative to g.

Definition 25 (Informative Content). $\mathit{info}_g(\varphi) := \bigcup\{\mathit{worlds}(s) \mid s \in [\varphi]_g\}$.

Also as before, the informative content of a sentence φ relative to an assignment g always coincides with the *truth set* of φ relative to g, $|\varphi|_g$, i.e., the set of worlds that satisfy φ in classical first-order logic relative to g. So as far as informative content is concerned, Inq_W does not diverge from classical first-order logic.

Fact 14 (Informative Content is Classical). *For any φ, g: $\mathit{info}_g(\varphi) = |\varphi|_g$*

In terms of the informative content of a formula, we define whether it is informative and/or inquisitive.

Definition 26 (Inquisitiveness and Informativeness in a State)

– φ *is* informative in s w.r.t. g *iff* $\mathit{worlds}(s) \cap \mathit{info}_g(\varphi) \neq \mathit{worlds}(s)$
– φ *is* inquisitive in s w.r.t. g *iff* $\mathit{worlds}(s) \cap \mathit{info}_g(\varphi) \not\models_g^\star \varphi$

As before, we also define absolute notions of informativeness and inquisitiveness.

Definition 27 (Absolute Inquisitiveness and Informativeness)

– φ *is* informative *iff for some g:* $\mathit{info}_g(\varphi) \neq W$
– φ *is* inquisitive *iff for some g:* $\mathit{info}_g(\varphi) \not\models_g^\star \varphi$

The following fact reports that the notions of informativeness and inquisitiveness in Inq_W correspond exactly with those in Inq_B.

Fact 15 (Informativeness and Inquisitiveneness in Inq_W and Inq_B)

- *φ is informative in Inq_W iff φ is informative in Inq_B*
- *φ is inquisitive in Inq_W iff φ is inquisitive in Inq_B*

All notions in Inq_B that are defined in terms of informativeness and inquisitiveness, such as the notions of assertions, questions, and hybrids, remain precisely the same in intension and extension. Only, now within each class there is a further distinction between witness sensitive and witness insensitive formulas.

Compliance. The notion of basic compliant responses that we had in Inq_B carries over straightforwardly to Inq_W. Recall that in Inq_B, the basic compliant responses to a sentence φ were characterized as those responses that provide precisely enough information to establish a state that supports φ, without raising any new issues. In Inq_W, states do not only contain information but also witnesses, and support sometimes requires the presence of such witnesses. Thus, in Inq_W the basic compliant responses to φ are naturally characterized as those responses that provide precisely enough information and precisely enough witnesses to establish a state that supports φ, without raising any new issues. This is captured by our earlier definition of basic compliant responses, definition 11, provided that the notion of entailment from Inq_B is replaced by the notion of entailment from Inq_W.

3.4 The Boundedness Problem Resolved

Now that we have discussed some of the basic logical properties of Inq_W, let us return to the problem that we set out to resolve. The boundedness formula was problematic for Inq_B in two crucial ways: first, it provided an example of a formula to which we could associate no basic compliant responses; second, it was semantically equivalent to its positive variant which, intuitively, licenses a different range of compliant responses. Let us start by observing that the latter problem no longer arises in Inq_W: thanks to the witness machinery, Inq_W is fine-grained enough to detect the differences between the two boundedness formulas.

Fact 16 (The Boundedness Formulas)
The boundedness formula and the positive boundedness formula are not equivalent in Inq_W.

Proof. Consider a state s such that:

- $\mathsf{worlds}(s) = \{w\}$, where $I_w(P) = \{0\}$
- $\mathsf{witn}(s) = \{id, 0\}$

This state factively supports both $\exists x.B(x)$ and $\exists x.(x > 0 \wedge B(x))$. However, while the boundedness formula is supported in s *tout court*, $s \models \exists x.B(x)$, the positive boundedness formula is not, $s \not\models \exists x.(x > 0 \wedge B(x))$. So, the boundedness formula and the positive boundedness formula are not equivalent in Inq_W (although they are *factively* equivalent). □

Now that we ensured that the two boundedness formulas can be distinguished, we would like our semantics to predict for each of them its expected range of basic compliant responses. The following fact shows that Inq_W achieves this as well.

Fact 17 (Basic Compliant Responses to the Boundedness Formulas)

- *For any* $n \geq 0$, $B(n)$ *is a basic compliant response to* $\exists x.Bx$
- *For any* $n > 0$, $B(n)$ *is a basic compliant response to* $\exists x.(x \neq 0 \wedge Bx)$, *but* $B(0)$ *is not a basic compliant response to* $\exists x.(x \neq 0 \wedge Bx)$.

How can Inq_W get this prediction right? Let us examine this more closely. In Inq_B, an assertion entails another simply in case the informative content of the former entails the informative content of the latter. Now, for any number n, the informative content of $B(n)$ entails the informative content of $B(n+1)$. Therefore, $B(n) \models B(n+1)$ in Inq_B. Thus, we have an infinite chain $B(0), B(1), B(2), \ldots$ of weaker and weaker responses, all of which resolve the issue raised by $\exists x B(x)$. None of these responses is minimally informative, and so, none is predicted to be a basic compliant response.

In Inq_W, on the other hand, entailment is more demanding. For an assertion ψ to entail another assertion χ, entailment of informative content is no longer sufficient: it should also be the case that ψ introduces all the witness that χ introduces. This prevents $B(n)$ from entailing $B(n+1)$, since, unlike the latter, the former does not introduce the witness $n+1$. Thus, from the point of view of Inq_W, every assertion $B(n)$ constitutes a *minimal* way to resolve the issue raised by $\exists x.Bx$, where the minimality concerns now not only the information provided, but also the witnesses introduced.

Notice that, by means of factive entailment, Inq_W is still capable of accounting for the fact that the informative content of one compliant response may entail the informative content of another. If this happens, the stronger response will be quantitatively preferred pragmatically over the weaker one. For instance, $B(1)$ and $B(135)$ are both basic compliant responses to $\exists x.Bx$. However, $B(1)$ factively entails $B(135)$. If the information state of the responder supports $B(1)$, then it would be misleading for her to actually choose $B(135)$ as a response. In general, if ψ and χ are two basic compliant responses to φ, and ψ factively entails χ, then ψ is preferred over χ as a response to φ.

Definition 28 (Comparing Basic Compliant Responses)
Let φ *be an inquisitive initiative, let* ψ *and* χ *be two basic compliant responses to* φ, *and let* σ *be an information state, i.e., a set of worlds. Then:*

1. ψ *is* preferred *over* χ *as a response to* φ *iff* $\psi \models^{\star} \chi$ *and* $\chi \not\models^{\star} \psi$.
2. ψ *is an* optimal response *to* φ *in* σ *iff*
   - $\sigma \subseteq \mathit{info}(\psi)$, *and*

– *for every basic compliant response ξ to φ that is preferred over ψ, $\sigma \not\subseteq \mathsf{info}(\xi)$.*

To illustrate the notion of an optimal response, consider an information state consisting of three worlds, one where the highest element of P is 5, one where it is 14, and one where it is 3. The optimal response to $\exists x.Bx$ in this information state is $B(14)$. This accounts for the intuition that, on the one hand, any response $B(n)$ with $n < 14$, even though compliant, would be *qualitatively* inappropriate, while any response $B(n)$ with $n > 14$ would be *quantitively* dispreferred. The only optimal response in this scenario is $B(14)$.

3.5 An Example from Natural Language

In this section we will briefly illustrate the potential of the proposed notion of compliance by means of an example from natural language. Consider a quantified question like *Who does every man like?*. As has been discussed widely in the literature (e.g., [5,13,21]), such questions allow for different types of responses, e.g., *Mary*, *himself*, or *his mother*. If this question is formally represented as $\forall x.\exists y.Rxy$ these different types of responses are accounted for in a uniform way. We illustrate this below for Inq_W. However, since boundedness issues do not play a role in this particular example, the same results are also obtained in Inq_B.

Consider what is needed for a state s to support $\forall x.\exists y.Rxy$. If we assume that $\mathsf{witn}(s)$ does not contain any witnesses, apart from the identity function, which is always an element of $\mathsf{witn}(s)$, then we must have that $\langle d, d\rangle \in I_w(R)$ for every $d \in D$ and every $w \in \mathsf{worlds}(s)$. The $\geq$-minimal state that satisfies this condition is one of the alternatives for $\forall x.\exists y.Rxy$. It is also the unique alternative for the response *himself* ($\forall x.Rxx$). Thus, this is a basic compliant response.

Now consider a state s such that $\mathsf{witn}(s)$ contains an object m, and such that $\langle d, m\rangle \in I_w(R)$ for every $d \in D$ and every $w \in \mathsf{worlds}(s)$. The $\geq$-minimal state that satisfies these conditions is another alternative for $\forall x.\exists y.Rxy$. It is also the unique alternative for the response *Mary* ($\forall x.Rxm$). Thus, this is another basic compliant response.

Finally, consider a state s such that $\mathsf{witn}(s)$ contains a 1-place function f which maps every individual in D to his mother, and such that $\langle d, f(d)\rangle \in I_w(R)$ for every $d \in D$ and every $w \in \mathsf{worlds}(s)$. The $\geq$-minimal state that satisfies these conditions is again one of the alternatives for $\forall x.\exists y.Rxy$. It is also the unique alternative for the response *his mother* ($\forall x.R(x, f(x))$). Thus, this is yet another basic compliant response.

4 Conclusions

In this paper, we addressed a problem that arises when the notion of compliance introduced in [20] is extended from a propositional setting to a first-order setting. This notion of compliance is formulated within the basic inquisitive semantic system, Inq_B. We argued that the problem does not lie in the particular way the notion of compliance is defined, but rather in the fact that the system Inq_B itself is not fine-grained enough to capture compliance conditions, since sentences with intuitively different sets of compliant responses are assigned the same semantic value. As a consequence, no satisfactory notion of compliance can be defined based on Inq_B.

To fix this problem we developed a more sophisticated semantics, Inq_W, in which the conversational context is taken to include not only a certain body of information, but also a certain set of witnesses. Along with information, sentences may then request or provide certain witnesses. This semantic refinement allows us to assign a different range of compliant responses to sentences having the same informative and inquisitive content, but requesting different sorts of witnesses. This extra semantic sensitivity allowed us to correctly characterize the set of basic compliant responses for those examples that were problematic for Inq_B.

But how general is our solution? Is the system Inq_W rich enough to account for *all* semantic distinctions that are relevant to determine compliant responses? Unfortunately, not quite. In fact, just before submitting this paper, we became aware of a problematic case. Consider the setting of the earlier boundedness formulas, but now suppose that our language has two unary predicates, P and Q. Let $B_P(x) := \forall y(P(y) \to y \leq x)$ and let $B_Q(x) := \forall y(Q(y) \to y \leq x)$. Consider the following two sentences:

1. $\exists x B_P(x) \wedge \exists x B_Q(x)$
2. $\exists x(B_P(x) \wedge B_Q(x))$

Intuitively, these two sentences have different basic compliant responses. Any sentence of the form $B_P(n) \wedge B_Q(m)$, with n and m two natural numbers, should count as a basic compliant response to the former, while only sentences of the form $B_P(n) \wedge B_Q(n)$ should count as basic compliant responses to the latter. However, the two come out exactly equivalent, not only in Inq_B, but also in Inq_W. It can be shown that the only basic compliant responses predicted by our semantics are those of the form $B_P(n) \wedge B_Q(n)$. What is going on here is that the witnesses for one formula interfere with the witnesses for another in a way they are not meant to. Inq_W is still too coarse, in that it pools all witnesses together in one witness set. What we need, it seems, is a system that treats witnesses in a more refined way, keeping track of the properties or relations that they are witnesses for. Evidently, the development of such a further refinement is a task for future work.

References

1. Asher, N., Lascarides, A.: Logics of Conversation. Cambridge University Press (2003)
2. Belnap, N.: Questions, their presuppositions, and how they can arise. In: Lambert, K. (ed.) The Logical Way of Doing Things. Yale University Press (1969)
3. Büring, D.: On d-trees, beans, and accents. Linguistics and Philosophy 26, 511–545 (2003)

4. Carlson, L.: Dialogue Games. Reidel, Dordrecht (1983)
5. Chierchia, G.: Questions with quantifiers. Natural Language Semantics 1(2), 181–234 (1993)
6. Ciardelli, I.: Inquisitive semantics and intermediate logics, Master Thesis, University of Amsterdam (2009)
7. Ciardelli, I.: A First-Order Inquisitive Semantics. In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (eds.) Logic, Language and Meaning. LNCS, vol. 6042, pp. 234–243. Springer, Heidelberg (2010)
8. Ciardelli, I., Cornelisse, I., Groenendijk, J., Roelofsen, F.: Computing Compliance. In: He, X., Horty, J., Pacuit, E. (eds.) LORI 2009. LNCS, vol. 5834, pp. 55–65. Springer, Heidelberg (2009)
9. Ciardelli, I., Groenendijk, J., Roelofsen, F.: Inquisitive semantics: a new notion of meaning (2012); forthcoming in Language and Linguistics Compass
10. Ciardelli, I., Groenendijk, J., Roelofsen, F.: Inquisitive semantics. NASSLLI Lecture Notes (2012), `http://www.illc.uva.nl/inquisitive-semantics`
11. Ciardelli, I., Roelofsen, F.: Inquisitive logic. Journal of Philosophical Logic 40(1), 55–94 (2011)
12. Clark, H.: Arenas of Language Use. University of Chicago Press, Chicago (1992)
13. Engdahl, E.: Constituent questions: The syntax and semantics of questions with special reference to Swedish. Reidel, Dordrecht (1986)
14. Farkas, D., Bruce, K.: On reacting to assertions and polar questions. Journal of Semantics 27, 81–118 (2010)
15. Farkas, D., Roelofsen, F.: Polar initiatives and polarity particle responses in an inquisitive discourse model. Presented at the Amsterdam Colloquium, December 2011, and at the Newcastle Workshop on the Syntax of Answers to Polar Questions (June 2012), `www.illc.uva.nl/inquisitive-semantics`
16. Ginzburg, J.: Dynamics and the semantics of dialogue. In: Seligman, J. (ed.) Language, Logic, and Computation, vol. 1 (1996)
17. Groenendijk, J.: The logic of interrogation. In: Matthews, T., Strolovitch, D. (eds.) Semantics and Linguistic Theory, pp. 109–126. Cornell University Press (1999)
18. Groenendijk, J.: Inquisitive Semantics: Two Possibilities for Disjunction. In: Bosch, P., Gabelaia, D., Lang, J. (eds.) TbiLLC 2007. LNCS, vol. 5422, pp. 80–94. Springer, Heidelberg (2009)
19. Groenendijk, J.: Erotetic languages and the inquisitive hierarchy (2011), in a Festschrift for Martin Stokhof, `www.vddoes.net/Martin/articles/Jeroen.pdf`
20. Groenendijk, J., Roelofsen, F.: Inquisitive semantics and pragmatics. Presented at the Workshop on Language, Communication, and Rational Agency at Stanford (May 2009), `www.illc.uva.nl/inquisitive-semantics`
21. Groenendijk, J., Stokhof, M.: Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, University of Amsterdam (1984)
22. Gunlogson, C.: True to Form: Rising and Falling Declaratives as Questions in English. Ph.D. thesis, University of California at Santa Cruz (2001)
23. Hamblin, C.I.: Mathematical models of dialogue. Theoria 37, 130–155 (1971)
24. Hamblin, C.L.: Questions in Montague English. Foundations of Language 10, 41–53 (1973)
25. Harrah, D.: The logic of questions. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosohical Logic, 2nd edn., vol. 8, pp. 1–60. Kluwer (2002)
26. Hulstijn, J.: Structured information states: raising and resolving issues. In: Benz, A., Jäger, G. (eds.) Proceedings of MunDial (1997)
27. Jäger, G.: Only updates: on the dynamics of the focus particle *only*. In: Dekker, P., Stokhof, M. (eds.) Proceedings of the 10th Amsterdam Colloquium (1996)

28. Karttunen, L.: Syntax and semantics of questions. Linguistics and Philosophy 1, 3–44 (1977)
29. Kubiński, T.: An outline of the logical theory of questions. Akademie Verlag (1980)
30. Mascarenhas, S.: Inquisitive semantics and logic, Master Thesis, University of Amsterdam (2009)
31. Roberts, C.: Information structure in discourse. In: Yoon, J., Kathol, A. (eds.) OSU Working Papers in Linguistics, vol. 49, pp. 91–136. Ohio State University (1996)
32. Roelofsen, F.: Algebraic Foundations for Inquisitive Semantics. In: van Ditmarsch, H., Lang, J., Ju, S. (eds.) LORI 2011. LNCS, vol. 6953, pp. 233–243. Springer, Heidelberg (2011)
33. Stalnaker, R.: Assertion. Syntax and Semantics 9, 315–332 (1978)
34. Wiśniewski, A.: The logic of questions as a theory of erotetic arguments. Synthese 109(1), 1–25 (1996)

Sitting, Standing, and Lying in Frames: A Frame-Based Approach to Posture Verbs

Thomas Gamerschlag, Wiebke Petersen, and Liane Ströbel

University of Düsseldorf
{gamer,petersen,stroebel}@phil.uni-duesseldorf.de

Abstract. Posture verbs which allow for an extended locative use, such as *sit*, *stand* and *lie*, make reference to specific parts of the localized object, to the orientation of prominent object axes and to positional information, which are perceived by means of cognitive modules such as gestalt recognition and spatial perception. These properties render posture verbs an excellent object for the investigation of cognition and language. This paper analyzes the three basic posture verbs of German (*sitzen* 'sit', *stehen* 'stand' and *liegen* 'lie') in terms of frame representations. It turns out that frames can serve as a highly flexible device for decompositional analyses that is at the same time a cognitively plausible knowledge representation format.

Keywords: posture verbs, extended locative use, frame analysis, object schemata, German, Korean, French.

1 Introduction

Posture verbs such as *sit*, *stand*, and *lie* basically denote particular postures of individuals. According to [1] virtually all languages have posture verbs and, in addition, often exhibit extended locative uses. For example, the English verb *sit* in (1) refers to the posture of an individual resting on the buttocks and also allows for specifying the location of this individual by means of a locative PP.

(1) *John is sitting on the swing.*

Posture verbs (henceforth PVs) with a locative extension cannot be analyzed in isolation, but need to be treated in the context of other locative expressions such as the locative PPs that figure as their complements. By consequence, any analysis of PVs has to be considered as part of an overall approach to the relation between space and language which aims at an understanding of how human language expressions make reference to space and location.

The last few decades have seen a considerable increase in the amount of studies devoted to this topic. Given the restrictions of this paper, we cannot summarize, let alone review, all the qualitative work that has been done in this area. From a cognitive perspective, the general approaches by [17] and [34] have been particularly influential. In addition, there are numerous comprehensive anthologies

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 73–93, 2013.

such as [7, 10, 13, 23], to name just a few. There are also numerous works on the spatial meaning of particular parts of speech, such as spatial prepositions ([2, 27, 36–39]), dimensional adjectives referring to the spatial properties of objects such as *wide* and *long* ([18–20]), and locative verbs ([1, 14, 26]), which comprise verbs such as *hang (at)* and *stick (to)* in addition to posture verbs.

The typological branch of the research area, one important exponent of which is the Language and Cognition Group at the Max Planck Institute for Psycholinguistics, has revealed that languages differ significantly with respect to their spatial reference systems ([22, 23]). According to Ameka and Levinson ([1]), this diversity is in conflict with Landau and Jackendoff's assumption ([17]) that spatial language is of a rather schematic nature which abstracts away from properties such as object shape and is mainly carried by prepositions as in English. Ameka and Levinson argue that languages with a large inventory of locative verbs, in particular, are problematic in this respect since they have a full set of contrastive locative verbs which often make specific reference to properties of the figure and the ground, such as the number of axes, the presence of a canonical orientation, and distinctions such as natural vs. cultural kind, flexible vs. rigid, tall vs. stout, and container vs. flat surface.

Any formal representation must be able to cope with the cross-linguistic diversity of spatial language. In this paper, we will show that frame representations in the sense of [3, 28] are ideal for this purpose as they provide us with a highly flexible device while at the same time being a cognitively plausible, variable-free representation format.

After a short introduction to our framework in the next section, we will apply the frame model to the three basic German PVs *sitzen* 'sit', *stehen* 'stand', and *liegen* 'lie' in section 3. Given the wide range of languages which have been investigated for their posture verb repertories by the Nijmegen Language and Cognition Group and others, this may seem rather unspectacular. However, there are two reasons for our choice of German as an object language. First, German is an instance of a language which uses a comparatively large set of about ten verbs in basic locative constructions ([16]). This makes German a good starting point for exploring the potential of a frame analysis that can then be extended to languages with larger inventories of locative verbs. Second, there are already a number of investigations of German locative verbs and the subclass of PVs on which we can build the frame approach ([4, 14, 16, 32] among others). In particular, we will take the decompositional approach by [14] as a basis for the frame representation of PVs and spatial prepositions. After the exemplary frame analysis of the three basic German PVs, we will outline some possible extensions of the frame approach in section 4.

2 The Framework

In our analysis of PVs, we will apply frame representations made up of recursive attribute-value structures. The introduction of frames as a cognitively plausible format of knowledge representation has led to a paradigm shift in cognitive science, artificial intelligence and other disciplines ([11, 25]), such that concepts are no longer represented as atomic units but as complex structures built up recursively of attributes with structured values. Feature lists and binary features represent a preliminary stage in this process (cf. [8]).

Our frame approach mainly follows [3] in that we claim that the values of an attribute in a frame may be arbitrarily complex frames themselves and that the attributes in a frame can exhibit a cyclic structure. Furthermore, we add two assumptions which are not explicitly found in [3]: first, we assume that attributes in frames are functional in the sense that they assign unique values. Second, we do not claim that the central frame node is necessarily a root of the frame graph (i.e., a node from which all other nodes can be reached via directed arcs). Frames can be represented by directed, labeled graphs with arcs corresponding to the attributes and nodes corresponding to the values (for details see [28]). Figure 1 shows the graphs of simplified frames for the concepts *rented apartment* and *sibling*.

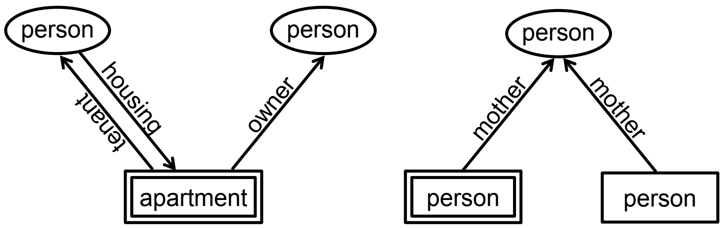


Fig. 1. Frame graphs of *rented apartment* (left) and *sibling* (right)

In the graphs, the labeled arcs represent the frame attributes while the node labels indicate the type of the attribute values. For example, in the frame graph of the concept *rented apartment*, the concept is modeled as an apartment which is further specified by two attributes, namely OWNER and TENANT. Both attributes have values of the type person. By typing frames and assuming a type signature, the class of admissible frames can be restricted (cf. [28]). In a type signature, the types are ordered in a hierarchy which is enriched by appropriateness conditions which constrain the domain and range of attributes. Thus the type signature determines which type of entities can have a certain attribute and of which type the values of each attribute are. In the frame graphs, the assumption that frame attributes are functional is modeled by the graph condition that a node may not have two outgoing arcs labeled with the same attribute. The central node of a frame, here the node labeled 'apartment', is marked by a double border. It indicates that this frame is a frame about an apartment rented to somebody and not about somebody renting or renting out an apartment. Note that the special notational treatment of the central node is necessary as it may not be a root or the only root of the frame graph ([30]). The frame graph on the left in Figure 1 has two roots while the one on the right has no root. Nodes corresponding to concept arguments are given a rectangular shape. As we treat nominal sortal concepts like *apartment* as one-place predicates, the central nodes of their frame graphs are argument nodes. The frame graph on the right in Figure 1 represents the concept *sibling* as a person for which a second person exists with which the first shares its mother. As *sibling* is a relational concept, its frame graph exhibits two argument nodes.

In our decompositional approach to PVs, we have decided to apply frames and not a formalism based on predicate logic as is commonly done since we consider frames to be cognitively more adequate. Frame theories have always been motivated cognitively: [3] argues that frames are used as a general format in accounting for the content of mental concepts and gives empirical evidence for attribute-value sets in cognition. [24] provides evidence for frames from a linguistic perspective. In [31] a biologically motivated model for the cortical implementation of frames is developed by applying the paradigm of object-related neural synchronization. Recently, [35] have shown that the attribute-value structure of frames provides an adequate formalization of the theory of grounded cognition. Moreover, the frame approach has already been successfully applied in the analysis of the inferential use of perception verbs such as *sound* and *feel (like)* ([12, 29]).

It is evident that the information represented in a frame graph can also be expressed in predicate logic. However, if one compares both approaches, there are advantages of frames which result from their variable-freeness: in contrast to predicates in predicate logic, which fix the number of arguments they take and their order, frames are more flexible. Additional attributes can be added and substructures can be addressed via labeled symbols instead of ordered argument positions. The main advantage is that in a decompositional frame analysis the unity of a concept is always preserved while in an analysis in predicate logic the elements constituting a concept can be scattered around only being connected by shared variables. We argue that the confinement to recursive attribute-value structures with attributes as basic elements will lead not only to more explicitness but also to a cognitively more plausible, variable-free analysis of PVs.

3 A Frame Analysis of German Posture Verbs

The properties that are relevant to the choice of a specific PV in German were established, among others, by [4, 14, 16, 32]. These properties include (i) the way the localized object is kept in its position (e.g., support from below in the case of *sitzen* 'sit' and support from above in the case of *hängen* 'hang'), (ii) the state of matter of the supporting medium (e.g., *schwimmen auf* 'be afloat on' versus *liegen auf* 'lie on') and (iii) the orientation of the most prominent object axis (e.g., *die Leiter steht* 'the ladder is standing' versus *die Leiter liegt* 'the ladder is lying'). [14] proposes an analysis in which these properties are explicitly implemented as conjuncts in predicate logic representations. Following [18], she assumes that part of the spatial requirements that are imposed by the PV on the localized object is captured in object schemata.

Our account of PVs builds heavily upon the analyses proposed in [14, 18–20]. In particular, we adopt two important ingredients of their approaches: the support relation and object schemata. The support relation captures the fact that PVs require the located object to be supported somehow in order to remain in its position. As will be shown below, PVs differ with respect to which part of the located object is supported. Object schemata are representations of the spatial knowledge of objects. They consist of a hierarchy of object axes which is determined by their saliency. Additionally, object schemata allow for further characterization of object axes, such as identifying the so-called ‘canonical vertical’, which is the axis that is aligned with the vertical if the object is in its prototypical spatial configuration. In the following, we present an analysis of the three German PVs *sitzen* ‘sit’, *stehen* ‘stand’, and *liegen* ‘lie’ in which the support relation and object schemata are directly translated into frame representations.

3.1 ***Sitzen*** **‘sit’**

The German PV *sitzen* ‘sit’ basically refers to the posture of an individual resting on the buttocks. Like English *sit*, *sitzen* allows for specifying the location of the sitting individual by means of a locative PP as in (2).

(2) *Hans sitzt auf der Schaukel.*
Hans sits on the swing
‘Hans is sitting on the swing.’

Kaufmann ([14, p. 103]) proposes the representation of *sitzen* in (3), which is formulated within the framework of Two-Level Semantics ([5, 6] among others).

(3) a. *sitzen* ‘sit’: $\lambda P \lambda x$ [SIT(x) & P(x)]
b. Int(SIT(x))= $\exists y$[support$_S$(d-us(y), buttocks(x))]

In (3a) the representation of *sitzen* at the level of semantic form is given. Semantic form is intended to be a minimal decomposition which is restricted to aspects of meaning relevant to grammar, in particular to argument realization. The representation in (3a) simply states that *sitzen* is translated into a conjunction of a one-place predicate SIT(x) and an additional predicate P(x) which is to be instantiated by the predicate contributed by the locative PP. The interpretation of SIT(x) at the level of conceptual structure is provided in (3b). In contrast to semantic form, conceptual structure is an elaborate semantic level, which can be made more fine-grained in any direction that matters. The representation in (3b) says that the predicate SIT is interpreted (‘Int’) as a relation between the figure x and some supporting entity y such that the deictic upper side (‘d-us’) of y supports the buttocks of x. In addition, the physical state of the supporter must be solid, which is indicated by the subscript ‘s’ for ‘solid’.

According to Kaufmann, the support relation is central to the interpretation of *sitzen*. As a consequence, the characteristic form or shape of a sitting person is rather epiphenomenal, resulting from the posture which has to be adopted in order for the buttocks to be supported from below. This view is corroborated by the fact that the verb *sitzen* cannot be applied directly to a person who has a sitting posture but is kept in this position by the support of a body part different from the buttocks. In German, one could refer to this by the complex construction *in einer sitzenden Haltung sein* 'be in a sitting posture', but *sitzen* as a matrix verb cannot be predicated directly of an individual in such a spatial configuration. Additionally, as the example in (2) demonstrates, a sitting person does not necessarily need to adopt a prototypical sitting posture. Imagine a child on a swing putting a lot of effort into swinging. Although it remains sitting on the swing, the shape of its body will nearly never correspond to a prototypical sitting posture.

Since our focus is on the conceptual properties of PVs, the frame representation of *sitzen* in Figure 2 below is based on the conceptual structure representation in (3b). The central node of the frame, which is marked by a double border, refers to the overall situation denoted by *sitzen*. The sitting individual is introduced as the value of the THEME attribute.

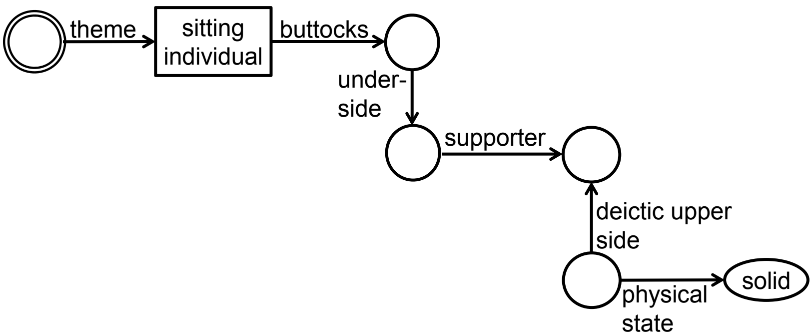


Fig. 2. Frame for *sitzen* 'sit'

In line with Kaufmann's analysis in (3b), the requirement that a specific part of the body is supported is integrated by the mereological attribute BUTTOCKS. The underside of the buttocks is the supported part of the body, i.e., it has a supporter which is the deictic upper side of the supporting entity. Moreover, the supporting entity has to be solid, which is implemented into the representation by means of the frame attribute PHYSICAL STATE and the value 'solid'.

Note that the complex frame for *sitzen* in Figure 2 is yielded by expanding the node *sitting individual* into several attribute-value pairs. However, the frame can also be collapsed down to this node without a loss of information since the type *sitting individual* is already defined as exhibiting these attribute-value pairs in the type signature. Thus, frames allow for zooming in and out of conceptual representations by expanding nodes referring to complex concepts. This makes them more flexible than the more rigid Two-Level Semantics representations illustrated above.

The frame of *sitzen* in Figure 2 does not represent the locative extension of the PV in which the location of the theme is specified in relation to another object introduced as internal argument of a locative PP such as *auf der Schaukel* 'on the swing'. Following Kaufmann, we assume that the support relation evokes or 'activates' further locational meaning which allows for merging the above frame with a figure–ground frame. This figure–ground frame integrates the locational information specified by the local PP. In general, for local prepositions we follow [36, 37] and others who assume that prepositions of this type single out specific regions with respect to the referent of the internal argument of the preposition and, in addition, predicate of an entity to be located in this specific region. According to this view, the semantic form of the nondirectional reading of the German preposition *auf* 'on' is represented as in (4), which is taken from [14, p. 111]. The representation states that *auf* denotes a relation holding between an object x which is located in the upper region of another object y. In addition, the second conjunct requires x to have contact with y since *auf* is a preposition which always involves contact. This conjunct is necessary in order to differentiate *auf* from *über* 'above', which also denotes location in the upper region of some object but does not imply contact.

(4) *auf* 'on' [-DIR]: $\lambda y \lambda x$ [LOC(x, UPPER_REGION(y)) & CONTACT(x,y)]

The representation in (4) can be translated into the frame in Figure 3, which contributes a figure–ground schema with the located entity and the reference object being introduced as values of the attributes FIGURE and GROUND, respectively. The meaning of the preposition is integrated into this frame as identifying the LOCATION of the figure with the UPPER REGION of the ground. The bidirectional broken arrow indicates that the instantiations of figure and ground are restricted to objects which are in physical contact with each other.

If *sitzen* is combined with a subject and a local PP headed by *auf*, the frames contributed by the three elements are merged into a complex frame in which the values of the THEME of the *sitzen* frame and the FIGURE of the *auf* frame are unified. This is illustrated by the frame for the sentence *Hans sitzt auf der Schaukel* 'Hans is sitting on the swing' in Figure 4.

The *sitzen* frame in Figure 2 necessarily involves contact of the supported object with the supporter since support in general cannot be conceived without contact. At the same time, the frame contributed by the preposition *auf* requires contact of the figure with the ground. Consequently, the ground is identified as the supporting entity, which is indicated by the thick arrow between the ground and the object whose upper side serves as supporter.

The region indicated by *auf* leaves some room for interpretation as to which part of the swing serves as the actual supporter: if Hans is located in the upper

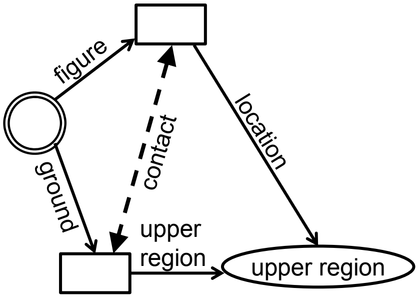


Fig. 3. Frame for *auf* 'on'

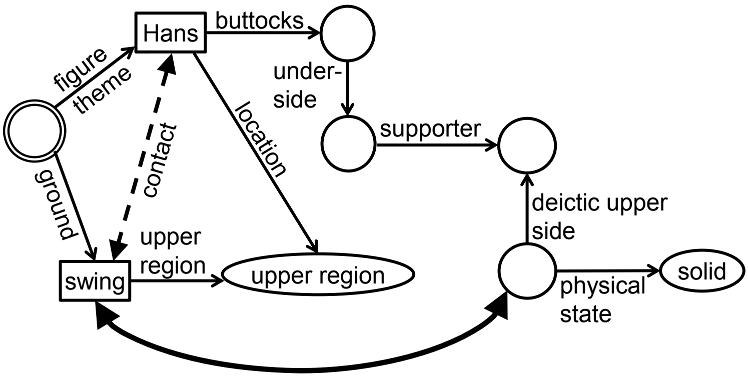


Fig. 4. Frame for *Hans sitzt auf der Schaukel* 'Hans is sitting on the swing'

region immediately above the seat, he is supported by this part of the swing. If, however, the design of the swing also consists of beams, its upper region additionally comprises the region above the top beam so that the frame in Figure 4 could also represent a situation in which Hans is located on the top beam. This second constellation is less expected since the design of a swing does not necessarily involve a beam construction and moreover its purpose requires being located on the seat. Nevertheless, the frame in Figure 4 allows for such a flexibility, which is in accordance with the interpretation of the natural language example in (2). Interestingly, the same ambiguity arises with other kinds of supporting entities such as armchairs and sofas which allow for (noncanonical) sitting on the back or armrests.[1]

Also note that the supporting entity and the ground are not necessarily identified. This becomes evident if *auf* is substituted by another preposition which does not imply contact, such as *unter* 'under' in *Hans sitzt unter dem Baum*

[1] We owe the observation concerning the interpretative flexibility of the frame in Figure 4 to an anonymous reviewer.

‘Hans is sitting under the tree’. In this sentence the supporting entity remains unrealized since it cannot be identified with the ground *tree* ‘Baum’.

The frame analysis of *sitzen* can easily be extended to the PV *knien* ‘kneel’, as in *Hans kniet auf dem Boden* ‘Hans is kneeling on the ground’. Like *sitzen*, *knien* requires solid support from below and only differs from *sitzen* in that the knees rather than the buttocks are supported.

3.2 *Liegen* ‘Lie’

In contrast to *sitzen* ‘sit’, *liegen* ‘lie’ cannot be sufficiently analyzed without making reference to object axes since *liegen* does not involve a specific part of an object, such as its back, as might be assumed. This is shown by the different positions of a brick illustrated in Figure 5 below.

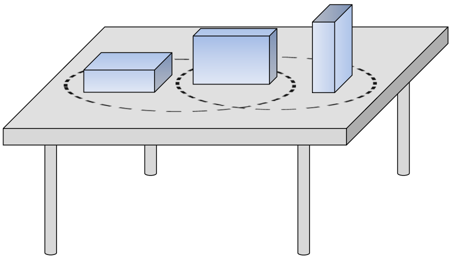


Fig. 5. Different orientations of bricks referred to by *liegen* ‘lie’ and *stehen* ‘stand’

The two bricks inside the left circle can both be referred to by the verb *liegen* whereas this is not possible for the remaining brick on the right, whose orientation can only be characterized by the PV *stehen* ‘stand.’ However, the brick in the middle can alternatively be referred to by *stehen* ‘stand’ so that it can be grouped together with the right brick as standing bricks (marked by the right circle). The fact that a brick usually does not have a clearly distinguishable part such as a back or a unique side shows that the choice of *liegen* and *stehen* does not depend on properties of this type. Instead, both PVs are sensitive to the orientation of object axes. *Liegen* requires alignment of the most prominent (longest) axis with the horizontal whereas *stehen* can be applied if either the longest or the second longest axis is oriented vertically. The conditions for *stehen* are, however, more intricate and will be discussed in the next section.

For *liegen*, [14, p. 108] proposes the semantic form in (5a) and the conceptual structure in (5b). The interpretation in (5b) states that *liegen* holds for an object if a side ‘s’ which is orthogonal to a nonprominent (‘nprom’) axis is supported from below by a solid object.

(5) a. *liegen* 'lie': $\lambda P \lambda x$ [LIE(x) & P(x)]
b. Int(LIE(x)) = $\exists$y [$support_S$(d-us(y)), s(nprom(x)))]

Given the conceptual structure in (5b), support from below entails that a non-prominent axis is aligned vertically. Since this also entails that the most prominent, i.e. maximal, axis is oriented horizontally, we will assume the simpler condition that *liegen* requires the maximal axis to be horizontal.

The PV *liegen* involves direct reference to object axes, which can be captured in the object schemata by [18–20]. As illustrated in Figure 6, the object axes are given in hierarchical order with the most prominent (or salient) object axis, 1D, to the left and the least prominent axis, 3D, to the right. In the second line, further information is specified: 1D is identified as the maximal (longest) axis, and 3D as the minimal (shortest) axis while 'Across' characterizes 2D as an axis which is oriented orthogonally to 1D.

The linear object schema can be translated directly into a frame of spatial objects as shown for *Ziegelstein* 'brick' in Figure 7.

1D	2D	3D
Max	Across	Min

Fig. 6. Object schema of *Ziegelstein* 'brick', linear representation

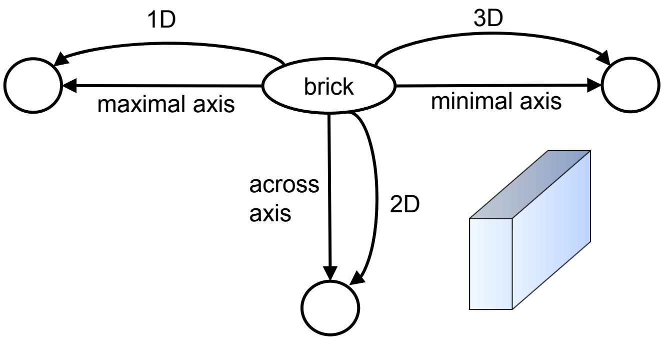


Fig. 7. Object schema of *Ziegelstein* 'brick', frame representation

1D, 2D, and 3D are captured as attributes of the brick. The values of these attributes are the axes identified by 1D, 2D, and 3D, respectively. The information which is specified for theses axes in the second line in the linear object schema above is also introduced by frame attributes, namely MAXIMAL AXIS, ACROSS AXIS, and MINIMAL AXIS. In the case of a brick, the values of these attributes are identified with the values of the attributes 1D, 2D, and 3D, respectively.

Figure 8 shows the frame for *liegen*. As can be seen, the portion of the object schema of the located figure relevant for *liegen* is specified as part of the meaning of the PV: a lying figure requires the most prominent object axis 1D to have a horizontal orientation. The remaining part of the meaning of a lying individual is almost identical to that of a sitting individual: the figure is kept in its position by a lower, supporting object which is solid. This component of the meaning of *liegen* only differs from *sitzen* in that it is not a specific part of the figure which is supported from below.

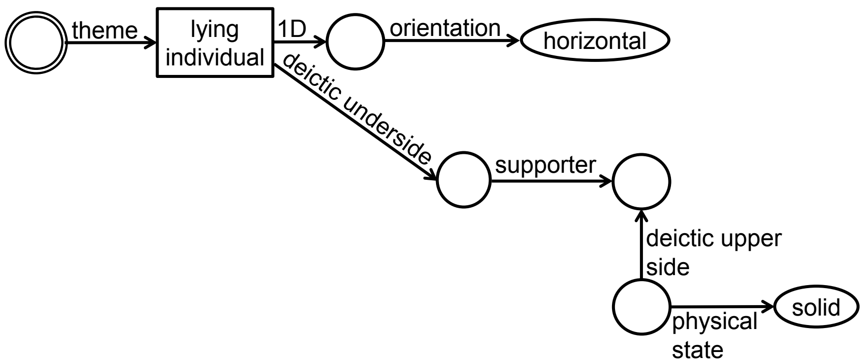


Fig. 8. Frame for *liegen* 'lie'

In the example in (6), the PV *liegen* is combined with the subject *Ziegelstein* 'brick' and the local PP *auf dem Tisch* 'on the table'. The corresponding frame in Figure 9 results from the unification of the frames contributed by the subject, the PV and the PP.

(6) *Der Ziegelstein liegt auf dem Tisch.*
the brick lies on the table
'The brick is lying on the table.'

As a further parallel to *sitzen*, the locational information contributed by the spatial preposition can be brought into the frame by a figure–ground frame, which is merged with the frame representing the meaning of the PV. Due the choice of the preposition *auf* 'on', which involves physical contact, the ground is identified as the supporting entity.

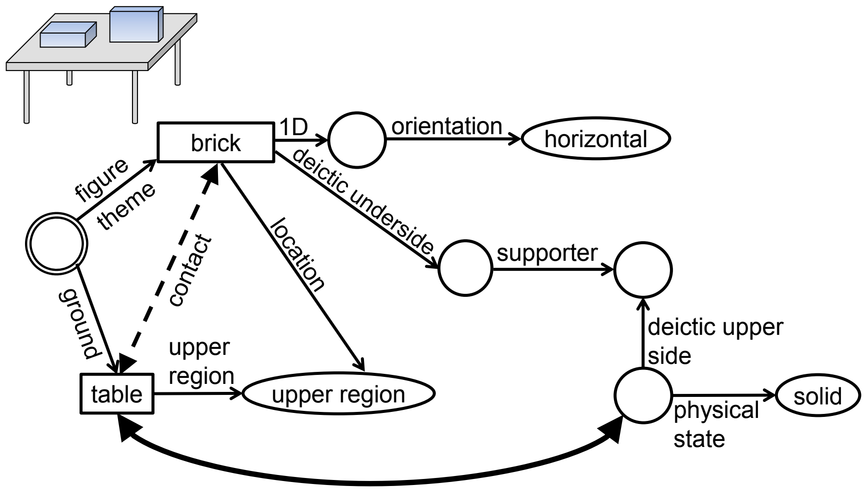


Fig. 9. Frame for *Der Ziegelstein liegt auf dem Tisch.* 'The brick is lying on the table.'

3.3 *Stehen* 'Stand'

As with *liegen*, the PV *stehen* 'stand' can be applied to two of the bricks in Figure 5 above. However, now the two bricks enclosed by the right circle exhibit orientations which can be referred to by *stehen* since the condition for the selection of *stehen* is that either the longest (i.e. maximal) axis or the second longest axis must be aligned vertically. Only the remaining brick on the left cannot be said to be standing since it is the shortest axis which is vertical in this case. In order to capture the two options for standing, one could assume two different frame representations which make explicit reference to the maximal axis 1D and the intermediate axis 2D and characterize them as aligned vertically. Alternatively, one can make use of the fact that both orientations of the brick which can be referred to by *stehen* exhibit a horizontal alignment of the minimal axis 3D. This alternative option is chosen in the frame for *stehen* in Figure 10.

Again, we will give an illustration of how the *stehen* frame in Figure 10 is combined with other frames when used in a full sentence: Figure 11 is a frame representation of the example in (7). The complex frame is yielded by unifying the single frames contributed by the subject, the PV *stehen* and the PP.

(7) *Der Ziegelstein steht auf dem Tisch.*
the brick stands on the table
'The brick is standing on the table.'

As can be seen, the frame differs only minimally from the frames for *sitzen* and *liegen*. Again, as with the preceding posture verbs, the frame contributed by the PV requires support from below by a solid object. In addition, the general location scenario activated by the verb allows for the integration of the

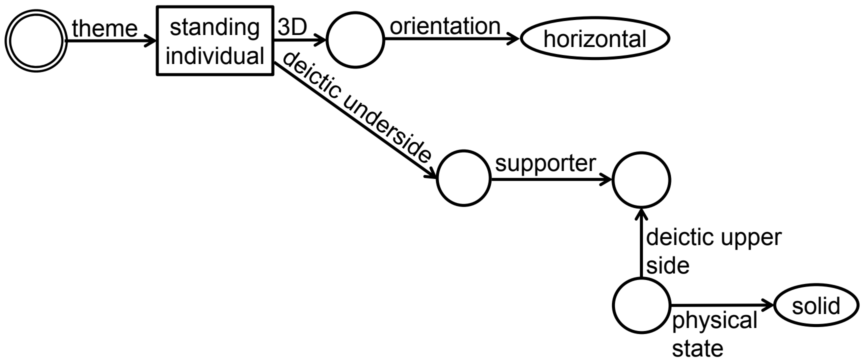


Fig. 10. Frame for *stehen* ‘stand’

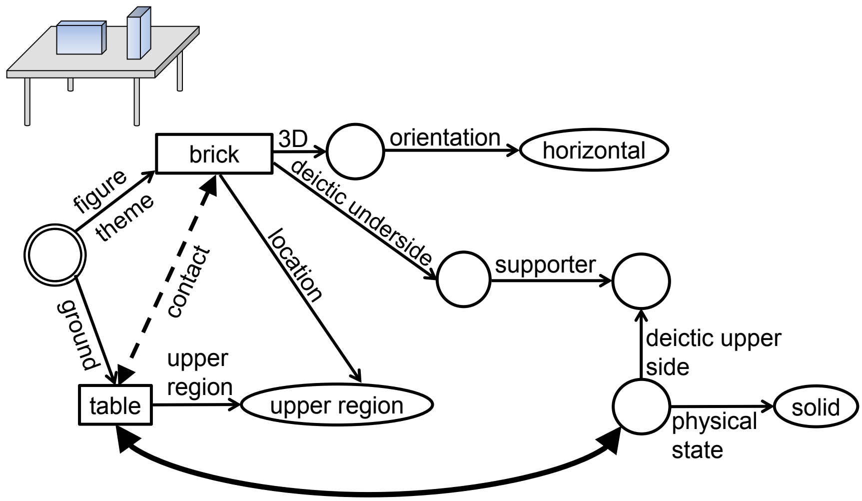


Fig. 11. Frame for *Der Ziegelstein steht auf dem Tisch.* ‘The brick is standing on the table.’

figure–ground subframe. As in the previous examples, the supporter is identified as the ground due to the choice of the preposition *auf* 'on', which implies contact between the figure and the ground.

A complication with *stehen* is that in German some objects can be said to be standing because of their canonical orientation, i.e., the way they are usually oriented. In this case, *stehen* can be applied even if the minimal axis is vertical as long as the object exhibits its canonical orientation. For example, a shoebox is oriented canonically when its lid is on top and can be opened, even though this usually involves a vertical orientation of the shortest axis. This specific use of *stehen* can be addressed by making reference to the so-called 'canonical vertical' ([18–20]), i.e., the axis which is vertical when the object exhibits its canonical orientation. By consequence, the object schema of a shoebox provided in Figure 12 below contains a specification of the minimal axis as canonical vertical. Technically, this is achieved by the frame attribute CANONICAL VERTICAL whose value is identical with the value of the frame attribute 3D.

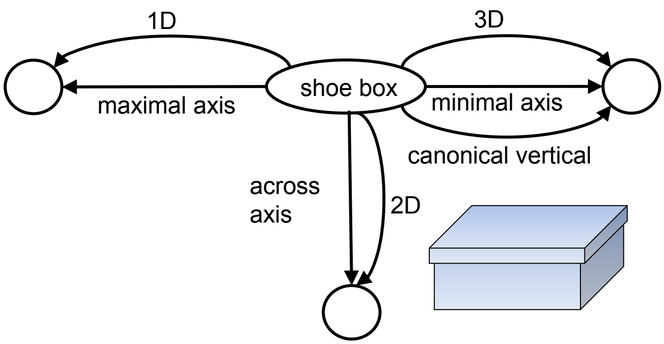


Fig. 12. Object schema for *Schuhkarton* 'shoebox'

For *stehen* to be applicable to objects with the shortest axis being the canonical vertical, the alternative frame for *stehen* in Figure 13 has to be assumed. This frame differs minimally from the frame for *stehen* in Figure 10. The only difference is that the figure is expected to exhibit an attribute CANONICAL VERTICAL which determines that the canonical vertical is vertically oriented. Only the relevant part of the alternative *stehen* frame is given in Figure 13 below. As can be seen, the alternative *stehen* frame does not make reference to the length of the axes but only requires the existence of a canonical vertical and its vertical orientation.

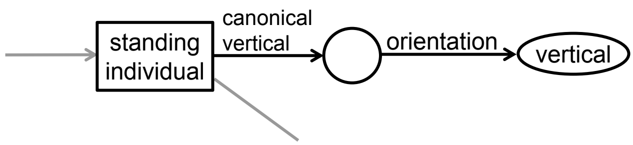


Fig. 13. Alternative *stehen* frame reduced to the relevant part

The frame in Figure 14 represents the example in (8). Since the object schema of a shoebox specifies the minimal axis 3D as canonical vertical, the frame resulting from the unification of the partial frames inherits the information that the minimal axis is the canonical vertical.

(8) *Der Schuhkarton steht im Schrank.*
the shoebox stands in.the wardrobe
literally: 'The shoebox is standing in the wardrobe.'

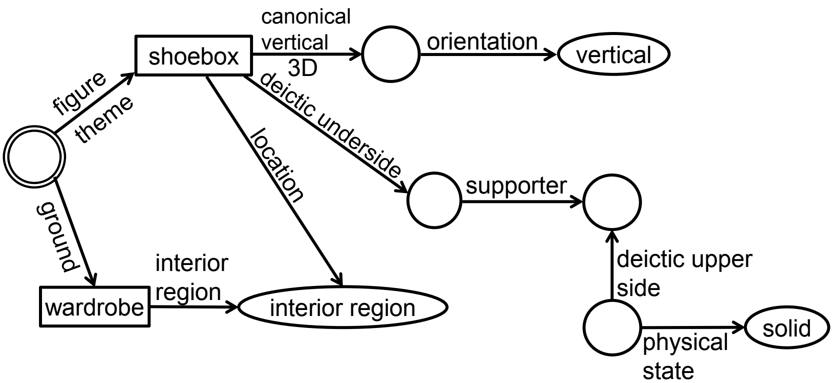


Fig. 14. Frame for *Der Schukarton steht im Schrank.* lit.: 'The shoebox is standing in the wardrobe.'

Again, the frame contributed by the local PP *im Schrank* 'in the wardrobe' is integrated into the figure–ground frame licensed by the overall locational scenario. As a contrast to the preceding examples, the preposition *in* 'in' does not imply contact but merely makes reference to the interior region of the wardrobe. By consequence, some interior part of the wardrobe, such as a shelf, can function as a supporter. Of course, this need not be the case since the box could also be located on some other object inside the wardrobe. However, our world knowledge of a wardrobe as a specific instance of a container gives us access to the information about potential supporters for objects which are located inside the wardrobe.

For objects with a canonical vertical which is not the maximal axis, the frames for *stehen* in Figure 13 and for *liegen* in Figure 8 compete if the canonical vertical is aligned vertically while at the same time the maximal axis is aligned horizontally. In this situation, the PVs *stehen* and *liegen* should be equally applicable. Yet, for an object like a shoebox the choice of *stehen* as in the example in (8) is natural whereas *liegen* sounds awkward. This difficulty can be overcome by imposing a specificity constraint on the choice of a PV which requires that the most specific PV is chosen. This would give *stehen* preference over *liegen* since the presence of a canonical vertical in the frame of *stehen* in Figure 13 can be considered more specific than the reference to the longest object axis in the frame of *liegen* in Figure 8.

Note that the different axes which can be aligned vertically are not directly addressed in Kaufmann's representation of *stehen* ([14, p. 108]) provided in (9) below.

(9) a. *stehen* 'stand': $\lambda P \lambda x$ [STAND(x) & P(x)]
b. Int(STAND(x)) = $\exists$y [support$_S$(d-us(y), s(prom(x)))]

The interpretation of *stehen* in (b) states that a side 's' which is orthogonal to a prominent axis ('prom') of the standing object x is supported by the deictic upper side of some other object y. The axis singled out by 'prom(x)' is defined as the maximal axis or the canonical vertical. However, it remains undefined how this process of axis selection works. In addition, it is not clear how objects like bricks are treated, which do not have a canonical vertical but can also be said to be standing if the second longest axis is vertical, as is the case with the brick in the middle in Figure 5 above.

4 Summary and Outlook

In this paper, we have shown that frame representations in the sense of Barsalou lend themselves naturally to the representation of posture verbs. The focus of our analysis has been on the German posture verbs *sitzen* 'sit', *liegen* 'lie', and *stehen* 'stand', which are already well described in the literature. In particular, we have drawn from the decompositional account of Kaufmann ([14]), who demonstrates that the choice of one of these verbs depends on (i) the body part which is supported to keep the located object in its position and (ii) the orientation of object axes including the so-called canonical vertical for objects which exhibit a canonical orientation. These factors have turned out to be easily translatable into attributes in the frame representation of the specific posture verbs. Likewise, properties of the located object and the locational information contributed by the spatial preposition correspond directly to a confined set of simple frame attributes. As a result, all the elements of the overall make-up of a posture/location scenario could by captured in the single, uniform format of frames. This is particularly true of the axis information provided in the object schemata by [18–20], which are extra-representational in Kaufmann's decompositional approach. Another advantage of the frame approach is that it allows for different degrees of explicitness, which is yielded by zooming in and out of conceptual representations by expanding nodes referring to complex concepts.

The analysis of German can be extended in several ways: first, in addition to the basic posture verbs discussed above there are other locative verbs such as *hängen* 'hang' and *lehnen* 'lean' which need to be considered in a comprehensive frame analysis of locative verbs. Second, posture verbs, like locative verbs in general, exhibit more abstract uses, often involving semantic drift or bleaching as in *Die Stadt liegt in einem Tal* 'The city lies in a valley' or *Der Verdächtige steht unter Beobachtung* 'The suspect is (literally: stands) under surveillance'. Here, a frame analysis that captures these semantic processes by means of reduction and reanalysis of single frame attributes seems promising. This approach may even be extended to the grammaticalized aspectual use of posture verbs as markers of the progressive aspect, which is not attested for German but for other Germanic languages such as Norwegian, Danish, Swedish, and Dutch ([9, 15, 21]). Third, the framework outlined above builds on the view that spatial prepositions are best analyzed in terms of regions ([36, 37] among others). However, [38, 39] argue that prepositions are better translated into vectors in order to deal with PP modifications like *The tree is ten meters behind the house* which involve the specification of a distance. The challenge for a frame approach is to cope with constructions of this type by introducing attributes such as DISTANCE and DIRECTION which are licensed or "activated" by attributes already contained in the frame of the spatial preposition.

Moreover, the frame account needs to be extended to other languages which differ significantly with respect to the repertory of posture verbs and the factors that govern their use. As a first step, we have contrasted the German data with posture verbs in French and Korean. The comparison between German and French already reveals numerous differences. For example, French does not have posture verbs directly corresponding to German *sitzen*, *liegen*, and *stehen*. Instead, it makes use of a variety of different strategies to refer to stative posture/location scenarios such as using the copula *être* or the unspecific locative verb *se trouver* 'be located', applying the resultative forms of change of posture verbs (e.g. *s'asseoir* 'sit down') and change of location verbs (e.g. *poser* 'place, put'), and also employing verbs which basically denote a change in spatial extension (e.g. *allonger* 'stretch out', which is preferably interpreted as 'lie' when its resultative form is combined with animate subject referents).

Like French, Korean does not have stative posture verbs but has a more systematic inventory of change of posture verbs whose resultative forms can be utilized to refer to stative location. However, the use of these verbs is constrained with respect to admissible subject referents. For example, *seta* 'stand' can only be combined with a nonhuman subject if the subject referent is at least as tall as a human (cf. [33]). As illustrated by the pair of examples in (10), *seta* can select a subject like *Pekhingem kwungcen* 'Buckingham Palace' but not a subject like *hwapwun* 'flower pot' since the height of the latter is usually far below the height of a human.

(10) a. *Pekhingem kwungcen-i nay nwun aph-ey*
Buckingham Palace-NOM my eye in.front.of-LOC
se-iss-ta.
stand-be-IND
literally: 'Buckingham Palace is standing in front of my eyes.'

b. **Ku hwapwun-i cengmwun-yeph-ey se-iss-ta.*
the flower.pot-NOM main.gate-side-LOC stand-be-IND
intended: 'The flower pot stands next to the main gate.'
(a-example from M.-H. Min p.c., b-example taken from [33, p. 361])

For the use of *seta* with nonhuman subject referents one can assume the frame in Figure 15, in which a value constraint on the attribute LENGTH requires that the vertical axis is as tall as or taller than a human. The example in (10a) shows that the vertical needs not be the longest axis since the vertical axis of Buckingham Palace is the shortest axis of the building. However, the vertical axis in (10a) is the canonical vertical. Consequently, the length attribute is built into the frame as an attribute of the canonical vertical.

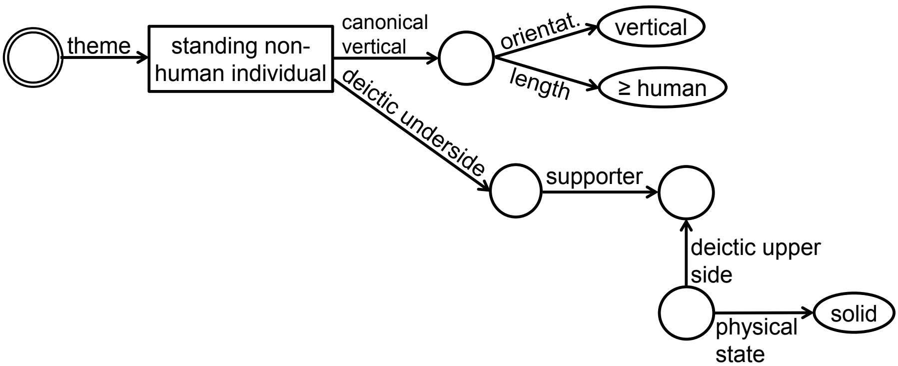


Fig. 15. Frame for Korean *seta* 'stand' combined with nonhuman subject referent

Finally, the analysis can be extended beyond linguistic matters. The frames presented above are built on linguistic analyses which investigate the necessary conditions for the use of specific verbs in a given language. The analyses are "minimal" in the sense that they focus on the factors which are relevant for the choice of a PV and refrain from representing detailed encyclopedic knowledge. Yet, given their flexibility frame representations can also be applied for purposes such as representing anatomical details and the motor activity which is required to remain in a specific posture. This may be useful in the domain of medicine as well as in other domains such as the recording and analysis of dance since these domains require making reference to different types of postures. Moreover, we have represented the relevant object axes and the part of the body which is supported as first-level (i.e. nonembedded) attributes of the theme argument of the PV. In a more structured frame representation these attributes would be part of attribute bundles referring to the composition and spatial configuration of the theme argument. For the sake of simplicity, we have not considered structural aspects of this kind which, however, are of major importance to purposes such as knowledge representation and inference systems.

The suggestions above indicate that the frame approach to posture verbs and locative verbs in general can be extended in many different directions. In spite of the sketchy character of the analysis outlined above, we consider it a promising framework for a further investigation of these phenomena.

Acknowledgments. This paper is a joint effort by the projects "Dimensional Verbs" and "Mathematical Modeling of Frames", which are part of the Collaborative Research Center "The Structure of Representations in Language, Cognition, and Science" supported by the German Science Foundation (DFG). We are grateful to our fellow researchers, in particular Jens Fleischhauer, Sebastian Löbner, Myeonghi Min, Albert Ortmann, Rainer Osswald, and Brigitte Schwarze. We are also indebted to the anonymous reviewer for helpful questions and suggestions. Moreover, we would like to thank the audiences at the the Fourth International Conference of the German Cognitive Linguistics Association, the Ninth International Tbilisi Symposium on Language, Logic and Computation, and the Fourth UK Cognitive Linguistics Conference for valuable comments.

Abbreviations

NOM	nominative
IND	indicative
LOC	locative
PV	posture verb

References

1. Ameka, F.K., Levinson, S.C.: The typology and semantics of locative predicates: posturals, positionals, and other beasts. Linguistics 45 (5/6), 847–871 (2007)
2. Asbury, A., Dotlačil, J., Gehrke, B., Nouwen, R.: Syntax and semantics of spatial P. John Benjamins, Amsterdam (2008)
3. Barsalou, L.W.: Frames, concepts, and conceptual fields. In: Lehrer, A., Kittay, E.F. (eds.) Frames, fields, and contrasts, pp. 21–74. Erlbaum, Hillsday (1992)
4. Berthele, R.: The typology of motion and posture verbs: a variationist account. In: Kortmann, B. (ed.) Dialectology meets typology, pp. 93–126. Mouton de Gruyter, Berlin (2004)
5. Bierwisch, M.: Semantik der Graduierung. In: Bierwisch, M., Lang, E. (eds.) Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven, pp. 91–286. Akademieverlag, Berlin (1987)

6. Bierwisch, M., Lang, E.: Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. Akademieverlag, Berlin (1987)
7. Bloom, P., Peterson, M.A., Nadel, L., Garrett, M.F.: Language and Space. MIT Press, Cambridge MA (1996)
8. Chomsky, N., Halle, M.: The sound pattern of English. Harper and Row, New York (1968)
9. Ebert, K.: Progressive markers in Germanic languages. In: Dahl, Ö. (ed.) Tense and Aspect in the languages of Europe, pp. 605–653. Mouton de Gruyter, Berlin (2000)
10. Evans, V., Chilton, P.: Language, cognition and space: the state of the art and new directions. Equinox Publishing, London (2010)
11. Fahlmann, S.E.: A system for representing and using real-world knowledge. PhD thesis. MIT (1977)
12. Gamerschlag, T., Petersen, W.: An analysis of the evidential use of perception verbs. In: Hart, C. (ed.) Selected Papers from UK-CLA Meetings. Vol. 1, pp. 1–18, http://uk-cla.org.uk/proceedings (2012)
13. Hickmann, M., Robert, S.: Space in Language: linguistic systems and cognitive categories. John Benjamins, Amsterdam (2006)
14. Kaufmann, I.: Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen: Die Kombinatorik lokaler Verben und prädikativer Komplemente. Niemeyer, Tübingen (1995)
15. Kuteva, T.: Auxiliation: an enquiry into the nature of grammaticalization. Oxford University Press, Oxford (2004)
16. Kutscher, S., Schultze-Berndt, E.: Why a folder lies in the basket although it is not lying: the semantics and use of German positional verbs with inanimate figures. Linguistics 45 (5/6), 983–1028 (2007)
17. Landau, B., Jackendoff, R.: 'What' and 'where' in spatial language and spatial cognition. Behavioral and Brain Sciences 16 (2), 217–238 (1993)
18. Lang, E.: Semantik der Dimensionsauszeichnung räumlicher Objekte. In: Bierwisch, M., Lang, E. (eds.) Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven, pp. 287–458. Akademieverlag, Berlin (1987)
19. Lang, E.: Spatial dimension terms. In: Haspelmath, M., König, E., Oesterreicher, W., Raible, W. (eds.) Language typology and language universals. An international handbook. HSK, vol. 20.2, pp. 1251–1275. Mouton de Gruyter, Berlin (2001)
20. Lang, E., Carstensen, K.-U., Simmons, G.: Modelling spatial knowledge on a linguistic basis. Lecture Notes in Artificial Intelligence, vol. 481. Springer, Berlin (1991)
21. Lemmens, M.: Aspectual posture verb constructions in Dutch. Journal of Germanic Linguistics 17 (3), 183–217 (2005)
22. Levinson, S.C.: Space in language and cognition: explorations in cognitive diversity. Cambridge University Press, Cambridge (2003)
23. Levinson, S.C., Wilkins, D.P.: Grammars of space. Cambridge University Press, Cambridge (2006)
24. Löbner, S.: Evidence for frames from human language. In: Gamerschlag, T., Gerland, D., Osswald, R., Petersen, W. (eds.) Concept types and frames in language, cognition and science (to appear)
25. Minsky, M.: A framework for representing knowledge. In: Winston, P.H. (ed.) The psychology of computer vision, pp. 211–277. McGraw-Hill, New York (1975)
26. Newman, J.: The linguistics of sitting, standing and lying. Benjamins, Amsterdam (2002)

27. O'Keefe, J.: The spatial prepositions in English, vector grammar, and the Cognitive Map Theory. In: Bloom, P., Peterson, M.A., Nadel, L., Garrett, M.F. (eds.) Language and Space, pp. 277–316. MIT Press, Cambridge (1996)
28. Petersen, W.: Decomposing concepts with frames. Baltic International Yearbook of Cognition, Logic and Communication 2, 151–170 (2007)
29. 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.) Concept types and frames in language, cognition and science (to appear)
30. Petersen, W., Osswald, T.: A formal interpretation of concept types and type shifts. In: Kosecki, K., Badio, J. (eds.) Cognitive Processes in Language. Lodz Studies in Language, vol. 25, pp. 183–191. Peter Lang, Frankfurt (2012)
31. Petersen, W., Werning, M.: Conceptual Fingerprints: Lexical Decomposition by Means of Frames – a Neuro-cognitive Model. In: Priss, U., Polovina, S., Hill, R. (eds.) ICCS 2007. LNCS (LNAI), vol. 4604, pp. 415–428. Springer, Heidelberg (2007)
32. Serra Borneto, C.: 'Liegen' and 'stehen' in German: a study in horizontality and verticality. In: Casad, E.H. (ed.) Cognitive Linguistics in the Redwoods: the Expansion of a New Paradigm, CLR 6, pp. 459–505. Mouton de Gruyter, Berlin (1996)
33. Song, J.J.: The posture verbs in Korean. In: Newman, J. (ed.) The linguistics of sitting, standing and lying, pp. 359–385. Benjamins, Amsterdam (2002)
34. Talmy, L.: Toward a cognitive semantics. Vol. 2. MIT Press, Cambridge MA (2000)
35. Vosgerau, G., Seuchter, T., Petersen, W.: Analyzing concepts in action frames. Manuscript. University of Düsseldorf (2012)
36. Wunderlich, D.: How do prepositional phrases fit into compositional syntax and semantics? Linguistics 29, 591–621 (1991)
37. Wunderlich, D., Herweg, M.: Lokale und Direktionale. In: von Stechow, A., Wunderlich, D. (eds.) Semantik. Ein Internationales Handbuch der zeitgenössischen Forschung. HSK, vol. 6, pp. 758–785. de Gruyter, Berlin (1991)
38. Zwarts, J.: Vectors as relative positions: a compositional semantics of modified PPs. Journal of Semantics 14, 57–86 (1997)
39. Zwarts, J., Winter, Y.: Vector space semantics: a model-theoretic analysis of locative prepositions. Journal of Logic, Language and Information 9, 169–211 (2000)

Alleged Assassins: Realist and Constructivist Semantics for Modal Modification

Bjørn Jespersen[1,⋆] and Giuseppe Primiero[2,⋆⋆]

[1] Institute of Philosophy, Department of Logic,
Czech Academy of Sciences, Prague;
Department of Computer Science,
Technical University of Ostrava (Czech Republic)
bjornjespersen@gmail.com

[2] Centre for Logic and Philosophy of Science,
Ghent University (Belgium)
giuseppe.primiero@ugent.be

Abstract. Modal modifiers such as *Alleged* oscillate between being subsective and being privative. If individual a is an alleged assassin (at some parameter of evaluation) then it is an open question whether a is an assassin (at that parameter). Standardly, modal modifiers are *negatively* defined, in terms of *failed* inferences or *non*-intersectivity or *non*-extensionality. Modal modifiers are in want of a positive definition and a worked-out logical semantics. This paper offers two positive definitions. The *realist* definition is elaborated within Tichý's Transparent Intensional Logic (TIL) and builds upon Montague's model-theoretic semantics for adjectives as representing mappings from properties to properties. The *constructivist* definition is based on an extension of Martin-Löf's Constructive Type Theory (CTT) so as to accommodate partial verification. We show that, and why, "a is an alleged assassin" and "Allegedly, a is an assassin" are equivalent in TIL and synonymous in CTT.

Keywords: Modal modification, property vs. propositional modification, *alleged*, *allegedly*, Transparent Intensional Logic, Constructive Type Theory.

1 Introduction and Overview

Kamp's seminal [10] seeks to draw a line between those adjectives whose meaning is a property and those adjectives whose meaning is a function that maps properties to properties. Kamp agrees with Montague's typing of properties as a function $\langle s, \langle e, t\rangle\rangle$ from a world/time pair s to a function from entity e to truth-value t. Montague [13, p.211] suggests that all adjectives have

⋆ Research Fellow, Project GACR 401/10/0792: *Temporal Aspects of Knowledge and Information.*

⋆⋆ Postdoctoral Fellow of the Research Foundation - Flanders (FWO).

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 94–114, 2013.

a property-to-property function as their meaning.[1] Kamp [10, pp.147ff] suggests that most adjectives have a property as their meaning. He admits that it would seem that some adjectives must occur in attributive position and are incapable of occurring in predicative position. Their meaning is a property-to-property function. Kamp's prime example of such a recalcitrant adjective is 'alleged':

> Is *alleged* a predicate, even in the most diluted sense? It seems not. [...] The same can be said to be true [...] of adjectives such as *fake*, *skilful*, or *good*. Where precisely we should draw the boundaries of the class of adjectives to which the second theory [property-to-property function] applies I do not know. For example, does *skilful* belong to this class? Surely we must always ask 'skilful what?' before we can answer the question whether a certain thing or person is indeed skilful [...].[2]

We agree with Kamp's linguistic observations. Kamp is concerned with a demarcation among adjectives. We hypothesize that his demarcation is in effect a demarcation between those adjectives that represent *properties* and those that represent *property modifiers*. Thus, to use Kamp's own example from [10, p. 123], "Every alleged thief is a thief" is not an instance of predication of two properties, *being alleged* and *being a thief*. Instead it is an instance of predication of one modified property, *being an alleged thief*. We must always ask 'alleged what', for nothing and nobody can be alleged, pure and simple. Nor is the logical form of the sentence, in predicate logic, anything like

$$\forall x((Alleged\ x \wedge Thief\ x) \rightarrow Thief\ x)$$

This form would, erroneously, trade a possible falsehood for a logical truth ([10, p.123]).

Kamp distinguishes between four kinds of adjectives in terms of their logical behaviour. Roughly the same taxonomy is known from modification theory. Interestingly, *Alleged* falls outside Kamp's taxonomy. In fact, *Alleged* and its ilk remain to this day a dimly lit corner of the research into adjectives and modifiers.[3] For instance, Partee [14, p.9] says that

> *Nonsubsective* adjectives may be either *modal* – expressing possibility or some other modal meanings – or *privative*, entailing negation. [...] There is no meaning postulate for the *modal* adjectives, since they have no entailments – an *alleged murderer* may or may not be a murderer, and similarly for adjectives like *possible*, *proposed*, *expected*, *doubtful*.

Yet everyone who is competent with the predicate 'is an alleged assassin' knows that it applies to someone who has been alleged to be an assassin and that they

[1] Beesely [1] argues, *contra* Montague, that also evaluative adjectives like 'good' and 'tall' have a property as their meaning. Beesley holds that "a is a good F" should be given the intersective, or conjunctive, analysis "a is good and a is an F". Beesley, however, does not extend his claim to 'alleged' and suchlike; nor is it obvious how to do so.

[2] [10, pp. 153-4].

[3] See [3], [11], [15], [22], [27].

may, or may not, actually be an assassin.[4] So there is some clearly circumscribed linguistic competence to account for. What has as yet not been established is how to provide a *positive definition* of the sort of modifier that *Alleged* typifies.

Modal modifiers such as *Alleged* are uniquely characterized by oscillating between being *subsective* and being *privative*. Formally, the rule of subsective modification (cf. Kamp's 'affirmative adjective') eliminates the modifier, while the rule of privative modification replaces the modifier by negation. Let a be an individual, F a property, M a property modifier, and $[MF]$ the property resulting from modifying F by M. Then a modifier is subsective if it validates this inference (in rudimentary predicate-logical notation, to begin):

$$\frac{[[M_sF]\ a]}{Fa}$$

For instance, if a is a *wine-drinking Georgian* then a is a Georgian, hence *Wine-drinking* is subsective. In extensional terms, a set of wine-drinking Georgians must be extracted from a set of Georgians. Hence if a belongs to a set of wine-drinking Georgians then a belongs *ipso facto* to a set of Georgians. Subsective modification is the simplest kind of modification and of little logical interest.[5]

A modifier is privative if it validates this inference:

$$\frac{[[M_pF]\ a]}{\neg Fa}$$

For instance, if a is a *fake banknote* then a is not a banknote, hence *Fake* is privative. In extensional terms, a set of fake banknotes must be extracted from the complement of a set of banknotes. Hence if a belongs to a set of fake banknotes then a belongs *ipso facto* to a set in the complement of a set of banknotes. Privative modification is logically much more delicate than subsective modification. As [2] shows, iterated privative modification cannot be modelled by iteration of propositional negation. It must be modelled by property negation. As a result, because a logic of multiple privation is a logic of *contraries*, a pair of privative modifiers is equivalent to a modal modifier. The above rule applies only to single privation.

In virtue of the oscillation between subsection and privation, if a is an alleged assassin then either a is an assassin or a is not an assassin. So the rule of inference defining modal modifiers would seem straightforward:[6]

---

[4] Partee (in personal communication at LOGICA 2012, Hejnice) points out that though she held that modal adjectives/modifiers lack a meaning postulate she did not hold that they lack meaning.

[5] Within subsective modification the simplest kind of modification is constituted by *trivial* modification: a is a lump of *genuine* gold iff a is a lump of gold. A trivial modifier returns the modified property unmodified, as it were. The polar contrary is privative modification. We note that our adoption of trivial modification is at variance with Kamp and Partee's *non-vacuity principle* [11, p.161].

[6] [20] may have given the impression that the above inference was the rule we proposed at the time for modal modifiers. We intended no such impression, however. See [20, p.269]. See also [8].

$$\frac{[[M_m F]\ a]}{Fa \vee \neg Fa}$$

But, of course, this classical tautology is trivially satisfied by *all* modifiers. A subsective modifier will invariably validate the left-hand disjunct. A privative modifier will invariably validate the right-hand disjunct. What is non-trivial is that a modal modifier will sometimes validate the left-hand disjunct and sometimes the right-hand disjunct. For any one instance of $[M_m F]\ a$, the paucity of the informational value of $[M_m F]\ a$, when true, is such that it cannot be inferred which side of $Fa, \neg Fa$ truth will come down on. This is what we mean by modal modifiers oscillating between subsection and privation. It is obvious, then, why the 'conjunctive' analysis $((\mathit{Alleged}\ x \wedge \mathit{Thief}\ x) \rightarrow \mathit{Thief}\ x)$ is to no avail. It eliminates modifiers from the analysis. And it prejudges in favour of subsection at the expense of privation, thereby missing the unique feature of modal modifiers.

It is not immediately obvious what a *positive* definition of modal modification would amount to. It is easy enough to characterize modal modification *negatively*. First, as we saw, a modal modifier fails to validate either of $Fa, \neg Fa$ as the conclusion of an argument whose only premise is $[M_m F]\ a$. Second, a modal modifier is *non-intersective* for failure to validate this argument, M_i an intersective modifier (Kamp: 'predicative'):

$$\frac{[[M_i F]\ a]}{M^* a \wedge Fa}$$

For instance, if a is a wine-drinking Georgian then a is a wine-drinker and a is a Georgian. In extensional terms, a set of individuals with the property $[M_i F]$ is the intersection of a set of Fs and a set of M^*s. Notice that M_i is a modifier whereas M^* is a property. A modifier cannot be detached from a context in which it modifies a property and be predicated of an individual. Instead it can be *pseudo-detached* in the following manner: if a is an $[MF]$, M an arbitrary modifier, then there is a property p such that a is an $[Mp]$. M^* is the schematic property $[Mp]$. The conclusion of the rule of inference defining intersective modification is formed by means of the rules of pseudo-detachment, subsection, and $\wedge$-introduction.[7]

Third, a modal modifier is *non-extensional* (or *intensional*, in the pejorative sense of 'intensional') for failure to validate this argument (adapted from [10, p.125]):

[7] The need for a rule of *left subsectivity* such as the one of pseudo-detachment tends to be overlooked in the Montagovian tradition. [14, p.3], for one, puts forward a meaning postulate to regulate intersective adjectives: For each intersective meaning ADJ', it holds that $\exists P_{\langle e,t\rangle} \forall Q_{\langle s\langle e,t\rangle\rangle} [ADJ'(Q)(x) \leftrightarrow P(x) \wedge {}^{\vee}Q(x)]$. The meaning postulate gets the truth-condition right, but fails to account for the transition from ADJ' to P. Beesleys' theory in [1] makes left subsectivity trivial, in virtue of his conjunctive analysis. The left conjunct is simply obtained by conjunction elimination. Beesley's task, of course, is to make a case for 'good' denoting a property rather than a modifier. See [7] for further discussion.

$$\frac{Fa \leftrightarrow Ga}{[MF]\ a \leftrightarrow [MG]\ a}$$

For instance, even if it so happens that all and only kings are philosophers, it may not follow that all and only belligerent kings are belligerent philosophers. An individual who is both philosopher and king may be a belligerent king (waging war in his capacity as king) without being a belligerent philosopher (waging war in his capacity as philosopher). Hence *Belligerent* is a non-extensional modifier. Logically, non-extensional modifiers are those that do *not distribute*, because they are logically sensitive to whether a belongs to F or G, even though the extension of F happens to be identical to the extension of G.[8]

So modal modifiers are non-intersective, hence non-extensional, possibly (non-) subsective and possibly (non-) privative. The *actual truth* of $[M_m F]\ a$ entails that one of two *possibilities* is realized: a being an F; a not being an F. Thus there is a striking similarity between *modal modifiers* and *non-factive attitudes.* Assume that b *knows whether* a is an assassin. As is well-known, *knowing whether* is invariant under complementation: a knows whether $A \equiv a$ knows whether $\neg A$.[9] If a is an assassin then b knows that a is an assassin; if a is not an assassin then b knows that a is not an assassin. But the fact that b knows whether a is an assassin entails that one of the same two possibilities as above is true. The truth of a knowing whether b is an assassin is compatible with either disjunct being true.[10]

The similarity between modal modifiers and non-factive attitudes suggests to us that modal modification should be modelled in terms of *possibility.* One conception of possibility is in terms of *alethic possibility*: reality may turn out in one of two contrary ways. Another conception is as *epistemic possibility*: something rather than the opposite may be known. We shall develop both conceptions below. The former is based on Tichý's Transparent Intensional Logic (TIL). Formally, TIL is a hyperintensional, partial, typed λ-calculus, whose syntax is

[8] Since modal modifiers are non-intersective they must also be non-extensional. Kamp [10, pp. 125-6] states, correctly, that all intersective modifiers (predicative adjectives) are extensional, and wonders whether the converse holds. The jury is still out. What seems obvious is that most, or all, logically, semantically and philosophically interesting modifiers are going to be non-extensional, because it is interesting in each individual case why they fail to distribute. For further discussion, see [2, p.11]. [20, p.253] uses 'intensional' interchangeably with 'modal'; but they are better used as labels for two different kinds of modifiers.

[9] See [5, §5.1.4].

[10] The link between modal modifiers and non-factive attitudes probably runs deeper than we let on in the present paper. [15, p.152] provides the following list of 'plain nonsubsective' (in effect, modal) modifiers/adjectives: *potential*, *alleged*, *arguable*, *likely*, *predicted*, *putative*, *questionable*, *disputed.* With the exception of *potential*, they all have something attitudinal about them. And all of those attitudes are non-factive. A bold hypothesis would be that almost all modal modifiers are parasitic on non-factive attitudes. Modal modifiers should not be filed under 'nonsubsective', for due to their oscillation between subsection and privation each modal modifier will be subsective on some occasions and privative on other occasions.

interpreted by means of a realist procedural semantics. The portion of TIL that concerns property modification is continuous with Montague's: a property modifier is a mapping from properties to properties. In TIL a property is logically a mapping from a logical space of possible worlds to a mapping from times to sets of individuals, where sets of individuals are characteristic functions. The latter conception, of possibility as epistemic possibility, is based on an extension of Martin-Löf's Constructive Type Theory (CTT). Formally, CTT is a typed calculus based on intuitionistic logic, endorsing the Curry-Howard isomorphism (propositions-as-types) and the equivalence between sets and propositions under a constructive syntax.[11] A modifier is obtained by interpreting in the appropriate way the assertion conditions under which a formula holds.

This paper builds on [20], which applies CTT and TIL to privative modification. Common features of TIL and CTT include:

- a functional approach based on the typed lambda calculus
- a typed universe
- an interpreted logical syntax
- a notion of meanings as constructions/procedures: a proof procedure for a proposition (CTT); a procedure for producing (in this case) a possible-world proposition/empirical truth-condition (TIL).

The key differences are that TIL offers a procedural semantics erected upon a model-theoretic structure for modifiers whereas CTT offers a proof-theoretic semantics.[12] In [20] a TIL property modifier is a function from properties to properties, whereas a CTT property modifier is a function from sets to sets. This latter difference is particularly important for our present purposes. A constructive set is a set of proof-objects for a proposition, and since constructive propositions are identified with their sets of proof-objects, a constructive property modifier is, in the final analysis, a function from propositions to propositions (hence a function between intensional entities). Modal modifiers require an interpretation of partially evaluated terms as the range of the function at hand. Therefore the CTT analysis of "a is an alleged assassin" is in effect an analysis of "Allegedly, a is an assassin", *Allegedly* being a propositional modifier. We point out below the equivalence (though not synonymy) of "a is an alleged assassin" and "Allegedly, a is an assassin" in TIL.

The rest of this paper is organized as follows. Section 2 presents the TIL explanation of modal modification. Section 3 presents the CTT explanation of modal modification. Section 4 compares the main results.

[11] For a recent introduction to the Curry-Howard isomorphism, see [25].

[12] The modal logic of TIL is $S5$ with a constant domain. See [5, ch.4]. The standard modal interpretation of Martin-Löf's Type Theory refers to $S4$: see [16]. For semantic considerations related to the possibility operator underlying the extension of CTT used in this paper, see the relations to usually modally defined knowledge operators in [21].

2 TIL: Types and Constructions

TIL comes with a ramified type hierarchy embedding a simple type theory.

Definition 1 (Type of Order 1)
Let B be a base, *where a base is a collection of pair-wise disjoint, non-empty sets. Then:*

i) Every member of B is an elementary type of order 1 over B;
ii) Let $\alpha, \beta_1, \ldots, \beta_m (m > 0)$ be types of order 1 over B. Then the collection $(\alpha\beta_1 \ldots \beta_m)$ of all m-ary partial mappings from $\beta_1 \times \ldots \times \beta_m$ into α is a functional type of order 1 over B;
iii) Nothing is a type of order 1 over B *unless it so follows from (i) and (ii).*

Remark. For the purposes of natural-language analysis, the following base of ground types is currently assumed:

– o: the set of truth-values $\{\mathbf{T}, \mathbf{F}\}$
– ι: the set of individuals (a constant universe of discourse)
– τ: the set of real numbers (doubling as temporal continuum)
– ω: the set of logically possible worlds (the logical space)

Functional types are defined over those ground types in the standard manner. A functional type with domain in possible worlds is an intensional type as known from possible-world semantics. The simple type theory suffices to type properties, propositions, property and propositional modifiers:

– property: $(((o\iota)\tau)\omega)$, abbreviated as '$(o\iota)_{\tau\omega}$'
– property modifier: $((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
– proposition: $((o\tau)\omega)$, abbreviated as '$o_{\tau\omega}$'
– propositional modifier: $(o_{\tau\omega}o_{\tau\omega})$

We model those empirical conditions as *possible-world intensions.* Formally, intensions are entities of type $(\beta\omega)$: mappings from possible worlds to an arbitrary type β. The type β is frequently the type of the *chronology* of α-objects, i.e. a mapping of type $(\alpha\tau)$. Thus α-intensions are frequently functions of type $((\alpha\tau)\omega)$, abbreviated as '$\alpha_{\tau\omega}$'. We shall typically say that an index of evaluation is a world/time pair $\langle w, t\rangle$. *Extensional entities* are entities of some type α where $\alpha \neq (\beta\omega)$ for any type β.

Logical procedures are so-called *constructions.* There are six different kinds of constructions, three of which are defined inductively below.[13] The basic idea is that functional abstraction is the very procedure of forming or presenting or *constructing* a function (rather than the resulting function); that functional application is the very procedure of applying function to argument (rather than the resulting functional value); and that variables provide input for those procedures to operate on.

[13] The constructions *Trivialization*, *Single* and *Double Execution* are not needed for present purposes. However, see [5, §1.3.2].

Definition 2 (Construction)

(i) *The* Variable x *is a* construction *that constructs an object O of the respective type dependently on a valuation v; it v-constructs O;*
(ii) *The* Composition $[XY_1 \dots Y_m]$ *is the following* construction*. If* X v-constructs *a function* f *of a type* $(\alpha\beta_1 \dots \beta_m)$*, and* $Y_1, \dots, Y_m$ v-construct *entities* $B_1, \dots, B_m$ *of types* $\beta_1, \dots, \beta_m$*, respectively, then the* Composition $[XY_1 \dots Y_m]$ v-constructs *the value (an entity, if any, of type* α*) of* f *on the tuple-argument* $\langle B_1, \dots, B_m \rangle$*. Otherwise, the* Composition $[XY_1 \dots Y_m]$ *does not* v-construct *anything and so is* v-improper*;*
(iii) *The* Closure $[\lambda x_1 \dots x_m Y]$ *is the following* construction*. Let* $x_1, x_2, \dots, x_m$ *be pairwise distinct variables v-constructing entities of types* $\beta_1, \dots, \beta_m$ *and* Y *a construction v-constructing an entity of type* α*. Then* $[\lambda x_1 \dots x_m Y]$ *is the* construction λ-Closure *(or* Closure*). It* v-constructs *the following function* f *of type* $(\alpha\beta_1 \dots \beta_m)$*. Let* $v(B_1/x_1, \dots, B_m/x_m)$ *be a valuation identical with* v *at least up to assigning objects* $B_1, \dots, B_m$ *of types* $\beta_1, \dots, \beta_m$*, respectively, to variables* $x_1, \dots, x_m$*. If* Y *is* $v(B_1/x_1, \dots, B_m/x_m)$*-improper (see iii), then* f *is undefined on* $\langle B_1, \dots, B_m \rangle$*. Otherwise, the value of* f *on* $\langle B_1, \dots, B_m \rangle$ *is the entity of type* α $v(B_1/x_1, \dots, B_m/x_m)$*-constructed by* Y*;*
(iv) *Nothing is a* construction*, unless it so follows from (i) through (iii).*

Explicit intensionalization and temporalization enables TIL to encode constructions of possible-world intensions, by means of terms for possible-world variables and times, directly in the logical syntax.[14] Where w ranges over ω and t over τ, the following logical form essentially characterizes the logical syntax of an empirical sentence:

$$\lambda w \lambda t[\dots w \dots t \dots]$$

Any instance of this schematic Closure constructs the set of $\langle w, t \rangle$ pairs at which the empirical truth-condition constructed by $[\dots w \dots t \dots]$ is satisfied.

Thus the Closure

$$\lambda w \lambda t[Assassin_{wt}\ a]$$

constructs the set of $\langle w, t \rangle$ pairs at which a is an assassin. The extensionalized property $[Assassin]_{wt}$ is applied to a to obtain a truth-value, which is subsequently abstracted over to obtain a possible-world proposition. $\lambda w \lambda t[Assassin_{wt}\ a]$ is a *hyperproposition* whose procedural product is a possible-world proposition.

To construct the modified property *being an alleged assassin*, the modifier *Alleged* is applied to the property *Assassin*, and the resulting property is extensionalized for application to a. Thus the meaning of "a is an alleged assassin" is the Closure

$$\lambda w \lambda t[[Alleged\ Assassin]_{wt}\ a]$$

14 See [5, §2.4.3] for a comparison between TIL and Montague Grammar.

2.1 TIL: Definitions of Modifiers

Kamp's definitions of his respective kinds of adjectives (predicative, affirmative, etc.) are model-theoretic. An adjective is, for instance, predicative, provided it satisfies the definition of predicative adjectives on all admissible interpretations. Kamp then proceeds to note that he doubts that at least any adjective that is privative on some admissible interpretation is privative on all of them [10, p.125]. The parallel phenomenon for modifiers is discussed in [9, §2.5]. For instance, *Nordic* gold is not gold (but a copper alloy) whereas a Nordic salmon is a salmon. And a *false* friend is not a friend (though pretending to be one) whereas a false proposition is still a proposition.[15] One could even argue that *Alleged* is not exclusively modal. If an *alleged proposition* is a proposition that someone alleges to be true then that occurrence of *Alleged* is as a subsective modifier. In general, we must not succumb to the naïve assumption that every modifier would be only intersective, only privative, etc.

In the procedural semantics of TIL we define first the respective *sets* of subsective, privative, intersective, and modal modifiers. Then we issue this conditional: *if* modifier M is intersective (etc.) *then* so-and-so follows. It is left up to a given interpretation to decide whether M is in fact intersective (etc). Any such interpretation will ideally be sensitive to the actual, extra-model status of M.[16] It is too strong to demand that M must be privative on all admissible interpretations in order to qualify as privative on one admissible interpretation.

So the logical status of the modifiers *Nordic*, *False*, etc, is a function of their respective arguments. E.g. the definition of *privative modifier* is a definition of the set of modifiers that are privative with respect to the argument property F. The type of a set of property modifiers is $(o((o\iota)_{\tau\omega}(o\iota)_{\tau\omega}))$. Modifiers will be defined in terms of *requisite properties.*[17] A requisite of a property G is a property F such that anything with property G must also have F.

Definition 3 (Requisite Relation Over $(o\iota)_{\tau\omega}$). *Let $X, Y/(o\iota)_{\tau\omega}$, and let x range over ι. Then*

$$[Req\ YX] = \forall w \forall t [\forall x [X_{wt}\ x] \supset [Y_{wt}\ x]]$$

[15] See also [14, p. 9]. Of course, Partee famously wishes to reduce all modifiers to subsective ones, leaving room only for (some) modal modifiers as not being necessarily subsective. See [24] for objections to Partee's proposal. In TIL the categories of subsective, privative, intersective, and modal modifiers are mutually irreducible. Yet in a derived or indirect sense privation is a variant of subsection. For example, extensionalize the property of being a banknote to obtain a set, then generate its complement, and apply comprehension to that set to extract its subset (perhaps empty, perhaps non-empty) of fake banknotes. In CTT, privative modifiers are an extreme kind of subsective modifiers: where the latter generate proper subsets, the former take the improper subset of functions mapping to the empty set. See [20] for details.

[16] It falls to linguistic analysis to decide what that status of M is. There is no logical mechanism for deciding it. (Thanks to Petr Šimon.)

[17] See [5, Ch. 4, esp. §4.1] and [9, §2.5ff] on the notion of requisite. *Privative modifier* was also defined by means of requisites in [20, §4.2].

Gloss *definiendum* as "Y is a requisite of X", and *definiens* as "Necessarily, any x instantiating X at any $\langle w, t\rangle$ also instantiates Y at $\langle w, t\rangle$." For instance, if the property *being a mammal* is a requisite of the property *being a whale* then if a happens to be a whale at $\langle w, t\rangle$ it is necessary that a also be a mammal at $\langle w, t\rangle$. Or if at $\langle w, t\rangle$ a has the modified property $[M_sF]$ then it is necessary that a also be an F at $\langle w, t\rangle$. The reason is because F is a requisite of $[M_sF]$: necessarily, whatever is an $[M_sF]$ is an F.

Definition 4 (Subsective, Privative, Modal, Intersective Modifiers). *Let g, g', g'', g''' range over $((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$; let g'''^* range over $(o\iota)_{\tau\omega}$; let x range over ι; $F/(o\iota)_{\tau\omega}$; $\exists/(o(o\iota))$; $\wedge/(ooo)$; $\neg/(oo)$. Then:*

Subsective w.r.t. F: $\lambda g[Req\ F[g\ F]]$
Privative w.r.t. F: $\lambda g'[Req[\lambda w\lambda t\lambda x\neg[F_{wt}x]][g'\ F]]$
Modal w.r.t. F: $\lambda g''[Req[\lambda w\lambda t[\lambda x[[\exists\lambda w'[[[\exists\lambda t'[[[M_m\ F]_{wt}\ x] \supset [F_{w't'}\ x]]]] \wedge [\exists\lambda w''[\exists\lambda t''[[[M_m\ F]_{wt}x] \supset \neg[F_{w''t''}x]]]]]]][g''\ F]]]]$
Intersective w.r.t. F: $\lambda g'''[Req[\lambda w\lambda t\lambda x[[g'''^*_{wt}x] \wedge [F_{wt}x]]][g'''\ F]]$.

A modal modifier behaves with respect to one and the same property F as subsective at one $\langle w', t'\rangle$ and as privative at another $\langle w'', t''\rangle$. No other modifier has the feature that its status (here, subsective vs. privative) depends on the given $\langle w, t\rangle$ of evaluation. The definition of modal modifiers defines the set of modifiers g'' that are modal with respect to F, such that if a is a $[g''\ F]$ at $\langle w, t\rangle$ then at $\langle w, t\rangle$, possibly, a is an F and, possibly, a is not an F. Put differently, whenever a has the property $[g''\ F]$ then a also has the property of being such that at one $\langle w', t'\rangle$ a is an F and at another $\langle w'', t''\rangle$ a is not an F. To compare subsective, privative and modal modifiers, every $\langle w, t\rangle$ is such that if x is an $[M_sF]$ then x is an F, and if x is an $[M_pF]$ then x is not an F. Furthermore, every $\langle w, t\rangle$ is such that if x is an $[M_mF]$ then it is possible that x be an F and it is possible that x not be an F. This last inference does not apply to $[M_sF]\ x$ or $[M_pF]\ x$. If x is an $[M_sF]$ then it is not possible that x not be an F (for it is necessary that x be an F). And if x is an $[M_pF]$ then it is not possible that x be an F (for it is necessary that x not be an F).

2.2 TIL: *Alleged*

To enable a direct comparison between TIL and CTT, TIL will also state an introduction rule and an elimination rule. The *definition* of modal modifiers, however, is the one provided in terms of requisites in Def.4.

What is required to acquire the property of being an alleged assassin? That somebody performs the *speech act* of alleging that a is an assassin.[18] Let f range

[18] The introduction rule assumes that any existential presupposition pertaining to the premise should be satisfied. If it is true that x alleges that the King of France is bald then it will be neither true nor false that the King of France has the property of being alleged to be bald, for there is currently no King of France around to instantiate that property. Thus the rule would not take a truth to a truth and so be invalid. We are suppressing the issues of existential presupposition, partiality, and truth-value gaps to keep the basic exposition as simple as possible.

over $(o\iota)_{\tau\omega}$; $Alleges/((o\iota o_{\tau\omega})_{\tau\omega})$: a relation-in-intension between individuals and propositions they allege to be true.[19]

Then the *introduction rule* for *Alleged* is

$$\frac{\exists\lambda x[Alleges_{wt}\ x\lambda w\lambda t[f_{wt}\ a]]}{[[Alleged\ f]_{wt}\ a]}$$

Gloss: "If somebody alleges that a is an f then a is an alleged f."

In fact, the set of properties f such that somebody alleges that a has f is identical to the set of properties f such that a is an alleged f:

$$\lambda f[\exists\lambda x[Alleges_{wt}\ x\ \lambda w\lambda t[f_{wt}\ a]]] = \lambda f[[Alleged\ f]_{wt}\ a]$$

However, it is not obvious how to generalize from this particular introduction rule to an introduction rule for any modal modifier. Sometimes a speech act is required, and sometimes an attitude, and sometimes something else. For instance, it is not obvious what the introduction rule for *Possible* would be, as soon as we want more than *ab esse ad posse*. As a hypothesis, however, TIL proposes this general type-theoretic pattern underlying an introduction rule: where the premise has an object of type $(o\iota o_{\tau\omega})_{\tau\omega}$ the conclusion must have an object of type $((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$.

The *elimination* rule for M_m can be stated in full generality, though. From Definition 2 we obtain the following rule, f ranging over properties:

$$\frac{[[M_m f]_{wt}\ x]}{\exists\lambda w'[\exists\lambda t'[[[M_m\ f]_{wt}\ x] \supset [f_{w't'}\ x]]] \wedge \exists\lambda w''[\exists\lambda t''[[[M_m f]_{wt}\ x] \supset \neg[f_{w''t''}\ x]]]}$$

Gloss: "From x being an $[M_m\ f]$ at $\langle w,t\rangle$, infer that there is a $\langle w',t'\rangle$ such that if x is an $[M_m\ f]$ at $\langle w,t\rangle$ then x is an f at $\langle w',t'\rangle$ and that there is a different $\langle w'',t''\rangle$ such that if x is an $[M_m\ f]$ at $\langle w,t\rangle$ then x is not an f at $\langle w'',t''\rangle$."

Absolute elimination of M_m in the conclusion is impossible due to the oscillation between subsection and privation, so the rule must be restricted to conditional elimination.

The set-theoretic counterpart of modal modification is the *union* of two disjoint sets. In the example of *alleged assassin*, the relevant union is the union of the set of assassins at $\langle w',t'\rangle$ and the set of non-assassins at $\langle w'',t''\rangle$. It is logically trivial that an alleged assassin, just like any other individual, is a member of that union, but it is not epistemically trivial which of its two subsets a given alleged assassin, or any other individual, belongs to.

[19] This typing is another simplification. TIL would tend to type *Alleges* as an empirical relation to a *hyperproposition*. But in this paper we have not introduced the ramified type hierarchy required to type hyperintensions. The simplification saves us from having to explain the descent from a hyperproposition that has been alleged to the proposition it constructs.

2.3 TIL: *Allegedly*

In [5, p.506] it is argued that *Allegedly*, as referred to in "Allegedly, a is an assassin", must be a *propositional property* – of type $(oo_{\tau\omega})_{\tau\omega}$ – rather than a *propositional modifier*, of type $(o_{\tau\omega}\ o_{\tau\omega})$. The argument is that a propositional property can, as a propositional modifier cannot, preserve the contingency of the proposition denoted by "Allegedly, ...". This argument is misconceived. The adverb 'allegedly' may well be analyzed as denoting a propositional modifier.

The reason is straightforward. The meaning of the above sentence is

$$[Allegedly\ \lambda w\lambda t[Assassin_{wt}\ a]]$$

The result of applying *Allegedly* to the proposition constructed by $\lambda w\lambda t$ $[Assassin_{wt}\ a]$ is another proposition, which is true at all those $\langle w, t\rangle$ where it is alleged that a is an assassin. The proposition constructed by $[Allegedly\ \lambda w\lambda t$ $[Assassin_{wt}\ a]]$ is as contingent as anything. Where p ranges over propositions, the *introduction rule* for *Allegedly* is

$$\frac{\exists\lambda x[Alleges_{wt} x\ p]}{[Allegedly\ p]_{wt}}$$

Gloss: "If somebody alleges that p, then p is alleged (is allegedly true)."

Let M'_m be a propositional modifier. Then the *elimination rule* for M'_m can be stated in full generality:

$$\frac{[M'_m\ p]_{wt}}{\exists\lambda w'[\exists\lambda t'[[M'_m\ p]_{wt} \supset p_{w't'}]] \wedge \exists\lambda w''[\exists\lambda t''[[M'_m\ p]_{wt} \supset \neg p_{w''t''}]]}$$

Gloss: "From p being modified by $[M'_m]$ at $\langle w, t\rangle$, infer that there is a $\langle w', t'\rangle$ such that if p is modified by $[M'_m]$ at $\langle w, t\rangle$ then p is true at $\langle w', t'\rangle$ and that there is a different $\langle w'', t''\rangle$ such that if p is modified by $[M'_m]$ at $\langle w, t\rangle$ then $\neg p$ is true at $\langle w'', t''\rangle$."

As with *Alleged*, the open question is whether the $\langle w, t\rangle$ of the premise is the $\langle w', t'\rangle$ or else the $\langle w'', t''\rangle$ of the conclusion. This is simply to say that an allegation may, or may not, be true. It is readily seen from the respective introduction and elimination rules for *Alleged* and *Allegedly* that the Composition

$$[Allegedly\ \lambda w\lambda t[F_{wt}\ a]]$$

and the Closure

$$\lambda w\lambda t[[Alleged\ F]_{wt}\ a]$$

are equivalent, but not synonymous, constructions of the same proposition. Those two different sentential meanings converge in the same truth-condition.[20]

[20] See [5, def. 2.3, p.154] for a definition of the individuation of meaning in terms of *procedural isomorphism*. In [4] we present a neo-Churchian Alternative $\frac{1}{2}$ that defines procedural isomorphism in terms of α- and η-conversion and an Alternative $\frac{3}{4}$ that adds a rule of restricted β-conversion.

3 CTT on Modal Modification

In the spirit of anti-realist semantics, a constructivist understanding of modal modification can be given by directly representing the conditions for the assertion of a modified judgement. Accordingly, we shall consider a modal modifier M_m as an operator applying to a predication Fa, where a property F is predicated of an individual a. The predication of a modally modified property will be abbreviated as '$M_m[Fa]$'. Notice the shift of position from the previously considered '$[M_m F]\ a$', a shift motivated by the application of the modifier to the predication as a whole, leading in the following to our analysis of *Alleged* as *Allegedly*.

In Section 1 we stressed that M_m oscillates between being privative and being subsective, thereby making it impossible to infer which of Fa and $\neg Fa$ holds when the premise is $M_m[Fa]$. Subsection for CTT is standardly given by proper subset formation; privative modification, as introduced in [20], is defined as a mapping to the empty set; the oscillation of modal modification between the two corresponds precisely to the contingent satisfaction of either the maximal proper subset satisfying the property involved by the predication, or of the contradictory construction. The only constructive way to understand this oscillation is to formulate Fa's truth value in terms of *contingently* satisfiable conditions, thus leaving open the question of which conditions are *actually* satisfied. For a constructivist semantics, where truth is defined by constructors and refutations are given by implication to contradiction, this is no obvious task. On such a strong understanding of proven and refuted contents, no space is left for a formal representation of *contingent* truths.

A way to formulate assertion conditions for contingent truths is offered by the formal distinction between *refuted contents* and *missing constructions*. This was already exploited in [12], where classical formulas were reduced to intuitionistic 'pseudo-truths' by double negation, the implication from $\neg\neg Fa$ to Fa being valid in a finite domain. In the same vein, the meaning of a valid judgement Fa *true* justifies the further conclusion that no construction for $\neg Fa$ *true* is possible. Then, a weaker form of predication is justified by inferring from $\neg\neg Fa$ the new judgement Fx, where the constructor x stands for a *refutable assumption*, corresponding to a place-holder for a (yet) missing, though admissible, construction of truth. In other words: no proof that Fa is a valid predication has been given; but, provided no refutation has been performed either, one such assumption can be made, up to refutation or confirmation being provided. We shall further explore this direction in order to provide appropriate rules for modal modification in a constructivist setting.

3.1 Modal Types for Contingent Truth

We rely on the previous work [18] for the complete presentation of a type theory with refutable assumptions. We shall consider briefly the modal fragment of that language, which is here adapted for the interpretation of modal modifiers. In the following, we shall revert to the standard type-theoretic notation that expresses a predication Fa in the form of an object of type $a:F$.

Let us start by considering a polymorphism of both types and constructors: we have one kind of expressions F *type*, by which we collect propositions *justified* by appropriate verifications $a, b, \ldots$; another kind of expressions F *type*$_{inf}$ collects propositions which are *assumed* to be true, as their constructions are not refuted and only their negation is not yet refuted. For these expressions a constructor is thus a variable $x, y, \ldots$, induced from a judgement $\neg(F \rightarrow \bot)$. Judgements of the first sort induce a constructive notion of truth (*true*); the second ones, a weaker predicate of *contingent* truth (*true**). Identity of terms holds within *type*, and its constructors are composed in the standard manner by way of listing, application, abstraction and pairing to define connectives (conjunction and implication) and quantifiers. Conversion rules are defined over terms of the *type*$_{inf}$ fragment, β-reduction of *type*$_{inf}$ terms to corresponding *type* terms (evaluation) and α-term equality.

Respecting the usual convention of distinguishing true from valid assumptions, we shall refer to a set of valid assumptions Δ as a set of constructions of the *type* kind used to infer another construction; a set of true assumptions Γ is a set of constructions of the kind *type*$_{inf}$, used to infer another construction. Contextual judgements are thus built by derivability from judgements $\Gamma = \{F_1 \; true^*, \ldots, F_n \; true^*\}$, which establishes the truth of F under refutable assumptions $F_1, \ldots, F_n$. When those $F_1, \ldots, F_n$ are fully verified (computationally, by reducing them by β-conversion to appropriate terms in *type*) the validity of F *true* is established. Derivability from refutable assumptions defines a truth predicate at a particular stage, depending on possible further states of knowledge. Derivability from valid assumptions expresses validity preserved under any further condition. In this way, we have extended the usual conceptual description of types in terms of two semantic notions of truth and derivability, respectively, for provable and refutable contents. To preserve this distinction two start rules are defined:

$$\frac{}{\Gamma, a:F, \Delta \vdash F \; true} \text{ Premise Rule}$$

$$\frac{}{\Gamma, x:F, \Delta \vdash F \; true^*} \text{ Hypothesis Rule}$$

The premise rule allows us to derive explicitly verified contents. The hypothesis rule reflects the derivation of contents that are only assumed to be true. The *true* predicate can be understood as validity (that is, truth in every situation) and it corresponds to truth by verification, whereas the predicate *true** corresponds to truth in a context of (true) assumptions.

To express such a distinction in the object language of the type system, we extend our analysis to a modal language.[21] This also allows restoring a unique semantic predicate. The modal operators are informally introduced following the previous explanation of truth and validity: provided that the conditions for

[21] For more on the following explanation of epistemic modalities, see [17].

having the right to express a judgement are satisfied, the notion of judgemental necessity $\Box(F\ true)$ corresponds to that of an *apodictic judgement*: what is known to be thus and so cannot be known to be otherwise. The constructive interpretation identifies provability, truth and knowledge. The basic condition for the truth of Fa is thus the individual (construction, proof) a that makes F true $(a{:}F)$; when F presupposes further types (propositions) to be valid (true), these represent the context in which F is formulated (instantiated, known). Conditions in such a context Γ can be seen as contextual or background knowledge. Hence, $\Box(F\ true)$ is knowledge for which no further contextual conditions are needed $(\Gamma = \emptyset)$. The corresponding interpretation of a judgemental possibility operator starts from the propositional equivalence $\Box F \leftrightarrow \neg\Diamond\neg F$; this leads in [26] to the other equivalence: $\Diamond(F\ true) \Leftrightarrow \neg\Box(\neg F\ true)$. If conditions needed for the knowledge of $F\ true$ can all be satisfied only with Γ empty, then this formula reduces to the conditions for $\Box(F\ true)$. Otherwise, truth is preserved in certain knowledge states in which appropriate conditions $\Gamma = (F_1\ true^*, \ldots, F_n\ true^*), n \geq 1$ are formulated. The latter amounts in our language to $x_i : F_i$ as a condition for $F\ true$. Hence, we infer modalities directly from our polymorphic constructors:

$$\frac{a:F}{\Box(F\ true)}\ \Box\text{-formation} \qquad \frac{x:F}{\Diamond(F\ true)}\ \Diamond\text{-formation}$$

The inference to the truth of contextual judgements requires generalization to contextual formulae:

Definition 5 (Necessitation Context). *For any context Γ, $\Box\Gamma$ is given by $\bigcup\{\Box F_i\ true \mid$ for all $F_i \in \Gamma\}$.*

Definition 6 (Normal Context). *For any context Γ, $\Diamond\Gamma$ is given by $\bigcup\{\circ F_i\ true \mid \circ = \{\Box, \Diamond\}$ and $\Diamond F_i$ true for at least one $F_i \in \Gamma\}$.*

Then a judgement valid under assumptions becomes a possibility judgement if its context remains normal, that is, at least one its assumptions is $true^*$.[22] For the use of contingent truths as the key to modelling modal modification, we are interested here in the rule that characterizes the use of Normal Contexts. Local validity (or derivability under true assumptions) is defined by introduction and elimination rules for the $\Diamond$-operator:[23]

$$\frac{\Gamma, x_i : F_i \vdash F\ true^*}{\Box\Gamma, \Diamond(F_i\ true) \vdash \Diamond(F\ true)}\ \text{I-}\Diamond$$

[22] *Necessitation* and *Normal Context* are equivalent to *Global* and *Local Context* as known from the literature in modal logic. Cf. [6].

[23] Structural rules such as Weakening, Contraction, Exchange hold in the form of theorems; also, a rule of substitution for truth predicates and terms can be proved, plus the local inversion of these modal rules with the appropriate $\Box$-counterparts, corresponding to their soundness and completeness. See [18].

This rule has a corresponding elimination rule that returns the $true^*$ predicate from a $\Diamond(F\ true)$ judgement occurring in the second premise.[24] For current purposes, we formulate the elimination rule to generate explicitly two predications resulting, respectively, from verifying or refuting the conditions:

$$\frac{\Box\Gamma, \Diamond(F_i\ true) \vdash \Diamond(F\ true) \qquad [x_i/a_i]:F_i}{\Box\Gamma, \Box(F_i\ true) \vdash \Box(F\ true)}\ \text{E-}\Diamond(1)$$

$$\frac{\Box\Gamma, \Diamond(F_i\ true) \vdash \Diamond(F\ true) \qquad F_i \rightarrow \bot}{\Box\Gamma, \Box(\neg(F_i\ true)) \vdash \Box(\neg(F\ true))}\ \text{E-}\Diamond(2)$$

3.2 CTT: *Alleged* and *Allegedly*

The basic idea informing our simulation of a modal modifier is to interpret it as a modal operator producing a non-terminating set of terms to provide assertion conditions for a *contingent* truth. Our judgemental $\Diamond$ operator and its introduction and elimination rules regulate precisely such epistemic conditions.

Let us start by reconsidering the example "a is an assassin". We start with a an individual constructor in the kind *type*, and F the property (*assassin*) predicated of a. A valid predication Fa corresponds, in our language, to a judgement of the form $a:F$, expressing the (proven) fact that there is an individual a who is an assassin:[25]

$[Fa]$: "(It is true that) a is an assassin"

with the truth predicate hidden by the formalism. It is crucial for our construction to unveil the nature of such a truth predicate, i.e. to establish which rules it obeys. When M is a modal modifier like *Alleged*, $[MF]\ a$ should be unpacked as

$[MF]\ a$: "(It is true that) a is an alleged assassin"

Our claim is that the modal modifier M over F applied to a can be simulated in terms of our I-$\Diamond$ rule by analytically defining the conditions under which the truth of Fa is asserted. The meaning of $[MF]\ a$ is in turn equivalent to $\Diamond(F\ true)$.

To get started, let us note the following. The constructivist epistemology underlying the present interpretation of modal modifiers rests entirely on the *perspectivalist* view of performing acts of judging propositions, according to which acts of judging (hence of knowing) are always acts performed from within a

[24] See [18].

[25] This would, ideally, be the individual a caught in the act of killing someone. As explained below, this rather unrealistic representation is replaced by the requirement that the true predication "John is an assassin" satisfies all the conditions that make John an assassin.

first-person perspective by an epistemic subject in an appropriate knowing context.[26] This means that judgements ground speech acts, which may remain implicit. Hence, the above-mentioned reading of *Alleged* readily transforms into the following:

$M[Fa]$: "Allegedly, (it is true that) a is an assassin"

In fact, where the apodictic form of judgement expressed by "a is an assassin" is grounded in a proof object independent of unsatisfied conditions, the modally modified judgement expressed by "Allegedly, a is an assassin" is grounded in a proof object dependent on refutable conditions, as explained above.

We thus exploit the nature of the predication as dependently defined. Every assertion can be formulated as depending on (possibly implicit) conditions. The most obvious conditions can be made evident in terms of an analytic deconstruction of the predicate, as e.g. by saying that "John is an assassin, provided it is true that he killed a human being on purpose". Besides this analytical form, a judgement can be turned into a dependent one by referring to conditions dictated by the *perspective* from which the act of judging occurs. As an example, let us assume that the proposition expressed by "John is an assassin" is asserted from within the perspective of a legal system where to be convicted as an assassin requires that the individual has been found guilty by the lowest to the highest courts. Then, an obvious formulation of our dependency relation would be of the form: "John is an assassin, provided it is true that he has been found guilty of killing a human being by each of the required courts". The list of conditions can be further modified by adding, for example, "under no mitigating circumstances".[27] If and when such conditions are proved to hold, we shall declare John an assassin. The modified form, predicating of John that he is an alleged assassin, holds in so far as it cannot be decided that all the conditions that make John an assassin hold.

In fact, $\Diamond(F\ true)$ expresses precisely the impossibility of reducing the constructor for F to *type*. This means that at least one of the assumptions under which $a\!:\!F$ is constructed remains unverified, hence making it impossible to assert that a has property F. Then, one cannot judge either Fa or $\neg Fa$ to be true. Notice that our $\Diamond$ rules prevent ill-behaving inferences. In particular, it is not possible to infer that John has allegedly killed someone from the fact that John is not an assassin: the latter has its own set of (satisfied, valid) assumptions, falsifying the open conditions for being an *alleged assassin*. It is also impossible

[26] On the perspectivalist epistemology underlying Martin-Löf's Type Theory, see e.g. [23]. For the use of such an approach in the epistemological debate on information and knowledge, see [19].

[27] The perspective can be easily changed so that also conditions change. For example, for someone who considers hunting an act of violence, the following might hold: "John is an assassin, for he killed a beast". Our concern is here only to assess the role of dependent judging in the formulation of a construction for modal modifiers. Hence we shall restrict ourselves to the more evident and less problematic formulation of such meanings.

to infer that John is an alleged assassin from *not knowing* that John has killed someone: the latter means that there is no predication at all with respect to $a : F$, hence also no account of the conditions for considering its truth. As a result, no judgement candidate has been laid down for acceptance or refutation.

The validity of the dependent judgement corresponds to the reduction of the construction to the *type* fragment and hence of its assertion conditions to $\Box\Gamma$, so as to finally execute an inference $\Box\Gamma \vdash \Box(F\ true)$. To do so, we require that none of the conditions under which $F\ true$ holds be falsifiable. The construction $\Diamond(F\ true)$ expresses instead the validity of $type_{inf}$ for at least one condition which does not reduce.

$M[Fa]$ is the modally modified predication of F of a. It is constructively expressed as a function $M(x)[x : F]$, saying that for at least one $F[x_i : F_i]$ it holds that $F_i\ type_{inf}$, and hence $F\ true^*$.[28] To preserve the functional aspect of M in the constructive notation, we will refer to $M(x)[x : F]$ as the type satisfied by some $f : F$ modified by having a judgement of the form $f : F$, for which at least one $f_i : F_i$ cannot be shown to reduce:

$$\frac{F\ type[\Gamma] \qquad F_i\ type_{inf} \in \Gamma \qquad M(x)[x : F]}{\Box\Gamma, (x_i : F_i) f : F \vdash F\ type[(x_i(f))(f_i) : F_i]}\ \text{Modal Modification}$$

Gloss: "Let there be an object type F that is satisfied, provided that all the object types in $\Gamma = \{F_1, \ldots, F_n\}$ have appropriate type constructors; let it be the case that for $F_i \in \Gamma$ a constructor is admissible but no reduction is provided, so that $F_i\ type_{inf}$ holds; then it is the case that, provided all the constructors in Γ apart from (at least) F_i are satisfied, a modifier M holds for F such that F is an object *type* if and only if F_i has an appropriate β-contractum, and it does not hold if F_i does not reduce." This rule is nothing other than an analytic definition of $type_{inf}$, inducing immediately the judgement $\Diamond(F\ true)$.

Three remarks are in place to appreciate why such a construction qualifies as one of a modal modifier:

1. the modal operator expresses the separation between terminating and non-terminating terms, a property which is not available in the standard format of CTT; by presenting the constructor $\Box\Gamma, (x_i : F_i) f : F \vdash F\ type[(x_i(f))(f_i) : F_i]$, we refer to a term f that is modified by the missing reduction for a term f_i on which it depends;
2. given the admissibility of f_i, this construction simulates a derivability relation that satisfies the *tertium non datur* of the non-modal alternatives $Fa \vee \neg Fa$, though the language does not allow its formal derivability from $\Diamond(F\ true)$. The appropriate way of expressing the meaning of a modal modifier such as *Alleged* in a sentence like "(It is true that) John is an alleged assassin" is to say: "(It is true that) John is an assassin if x is true" or "(It is false that) John is an assassin if x is not true";
3. the resolution of the contingency of the modal predication is possible by one of the two elimination rules.

[28] Because M applies to an a already predicated in F, CTT has no need for a counterpart of pseudo-detachment or any other rule of left subsectivity.

4 Conclusion

The uniquely defining feature that any theory of modal modifiers must accommodate is their oscillation between subsection and privation. It is relative to a particular context of evaluation whether a particular modal modifier is subsective or else privative. No other modifier – be it subsective, privative or intersective – shares this feature of context-sensitivity.

We suggested above an intimate connection between modal modifiers and non-factive attitudes. From b knowing whether a is an F it follows only that it is possible that a be an F and that it is possible that a not be an F. If a is an alleged assassin, the same two possibilites hold. We worked out an account of modal modifiers in two different directions. Within TIL we worked out an account of alethic possibility. The truth of a being an alleged assassin is logically compatible with one of two possible states-of-affairs obtaining: a being an assassin, a failing to be an assassin. Within CTT we worked out an account of epistemic possibility. The knowledge that a is an alleged assassin is compatible with either of two possible pieces of knowledge: knowing that a is an assassin, knowing that a is not an assassin. Neither has been refuted or verified.

The TIL definition of modal modifiers M_m says that, necessarily, if x has the property $[M_m\ F]$ at $\langle w, t\rangle$ then, possibly, x is an F and, possibly, x is not an F. Possibly, there is a $\langle w', t'\rangle$ such that if x is an $[M_m F]$ then x is an F, and possibly, there is an alternative $\langle w'', t''\rangle$ such that if x is an $[M_m F]$ then x fails to be an F. It falls to empirical inquiry to establish whether $\langle w, t\rangle$ is like $\langle w', t'\rangle$ or like $\langle w'', t''\rangle$. The meaning of an adjective denoting a modal modifier is a procedure whose product is a mapping from properties to properties. From the definition of modal modifiers we obtained a conditional elimination rule for M_m. We also provided an introduction rule for *Alleged*, while pointing out that it may not generalize to all other modal modifiers.

The CTT definition of modal modifiers M_m is given in terms of rules for a modal operator $\Diamond$ that applies to a judgement of the form $(F\ true)$. The introduction rule spells out how to form such a judgement $\Diamond(F\ true)$ from laying down its assertion conditions, which are neither verified nor refuted. The elimination can be given in one of two forms: by verifying or by refuting conditions. The rule of Modal Modification expresses this dependency from open assumptions in the form of a function from constructions to constructions.

It follows readily from the TIL introduction and elimination rules for the property modifier M_m that the sentence "a is an alleged assassin" is equivalent with the sentence "Allegedly, a is an assassin", *Allegedly* a propositional modifier. Those are not synonymous sentences, however, since their respective meanings are two different procedures that produce the same truth-condition (possible-world proposition). In CTT the relationship between those two sentences is synonymy, because one is the logical analysis of the other. The logical form of "a is an alleged assassin" is, in the final analysis, $M_m[Fa]$ and not $[M_m\ F]a$. $M_m[Fa]$ is the modally modified judgement that the proposition that a is an F is true. This judgement can be made, defeasibly, as long as it has not yet been established whether a is an assassin.

Acknowledgements. The authors wish to thank Marie Duží and Petr Šimon for helpful comments and remarks. The two authors of this paper and of [20] contributed equally to both papers. Giuseppe Primiero wishes to acknowledge the support of the Isaac Newton Institute for Mathematical Science, Cambridge (UK), where he was a Visiting Fellow for the scientific programme *Semantics and Syntax: a Legacy of Alan Turing* during the final drafting of this paper.

References

1. Beesley, K.R.: Evaluative adjectives as one-place predicates in Montague Grammar. Journal of Semantics 1, 195–249 (1982)
2. Carrara, M., Jespersen, B.: Double privation and multiply modified artefact properties (in submission)
3. Cresswell, M.: Prepositions and points of view. Linguistics and Philosophy 2, 1–41 (1978)
4. Duží, M., Jespersen, B.: Procedural isomorphism, analytic information and β-conversion by value. Logic Journal of the IGPL (2012), doi: 10.1093/jigpal/jzso44
5. Duží, M., Jespersen, B., Materna, P.: Procedural Semantics for Hyperintensional Logic: Foundations and Applications of Transparent Intensional Logic. Logic, Epistemology, and the Unity of Science, vol. 17. Springer, Berlin (2010)
6. Fitting, M.: Basic modal logic. In: Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 4, pp. 368–449. Oxford University Press (1994)
7. Jespersen, B.: Left subsectivity (in submission)
8. Jespersen, B., Carrara, M.: Two conceptions of technical malfunction. Theoria 77, 117–138 (2011)
9. Jespersen, B., Carrara, M.: A new logic of technical malfunction. Studia Logica, doi:10.1007/s11225-012-9397-8
10. Kamp, H.: Two theories of adjectives. In: Keenan, E. (ed.) Formal Semantics of Natural Language, pp. 123–155. Cambridge University Press, Cambridge (1975)
11. Kamp, H., Partee, B.: Prototype theory and compositionality. Cognition 57, 129–191 (1995)
12. Kolmogorov, A.: On the principle of excluded middle. In: van Heijenoort, J. (ed.) From Frege to Gödel: a Source Book in Mathematical Logic 1879-1931, pp. 414–437. Harvard University Press (1967)
13. Montague, R.: English as a formal language. In: Visentini, B., et al. (eds.) Linguaggi Nella Societá e Nella Tecnica, pp. 189–224. Milan (1970); Reprinted in Thomasson, R.H. (ed.), Formal Philosophy. Yale University Press, New Haven (1974)
14. Partee, B.: Privative adjectives: subsective plus coercion. In: Bäuerle, R., Reyle, U., Zimmermann, T.E. (eds.) Presuppositions and Discourse. Elsevier, Amsterdam (2001), `http://people.umass.edu/partee/docs/ParteeInPressKampFest.pdf`
15. Partee, B.: Compositionality and coercion in semantics: the dynamics of adjective meaning. In: Bouma, G., et al. (eds.) Cognitive Foundations of Interpretation, pp. 145–161. Royal Netherlands Academy of Arts and Sciences, Amsterdam, `https://udrive.oit.umass.edu/partee/Partee2007CompositionalityCoercionAdjs.pdf`
16. Pfenning, F., Davies, R.: A judgemental reconstruction of modal logic. Mathematical Structures in Computer Science 11, 511–540 (2001)
17. Primiero, G.: Epistemic modalities. In: Primiero, G., Rahman, S. (eds.) Acts of Knowledge - History, Philosophy and Logic, pp. 207–232. College Publications, London (2009)

18. Primiero, G.: Constructive contextual modal judgements for reasoning from open assumptions. Logique & Analyse 220 (2012)
19. Primiero, G.: Offline and online data: on upgrading functional information to knowledge. Philosophical Studies (2012), doi:10.1007/s11098-012-9860-4.
20. Primiero, G., Jespersen, B.: Two Kinds of Procedural Semantics for Privative Modification. In: Nakakoji, K., Murakami, Y., McCready, E. (eds.) JSAI-isAI 2009. LNCS (LNAI), vol. 6284, pp. 252–271. Springer, Heidelberg (2010)
21. Primiero, G., Taddeo, M.: A modal type theory for formalizing trusted communications. Journal of Applied Logic 10(1), 92–114 (2012)
22. Rotstein, C., Winter, Y.: Total adjectives vs. partial adjectives: scale structure and higher-order modifiers. Natural Language Semantics 12, 259–288 (2004)
23. van der Schaar, M.: The cognitive act and the first-person perspective: an epistemology for constructive type theory. Synthese 180(3), 391–417 (2011)
24. van der Schaar, M.: Wooden horses and false friends; on the logic of adjectives (to appear)
25. Sørensen, M.H., Urzyczyn, P.: Lectures on the Curry-Howard Isomorphism. Studies in Logic and the Foundations of Mathematics, vol. 149. Elsevier (2006)
26. Sundholm, B.G.: Mind your P's and Q's. On the proper interpretation of modal logic. In: Childers, T., Majer, O. (eds.) The Logica 2002 Yearbook, pp. 101–111. Filosofia, Czech Academy of Science (2003)
27. Winter, Y.: Cross-categorical restrictions on measure phrase modification. Linguistics and Philosophy 28, 233–267 (2005)

An Outline of a Dynamic Theory of Frames

Ralf Naumann

Institut für Sprache und Information
Universität Düsseldorf
Germany

Abstract. In this article we present an extension to the theory of frames developed in Petersen (2007). Petersen's theory only applies to concepts for persistent objects like trees or dogs but not to concepts for actions and events that are inherently dynamic because they describe factual changes in the world. Basic frames are defined as Kripke-models. In order to represent the dynamic dimension one needs in addition both combinations of and transformation between such models. Combinations of Kripke-models are used for *temporalization* (representing stages of objects and the temporal development of events) and *refinement* (representing the internal structure of objects). Such combinations are defined using techniques from Finger & Gabbay (1992) and Blackburn & de Rijke (1997). Transformations between Kripke-models are used to represent the factual changes brought about by events. Such transformations are defined using strategies from Dynamic Logic and Dynamic Epistemic Logic, Van Benthem et al. (2005).[1]

Keywords: dynamic frame theory, Kripke models, combining systems, simulations, dative alternation.

1 Introduction

Barsalou (1992, 1999), following the work of Fillmore (1982), extended Fillmore's concept of frame, arguing that it is the fundamental representation of knowledge in human cognition which underlies the content and structure of concepts. He defines a frame as a recursive attribute-value structure in which attributes denote properties of objects, like colour or height, whose manifestations are represented by the values, e.g. blond or black for the colour of the hair of a person. Values need not be atomic but can be frames themselves. For example, the attribute *BIRTH* of a person can be a frame consisting of attributes like *DATE* and *PLACE*. A formal theory of frames in the sense of Barsalou was developed in Petersen (2007). Extending the notion of a typed feature structure in Carpenter

[1] The research presented here has been supported by the Collaborative Research Center 991 funded by the German Research Foundation (DFG). I would like to thank an anonymous reviewer and Rainer Osswald for helpful and stimulating comments and criticisms.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 115–137, 2013.

(1992), she defines a frame as a directed connected graph satisfying the following three conditions: (i) there is a central node (depicted by a double border), (ii) each node is of a particular type indicating the sort of the value and (iii) arcs are labeled with functional attributes.[2] In Figure (1) the (simplified) frame for the sortal concept *tree* is given (see Petersen & Osswald 2012 for details). Such frames will be called *Petersen-frames*.

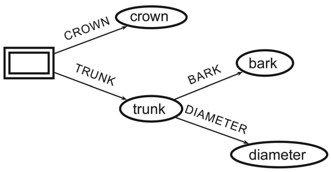


Fig. 1. Frame for the sortal concept *tree*

Both Barsalou's approach and Petersen's formalization of it only apply to static concepts which are related to persistent objects like trees, mothers and dogs and which are usually expressed in natural language by common nouns. What is missing is an account of dynamic concepts that express changes and which are related to non-stative or action verbs like *write*, *go to*, *kick* or *arrive*. Both types of concepts are not independent of each other. Static concepts provide links to dynamic concepts, e.g. in the form of actions that can be performed by or with a given type of objects. Dynamic concepts for action and events, on the other hand, are 'applied' to persistent objects having particular properties the values of which get changed during the execution of the action or the occurrence of the event.

A central question for a frame theory therefore is: is it possible to model dynamic concepts in the same (or at least similar) way as static concepts? Löbner (2011) distinguishes the two options below.

1. There is a uniform format underlying both static and dynamic frames.
2. The format of frames is used only for static concepts. Dynamic concepts are conceived of as procedures operating on static frames. On this view a theory of frames consists of a space of (static) frames and a set of dynamic operations on, or in, that space.

The two options are illustrated by the concept for *x going from A to B*. One attempt at modelling this concept as a frame is depicted in Figure (2).

The central node represents the event of going. This event is related by three attributes (modeling thematic relations or roles) to three objects x, A and B, respectively. However, there is no relation between the Source and Goal attributes and the Location attribute of the actor at the beginning and the end of the

[2] Thus, in contrast to Carpenter, Petersen does not require that a feature structure be rooted. A second difference, neglected in the present context, is that attributes are defined as a special kind of type; see Petersen (2007) for details.

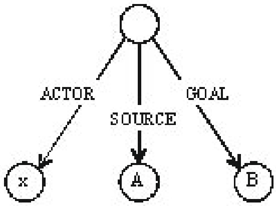


Fig. 2. Fillmore frame

going-event, respectively. According to option (1), this lack can be overcome in the following way. Frames for events and actions are defined in such a way as containing attributes representing temporal transitions, e.g. a change of the actor's location from value A to value B. This is shown in Figure (3).

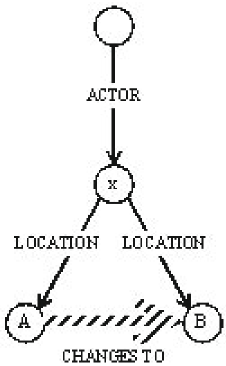


Fig. 3. Frame with temporal transition

By contrast, using option (2), events are modeled as procedures that operate on (static) frames. In our example such a procedure would map a frame for the object x with the attribute Location having the value A to a frame for the same object with the attribute having the value B. This is depicted in Figure (4) on the next page.

As noted by Löbner (2011), option (1) calls for an inventive new account of frames with the main problem being the representation of time within frames for verbs. The main difficulty for the second option consists first in deriving frames from a procedural verb representation and second in formally defining the procedures themselves.

The rest of the paper is organized as follows. In the next section the theory is presented in an informal way. In the following section a possible formalization both of the structures and of the combinations and mappings between them is

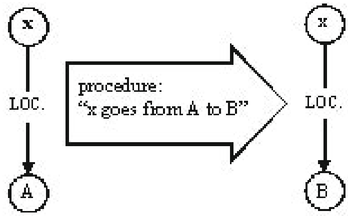


Fig. 4. Transition between frames

given.[3] In the last two sections the theory is related to the notion of simulation in Barsalou (1999) and to the frame-based analysis of *send*-verbs in Kallmeyer & Osswald (2012).

2 Static and Dynamic Frames

A key problem with the frames depicted in Figures 2-4 in the preceding section is that they (i) capture only one aspect (frame in Figure 2), (ii) try to capture different aspects in a single frame (frame in Figure 3) or (iii) that some aspect is not represented as a frame at all (frame in Figure 4). Thus, they either achieve too little or they try to achieve too much within single frames.

By contrast, the approach to a dynamic theory of frames to be developed in the sequel distinguishes different levels of frames: besides basic frames, e.g. Petersen-frames, one also needs combinations of and transformations between frames. One thus arrives at a hierarchical structure in which the different aspects of the dynamic dimension of concepts for actions and events are modelled.

The basic level are frames like those shown in Figures 1 and 2. They are used to model the static dimension, e.g. the relation between an event and its participants. Petersen-frames already represent a second level of frames: refinement. They can be taken to give a detailed or internal (though partial) representation of an object whose type is determined by the central node. On this view, refinement is a relation between a persistent object, taken as an atom, and a representation of it where it is seen as having various properties having certain values. Refinement is a first step to model the dynamic aspect. Action and events change a particular aspect (or particular aspects) of an object, for example its volume or its degree of dryness, leaving other aspects unchanged. Different types of events change different aspects of the same object (dry a shirt vs. send a shirt to Mary). It is therefore necessary to represent those properties the values of

[3] Due to lack of space the discussion is restricted to the question of what structures and combinations between them are needed. The equally important issues of how those structures are used to construe the meanings of verbs and of what are appropriate logics (languages) to talk about those structures must be left to another occasion. Preliminary results can be found in Naumann (2012a,c) and section (5) below.

which get affected during the event.[4] However, refinement alone is not enough for modelling the dynamic dimension because one needs two different representations of the same object: one at the beginning of the event and a second one at its end.[5] Thus, events must be represented in their temporal evolution (occurrence) and Petersen-frames must be related to (appropriate) stages of those evolutions. This level of frames is called *temporalization.* Finally, a fourth level is needed, which represents the dynamics proper, so to speak. At that level the transition (transformation) between one Petersen-frame to the next Petersen-frame in the temporal evolution of an event is interpreted as the result of an update construction between the first Petersen-frame and a particular type of event frame. In the remainder of this section, we will give an informal account of the theory.

Concepts for persistent objects like trees, dogs and mothers, as modelled by Petersen-frames, only have a static dimension in the following sense. These concepts describe what is the case (holds) for an object at a particular moment in time. There is therefore no explicit temporal (or dynamic, change) component. Such concepts (recursively) relate a central node, which determines the type of the object, to a set of attribute-value pairs that represent properties of the object and their values at a particular moment of time, respectively. An example of such a frame is given in Figure (1) above for the (sortal) concept *tree.*

Similar to concepts for persistent objects, concepts for actions and events have a static dimension. This dimension represents the relation between an action/event and the (persistent) objects participating in it. In the present context these relations are defined in terms of thematic roles like *Actor*, *Theme* or *Recipient.* These relations between an event and its participants too are static in the sense that they do not change during the occurrence of the event. This component can be represented by frames of the type in Figure (5), which are a variant of a Fillmore frame. Such frames will be called *static event frames (SEFs).*

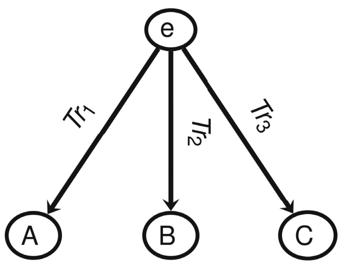


Fig. 5. Static event frame

[4] As will be shown in section (4), this aspect is also closely related to the notion of a *simulation.*

[5] For many types of events it is also important to model what happens *during* the occurrence of the event; see below for details.

In an SEF both the event and the objects it is related to are taken as atomic, undivisible entities with no internal structure. In order to arrive at the dynamic dimension of the concepts for actions and events both types of entities, persistent objects and actions/events, must be assigned some kind of internal structure, which is then used in representing those entities over (or in) time (temporalization). The key idea to be used is the concept of *refinement*.

The idea of refinement can be illustrated by the following example taken from Blackburn & de Rijke (1997). When working with a graphical user interface on a computer, the desktop usually contains a number of icons. These icons are just 'blobs' as long as the user does not click on them. However, when one wants to perform a certain task, one gets a more refined view. One double-clicks on an icon and as an effect one zooms into another level of structure.[6] So in a refinement there are two levels which are linked by a relation. At the higher level an entity is seen as an atom without internal structure, whereas at the lower level one gets a more detailed (or fine-grained) view of the entity. For persistent objects this means that information about (some of) their properties and their corresponding values (at a particular moment in time) is provided. In the case of actions and events the information concerns the temporal (dynamic) development of the event in time. Thus, in a frame theory refinement both for persistent objects and events has to do with a relation between those entities and time.[7]

For persistent objects, a Petersen-frame is already a refinement of the object at the central node. For actions and (non-boundary) events one has to take into consideration that, contrary to persistent objects, they occur in time. They have a beginning (left boundary) and an end (right boundary) point.[8] During any proper part of the time span corresponding to those two points only part of the event exists (or occurs).[9]

There are different ways of how the boundary of an event can be defined. For example, it can be taken to be a time point. An alternative view, investigated in Pinon (1997), consists in taking the beginning and the end of an event to be a special sort of event, called *boundary events*, which have no temporal extension in the sense that their run-time is a singleton. In the sequel we will adopt this latter

[6] An example from linguistics, also discussed in Blackburn & de Rijke (1997), is GPSG. In this grammar formalism feature structures are used to refine the notion of grammatical category. Nodes in a parse tree are not just decorated with atomic information about categories (like np or vp, for example). Rather these atomic categories become refined by being assigned a feature structure that contains information about various subatomic features and values.

[7] For persistent objects one may speak of a temporalized perspective or a stage (for the latter, see Osswald & Van Valin 2012 and the references cited in that article).

[8] Thus, we do not consider infinite events.

[9] The difference to persistent objects is the following. Although persistent objects have a beginning (say birth or creation) and an end (say death or destruction) too, they do not occur in time in the sense that for a proper part of their lifespan only a proper part of them exists. Rather, they are completely present at any moment during that time span. For more on this distinction, see for example Wiggins (1980).

alternative since by using it we do not need to explicitly introduce a separate domain of time points (or, alternatively, of time intervals).[10]

One way of modelling this relation between an event and its left and right boundary as a frame is given in Figure (6), where α and β are attributes that are interpreted by two functions assigning to an event its left and right boundary, respectively.

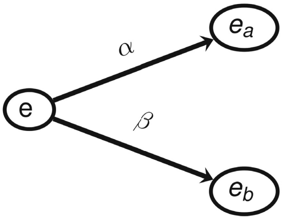


Fig. 6.

An alternative way, to be adopted in the sequel, consists in assigning to an event a frame which is a linear (or sequential) transition structure in which the left and right boundary are nodes that are linked by the event itself. Such frames will be called *Temporal Event Frames*.

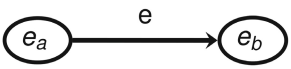


Fig. 7. Temporalized event frame

For events, the above way of refinement already amounts to temporalization because the internal structure represents the event as occurring in time. Temporalizing their participants requires another strategy. A first attempt at defining temporalization in the domain of persistent objects could consist in assigning to each such object in an SEF a Petersen-frame. However, this attempt fails for at least two reasons. First, it only captures the relation between the object and its set of property-values pairs at a particular moment in time during the occurrence of the event. Second, since the root of an SEF has, at least in general, an extended temporal extension (a proper interval), it would be unclear to what stage of the event the Petersen-frame should be applied. Thus, assigning to each leaf in an SEF a Petersen-frame by a refinement-relation is not what we are looking after.

[10] Though we occasionally will refer to time points in the sequel. One way of relating boundary events to a flow of time consisting of time points is to assume that each boundary event is assigned exactly one time point as its run-time.

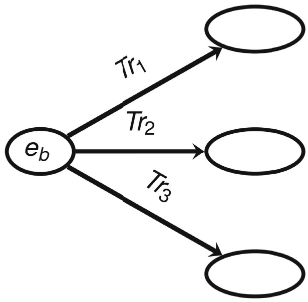


Fig. 8. Temporalized static event frame with Petersen-frames as values

The key observation is that a *temporalized SEF (t-SEF)* has to be assigned not to the event itself but to its temporalized representation, that is, to its left and right boundary e_a and e_b, respectively. For example, the t-SEF assigned to e_a represents what holds of the objects participating in the event e (modelled by the SEF assigned to e) at the left boundary of e. Thus, the general idea is that for a boundary event a thematic relation links it to a refined (or temporalized) representation of an object to which the event of which it is a boundary is related by the same thematic relation. This representation models the properties of the object at a particular moment in time, namely at that moment corresponding to the boundary. If such a t-SEF is assigned to each node in a TEF, one gets a description of how the object develops (or changes) during the occurrence of the event. This is shown in Figure (8) where the values are Petersen-frames.

Thus, the general architecture is the following. Non-boundary events are related to persistent objects represented as atomic objects. This relation is captured by SEFs. By contrast, boundary events are related to a more fine-grained (or temporalized) representation of the object. This relation is captured by t-SEFs.

The dynamic dimension of concepts for actions/events can now be defined as an operation on (or transformation between) t-SEFs: the t-SEF assigned to e_a is transformed by the event e into the t-SEF assigned to e_b. If this operation is to be modeled by a frame two questions that have to be answered are: (i) what is represented by such a frame? and (ii) how is the operation between a t-SEF and this type of frame be defined? Beginning with the first question, such a frame has to represent what change is brought about by the event. One way of how this can be done consists in specifying what has to hold for the event to occur (its precondition) and what holds after the event has occurred (its postcondition).[11]

[11] In Dynamic Logic, a program has both a (weakest) precondition specifying under what conditions this program can be executed and a (strongest) postcondition specifying what holds after a (terminating) execution of the program; see e.g. Harel et al. (2000) for details.

For example, the change expressed by *become dry* requires that at the right boundary of the event the object undergoing the change (say some piece of clothing like a shirt) is dry, i.e. it has the maximal value on the dryness scale, say 0, whereas at the left boundary of the event this object had a non-zero degree of dryness. This kind of change is definite in the sense that a unique value for the right boundary of the event is determined. By contrast, the change expressed by *become drier* is indefinite in the sense that the only condition imposed on the right boundary of the event is that the degree of dryness be lower than the degree at the left boundary. The pre- and postcondition are not independent of each other. First, they are both formulated with respect to the same property of an object participating in the event and second the values of this property must be distinct.

An answer to the second question must account for the following constraints. First, the (input) t-SEF at e_a must satisfy the precondition imposed by e because otherwise the event e cannot occur. Second, the transformation consists in assigning to the property that gets changed a new value, namely the value resulting from the change brought about by the event. The frame corresponding to the dynamic dimension of the concept for an action/event can now be defined as having two attributes corresponding to the pre- and postcondition, respectively. Such frames will be called *update frames*.

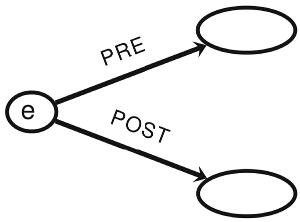


Fig. 9. Update frame

Both the pre- and postcondition can be given as formulas of the language that is used to talk about t-SEFs. Applied to the example of *becoming dry*, one gets (1), yielding the frame representation in Figure (10). Here we use modal logic as a language to talk about frames. See section (3) and Blackburn (1993, 1994) for details.

(1) a. precondition : <THEME><DRYNESS>¬ 0
    b. postcondition: <THEME><DRYNESS>0

The operation yields an output t-SEF only if the test whether the input t-SEF satisfies the precondition imposed by the update frame succeeds. The resulting t-SEF is constructed from the input t-SEF by a substitution (or assignment) operation: the property affected by the change is assigned a new value (or, the old value (at e_a) is replaced by a new value given by the postcondition). This

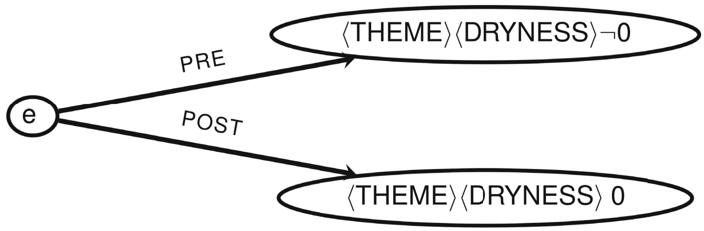


Fig. 10. Update frame with values

is shown in Figure (11). Applied to (1), one gets: Tr = THEME and PROP = DRYNESS (Tr = thematic relation). The value v can be any value other than 0 on the dryness-scale so that the precondition $v^* \neq 0$ is satisfied. The result of the update construction, i.e. of applying the update frame with root e to the t-SEF on the left, is the t-SEF on the right where the value v is replaced by v'.

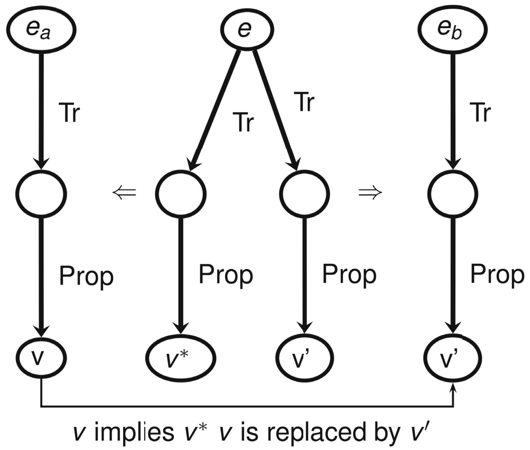


Fig. 11. Update construction

Summarizing the informal account of this section, one has: Concepts for actions and events consist of

- a static dimension (modelled by a static event frame SEF)
- a temporal dimension (modelled by a temporal event frame TEF)
- temporalized static event frames (t-SEF), which combine the static and the temporal dimension
- an update dimension (modelled by an update frame UF)
- an update construction which maps t-SEFs and an update model to t-SEFs

3 A Possible Formalization

In this section we will sketch a possible formalization of the ideas presented in section (2). We will concentrate on defining the different structures and their combinations. Concerning the question of which logics (languages) are appropriate for those structures we only give an example of a simple modal logic. The discussion of more expressive logics (which are certainly needed) must be left to another occasion. We will begin by giving the definition of a feature structure in Blackburn (1993).

Definition 1 (Feature Structure). *A feature structure of signature $<L,A>$ with L a non-empty set of possible labels and A a non-empty set of (atomic) information, is an ordered triple $< N, \{R_l\}_{l\in L}, \{Q_a\}_{a\in A} >$, where N is a non-empty set of nodes; for each $l \in L$, R_l is a binary relation on N that is a partial function; and for each $a \in A$, Q_a is a unary relation on N.*

According to this definition, feature structures are a special kind of Kripke models consisting of a set of nodes together with a collection of binary relations and a collection of unary relations on these nodes. A simple modal logic for talking about such structures can be defined in the usual way. The standard truth definition is as follows.

1. $M, n \models p_a\, iff\; n \in Q_a$
2. $M, n \models \neg\phi\, iff\; not M, n \models \phi$
3. $M, n \models \phi \vee \psi\, iff\; M, n \models \phi\; or M, n \models \psi$
4. $M, n \models < l > \phi\, iff\; \exists n'\; with\; (n, n') \in R_l\; and\; M, n' \models \phi$

Kripke-models are basically used to represent ontologies like those of persistent object (Petersen-frames) and action/events. Many-sorted Kripke-models are used to represent relations between different ontologies where the entities are taken as atoms. In Petersen-frames the domain is a (non-empty) set of persistent objects. In addition, there is a distinguished node, called the central node (cn)[12] L is a set of (functional) relation symbols like *CROWN* or *TRUNK* and A is a set of (unary) predicate symbols like *diameter* or *trunk*.

The definition of an event-structure is given below.

Definition 2 (Event Structure). *An event structure E is a sixtuple $< E_{nb}, E_b, \alpha, \beta, \sqcup, \sqsubseteq >$ where (i) E_{nb} is a non-empty set of non-boundary events, (ii) E_b is a non-empty set of boundary events, (iii) α (β) is a total function from E_{nb} to E_b that assigns to each non-boundary event e its left (right) boundary $\alpha(e)$ ($\beta(e)$).[13] $\sqcup$ and $\sqsubseteq$ are the usual join and part-of relations on the domain of non-boundary events.*

[12] Missing from this definition is the condition that the interpretation of atomic formulas is constrained by a sort-hierarchy.

[13] It is possible to extend the definition to boundary events by setting $\alpha(e_b) = \beta(e_b) = e_b$, i.e. each boundary event is its own left and right boundary.

Static event frames are defined as two-sorted Kripke structures with two disjoint domains: a domain of non-boundary and boundary events and a domain of persistent objects.

Definition 3 (Static Event Frame). *A static event frame (SEF) of signature* $<T>$ *with* T *a non-empty set of thematic relation symbols is a triple* $< E, O, \{TR_t\}_{t \in T} >$, *where* E *is a non-empty set of non-boundary or boundary events,* O *is a non-empty set of persistent objects and each* TR_t *is a partial function on* $E \times O$.

Instead of having two domains; E and O, it is possible to work with a single domain W and two special unary constants, say *event* and *object*, in the underlying language. These constants are defined in such a way that the domain W is partitioned into two disjoint subdomains. Formally, this constraint can be enforced by axioms like (i) $event \vee object$ and (ii) $\neg(event \wedge object)$.

In order to model the dynamic dimension, it is necessary to use in addition to Kripke-models various operations on such models. A first operation that will be used is the combination of Kripke-models in the sense of Finger & Gabbay (1992) and Blackburn & de Rijke (1997). If $\mathcal{A}$ and $\mathcal{B}$ are two classes of Kripke-models (or, more generally, structures), and $\mathcal{Z}$ is a collection of relations between the elements of $\mathcal{A}$ and those of $\mathcal{B}$, the triple $< \mathcal{A}, \mathcal{Z}, \mathcal{B} >$ is called a *trio* with $\mathcal{A}$ and $\mathcal{B}$ the left and right continent, respectively, and $\mathcal{Z}$ the bridge between the two continents. In the present context, trios are used for refinement and temporalization.

In a refinement relation entities belonging to the domains of the elements of the left continent $\mathcal{A}$ are assigned a structure belonging to the right continent $\mathcal{B}$. For action and events, refinement will be defined as the relation between an action/event and its temporal developments (or evolutions). For example, an event of drying can be decomposed into its inchoation (the beginning of the drying), followed by a development portion (the theme becoming less and less drier) and a culmination (the theme is dry) followed by a consequent state during which the theme remains dry.[14] For events described by a verb like *send* a possible temporal decomposition consists of an action undertaken by the actor causing a movement of the theme to the recipient as its destination.[15] In both cases the occurrence of an event is described as an ordered sequence of different phases.[16]

Temporal event frames (TEFs) are defined in terms of the sequential decomposition of a non-boundary event e.

Definition 4 (sequential Decomposition of a Non-boundary Event). *A sequential decomposition (SD) of a non-boundary event e is a finite sequence of*

[14] Such a decomposition is similar to the concept of a nucleus structure in Moens & Steedman (1988).

[15] Such a temporal decomposition is similar to event templates like *x CAUSE z GO_ TO y)*. This is not the only decomposition for *send*; see section(5) for details.

[16] Thus, in the domain of action and events refinement can also be regarded as a special form of temporalization.

non-boundary events $e_1 ... e_n$ *for some* n *s.t. (i)* $e = \sqcup E$ *and* $E = \{e_1, ..., e_n\}$*, (ii)* $\alpha(e_1) = \alpha(e)$*, (iii)* $\beta(e_n) = \beta(e)$ *and (iv)* $\beta(e_i) = \alpha(e_{i+1})$ *for* $1 \leq i < n$*.*

Two subtypes of SD of non-boundary events are distinguished: *type-identical* and *non-type identical* SDs. For a type-identical SD, each event in the sequence is of the same type as e. Thus, if P_v is the set of all events of type v (e.g. if v = *dry*, P_v is the set of all drying-events), then if $e \in P_v$ one also has $e_i \in P_v$ for $1 \leq i \leq n$. By contrast, for a non-type identical decomposition, the e_i are not of the same type as e. This is the case for events like *sending*, for instance, where the event is decomposed into a causing event and a resulting effect event, which both are not sendings.

A second dimension with respect to which SDs can be classified is the way the postcondition is evaluated on it. To take a drying-event of a shirt as an example: the shirt is only dry at the left boundary of the event. Thus, the postcondition only holds at the end of the event but at no stage preceding it. This dimension can be represented by using program constructs from Dynamic Logic. For example, the above example of a drying in which the postcondition only holds at the right boundary of the event, can be modelled by a *while-loop*. Other types of how the postcondition is evaluated can be defined in terms of (combinations of) other programs. How this can be done, in particular for aspectual distinctions, has been shown in Naumann (2001).

The SD of an event is in general not unique. It is always possible to set $\sqcup E = \{e\}$. This is the coarsest SD. For a drying-event, the finest SD consists of atomic drying events, i.e. one has: $\forall e'(e' \sqsubset e_i \rightarrow e' \notin P_{dry}))$.

Given the definition of an SD of an event, a temporal event frame is defined as follows.

Definition 5 (Temporal Event Frame). *A temporal event frame (TEF) is a quadruple* $< E_b, E_{nb}, R, e >$ *s.t. (i)* $E_b \sqcup E_{nb} = E$ *is a sequential decomposition of e and (ii)* R *is defined by* $R(e_a, e_b, e)\, iff\, \alpha(e) = e_a \wedge \beta(e) = e_b$*.*

Finally, refinement for events is defined as given below.

Definition 6 (Refinement for Events). *Refinement for events is a trio* $< E_{nb}, Z, \{TEF_q\}_{q \in Q} >$ *s.t. (i)* E_{nb} *is a non-empty set of non-boundary events, (ii) each* TEF_q *is a temporal event frame and (iii)* Z *is defined by:* $(e, TEF_q) \in Z$ *iff* TEF_q *is a sequential decomposition of* e*.*

Temporalization is used for the domain of persistent objects. Although elements of this domain persist through time, they usually undergo changes. For example, a wet shirt becomes dry or Bill gets sent a book by John and therefore now possesses this book whereas the book changed its location. Temporalization is defined in two steps. First, a persistent object is assigned a frame which partially describes what holds at the object at a particular stage during the occurrence of an event. Taken in isolation, this step can be seen as an instance of refinement because the object is described as having an internal structure given by the property/value pairs of the frame. By repeating this assignment for each phase of the event, one arrives at a sequence of frames for the object which depicts its temporal development during the occurrence of the event, in particular how some of its properties change as an effect of the object participating in the event. The first step is captured by temporalized static event frames.

Definition 7 (Temporalized Static Event Frame). *A temporalized static event frame (t-SEF) is triple* $< SEF, Z, \{P_f\}_{f \in F} >$ *where* SEF *is a static event frame based on a domain of boundary events and a domain* O *of persistent objects,* $\{P_f\}_{f \in F}$ *is a set of Petersen-frames having cardinality* $|O|$ *and* Z *is an injective function that assigns to each element of* O *a Petersen-frame from* $\{P_f\}_{f \in F}$.

The second step consists in assigning to each $e_x \in E_b$ of a TEF its corresponding t-SEF. More formally: For a given TEF, let Z' be a function from E_b that assigns to $e_x \in E_b$ its corresponding t-SEF. The corresponding trio is then defined by $< TEF, Z', range(Z'(E_b)) >$

Refinement and temporalization relate structures to each other. In order to model the dynamics proper, mappings (or transformations) between Kripke-models are needed. For Kripke-models, either the domain (the set of states), the accessibility relations or the valuation can be changed. For modelling the change brought about by an action or an event, only the valuation needs to be changed. Changes in the valuation are defined using the notion of a *substitution*. The following definition is taken from Van Benthem et al. (2006). Let $\mathcal{L}$ be an appropriate language for talking about Petersen-frames.

Definition 8 (Substitutions). $\mathcal{L}$ *substitutions are functions of type* $\mathcal{L} \rightarrow \mathcal{L}$ *that distribute over all language constructs, and that map all but a finite number of basic propositions to themselves.* $\mathcal{L}$ *substitutions can be represented as sets of bindings* $\{p_1 \mapsto \phi_1, ..., p_n \mapsto \phi_n\}$ *where all the* p_i *are different. If* σ *is a substitution, then the set* $\{p \in P \,|\, \sigma(p) \neq p\}$ *is called its domain, notation dom(*σ*). The identity substitution is denoted by* ϵ*.* $SUB_{\mathcal{L}}$ *is the set of all substitutions.*

Using the notion of a substitution, the notion of a *Petersen-frame under a substitution* is defined as follows.

Definition 9 (Petersen-Frame under a Substitution). *If* $P =< W, V, \{TR_t\}_{t \in T} >$ *is a Petersen-frame and* σ *is a substitution (for an appropriate language* $\mathcal{L}$*, then* V_P^σ *is the valuation given by* $\lambda p.[[\sigma(p)]]^M$*. In other words,* V_P^σ *assigns to* p *the set of worlds* w *in which* $\sigma(p)$ *is true. For* $P =< W, V, \{TR_t\}_{t \in T} >$*,* M^σ *is the model given by* $P =< W, V_M^\sigma, \{TR_t\}_{t \in T} >$.

The idea underlying a substitution and a Petersen-frame under it can be illustrated by the following example. Let *is_zero* be an atomic proposition that is true of a node of type *dryness* just in case the value of the corresponding property of an object (say a shirt) is the maximal element of the dryness-scale (i.e. 0). Suppose furthermore that there is a single node, say n, in the Petersen-frame of that type. Then for a drying-event the input Petersen-frame M has $\neg$*is_zero* for the node n that is the value of the path $< THEME >< DRYNESS >$ so that

$V(is\_zero) = \emptyset$ in the frame M. The required substitution is $\sigma(is\_zero) = \phi$ with $V(\phi) = \{n\}$. If $dom(\sigma)$ is a singleton, this means that there is exactly one postcondition. If there is more than one postcondition, $|dom(\sigma)| > 1$ (this is the case for events like sending; see section(5) for details).

Update models for events specify the pre- and postconditions for each event in the model, where the latter are defined using the notion of a substitution.

Definition 10 (Event Update Model). *An event update model with language $\mathcal{L}$ is a triple $< E, pre, post >$ where (i) E is a non-empty set of non-boundary events, (ii) $pre : E \to \mathcal{L}$ assigns a precondition to each event and (iii) $post : E \to SUB_{\mathcal{L}}$ assigns an $\mathcal{L}$ substitution to each event.*

Update execution is now modelled by the following construction.

Definition 11 (Update Execution). *Given a Petersen-frame $P =< W, V, \{Tr_t\}_{t \in T} >$ with central node $w \in W$ and an update model $< E, pre, post >$, with $P, w \models pre(e)$, the update triggered by e in P, w is the model M^{σ}.*

Thus, for a single postcondition the occurrence of an event e has the effect of transforming the Petersen-frame at its left boundary to another Petersen-frame at its right boundary that differs from the former only in the value that is assigned to the node specified by $e's$ postcondition.

A schematic overview of the theory is depicted in Figur (12) (R = refinement; T = temporalization). The event e at the root of the SEF on the left is refined to the TEF on the right, yielding a temporal sequential decomposition of e. Each boundary event e_x in this decomposition is the root of a t-SEF the Petersen-frame of which gives a (partial) representation of the object bearing Tr to e (or e_x) at the stage e_x of the event e. On this perspective the object $Tr(e)$ gets temporalized. Viewed from $Tr(e)$, the Petersen-frame is a refinement, describing one of its stages. Each e_i brings about a partial change with respect to $Tr(e)$. This change is modelled by the update construction $\otimes$ between the Petersen-frame at $\alpha(e_i)$ and the update model corresponding to e_i, yielding the (updated) Petersen-frame at, modelled by a Petersen-frame under substitution, $\beta(e_i)$. When taken together, one gets the overall change effected by the event e (w.r.t. $Tr(e)$). The correspondinmg update construction is shown at the bottom of the figure.

Figure (12) also shows how it is possible to model the meaning of a verb in a single (higher-order) frame. Note that for the meaning of a verb the sequential decomposition plays no role in semantically classifying the verb. What is needed, instead, is a more abstract, higher-level characterization of that sequence. One way of arriving at such an abstraction is to use the notion of a nucleus structure (Moens and Steedman 1988). Such a structure defines a sequential decomposition of an event in terms of notions like *development portion*, *culmination* and *consequent state*. For example, a degree achievement like *dry* has the nucleus structure depicted in Figure (13).

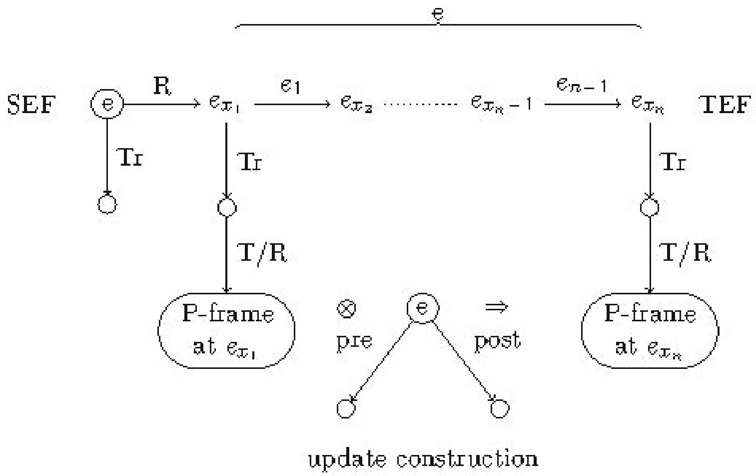


Fig. 12. Schematic overview of the theory

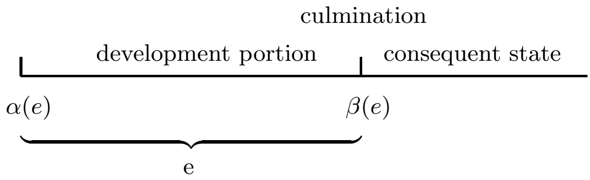


Fig. 13. Nucleus structure for *dry*

It is possible to define the development portion and the culmination in terms of the event e in the following way: the development portion is e without its right boundary and the culmination is the right boundary. At each part of the nucleus structure the property affected by an event of the type denoted by the verb is assigned a particular value. The assignment of these values can be characterized by a formula in Dynamic Logic. For example, for *dry* the postcondition $< THEME >< DRYNESS > 0$ is false for the development portion and true for the culmination and the consequent state.

On this perspective the sequential decomposition of e is replaced by a particular nucleus structure characterizing the verb (or its aspectual class), i.e. the TEF is a nucleus structure. Each part of this structure is the root of a SEF so that it is assigned a Petersen-frame displaying the value of the property that gets changed. This definition of a sequential decomposition corresponds to the second dimension discussed above at which such a decomposition can be classified. Thus, instead of working with a sequential decomposition, one uses a particular dimension at which this decomposition can be characterized.[17]

[17] For details on this perspective, see Naumann (2012c).

4 Simulations in a Dynamic Theory of Frames

In this section we will relate our dynamic theory of frames to Barsalou's main motivation for introducing the frame concept. According to Barsalou (1999), perceptual representations rather than amodal logic-based propositions are the building blocks of cognition. Perceptual representations (or simulations) are the key concept in a theory of grounded cognition. During the interaction with the world traces of perceptions and experiences of objects and events become associated with words (e.g. verbs and nouns) and are stored in the memory repository of the brain. During language comprehension those traces are retrieved from memory and are reactivated to produce a perceptual representation (or simulation) of the situation described by the sentence or discourse. For example, when reading the sentence *The ranger saw an eagle in the sky* comprehenders will simulate the eagle as having its wings outstretched (as opposed to having them drawn) because it was flying and not, say, perched in a nest. In a series of experiments Zwaan and his colleagues (Zwaan & Stanfield 2001, Zwaan et al. 2002) tested this approach to language comprehension They predicted that there should be a mismatch effect when subjects are presented with the above sentence followed by a picture of an eagle with its wings drawn. This hypothesis was tested in two experiments. After reading a sentence, comprehenders were presented with a line drawing of the object in question. In the first experiment they had to judge whether the object had been mentioned in the sentence whereas in the second experiment they had to simply name the object. The authors found that in both experiments responses were faster when the shape of the pictured objects matched the shape implied by the sentence compared to when there was a mismatch.

These experimental findings can be taken as evidence that the amodal (propositional) representation of the sentence *The ranger saw an eagle in the sky* given in (2) does not capture the fact that the eagle is represented with its wings outstretched.

Despite its weakness the logic representation in (2) captures the aspect that the eagle is the constant theme of the seeing event and the constant actor of the flying event. This aspect is important for the identity of persistent objects over time. Consider e.g. the short discourse in (3).

(2) The eagle was flying in the sky. After a few minutes it arrived at its nest and began to feed its offspring.

In this discourse the simulated shape of the eagle changes as a function of its location and, more importantly, the action it is involved in. But despite those differences it is the *same* eagle that is talked about and that has to be simulated (a fact which is, among other things, reflected in the use of the pronouns *it* and *its*).

In our approach the aspect of identity over time (or across sentences) is captured by SEFs which relate an event to a set of objects taken as atoms that persist over time. The aspect of a modal perceptual representation is captured by t-SEFs in which the persistent object is related to a boundary event and represented as having certain properties together with the corresponding values that can undergo a change due to external forces and events.

The results of Zwaan and colleagues also show that in the Petersen-frame of an object in a t-SEF not all of the object's properties need to be represented (or activated) but only a certain subset consisting of those properties that are related to the type of event described by the sentence. Thus, we get the following thesis:

> What is represented in a Petersen-frame of an object in a t-SEF depends on the type of the event that is described.

The way an object participating in an event is simulated depends, at least partly, on the type of the event. For example, although all three events in the discourse in (3) (flying, arriving and feeding) are related to the same eagle, the Petersen-frames used in the corresponding t-SEFs differ. For example, the shape of the eagle's wings will be simulated differently in the flying and in the arriving event (difference in the value of the attribute describing the shape of the wings of the eagle)[18] and in the feeding event the shape of the wings need not be represented at all. Instead the shape of its head (mouth) will be represented simulating the feeding activity.[19] Thus, a verb not only imposes a constraint on the type of object (e.g. the ability to fly) but it also primes the values of some of its properties, like the shape of its wings, for example. As a consequence, the semantic representation of a verb cannot be restricted to SEFs but must in addition also contain t-SEFs in which those constraints on how the object is simulated during the event are expressed.

The above considerations show that simulation and refinement are closely related. Refinement is needed in order to have access to those aspects of an object that change due to the event. But such aspects are also needed to account for the way an object is simulated during sentence comprehension. With respect to the different levels of frames, one gets the following correlations:

> SEF → represented as an atom → identity over time
> t-SEF → represented as having an internal structure → simulation + refinement + temporalization

The relation between simulation and the theory presented in this article is discussed in more detail in Naumann(2012b,c).

[18] Of course, other aspects of a simulation depend on additional factors like past experiences and/or the preceding linguistic context.

[19] This dependency of how an object is simulated on the object it participates in is one argument for defining t-SEFs with boundary events as roots and not with time points. This makes it possible to have an object being simulated in different ways when it is involved in several events at the same time.

5 The Dative Alternation with *Send* Verbs in English

In this section it is shown how the theory presented in the preceding sections can be applied to the frame-based analysis of *send*-verbs in Kallmeyer & Osswald (2012). In English, *send* occurs both in the double object (DO) and the prepositional object (PO) construction as exemplified by (4).

(3) a. John sent Mary the book.
  b. John sent the book to Mary.

Using decompositional schemas (see e.g. Levin & Rappaport-Hovav 2005), the two interpretations can be represented as shown in (5).

(4) a. $[[x\,ACT]\,CAUSE\,[y\,HAVE\,z]]$
  b. $[[x\,ACT]\,CAUSE\,[z\,GO\,TO\,y]]$

In both cases *send* is analyzed as having a causal component: the actor x does something which causes a change in the theme z. In the DO construction the effect of the causation is a change of possession whereas in the PO construction it is a change of location (the theme is at (or arrives at) the recipient conceived of as the destination). However, neither in the DO nor in the PO construction is the effect lexically entailed as shown by the non-contradictory examples in (6).

(5) a. John sent Mary the book. But she never got it.
  b. John sent the book to Mary. But it never arrived there.

Send only lexicalizes a caused motion towards the destination. By contrast, the arrival at the destination and the change of possession are only prospective (see Kallmeyer & Osswald 2012 and Beavers 2011 for details). Based on this analysis, Kallmeyer & Osswald (2012) propose the frame representation in Figure (14) for *send*.

$$\begin{bmatrix} \textit{sending} & \\ \text{EFFECTOR} & \boxed{8} \\ \text{THEME} & \boxed{9} \\ \text{GOAL} & \boxed{10} \\ \text{CAUSE} & \begin{bmatrix} \textit{activity} & \\ \text{EFFECTOR} & \boxed{8} \end{bmatrix} \\ \text{EFFECT} & \begin{bmatrix} \textit{change-of-loc} & \\ \text{THEME} & \boxed{9} \\ \text{DESTINATION} & \boxed{10} \end{bmatrix} \end{bmatrix}$$

Fig. 14. Lexical frame for *send*

The frame representation in Figure (14) not only captures the fact that *send* expresses a causation whose effect is a change of location but also the fact that the theme (prospectively) arrives at the destination.[20] The other meaning components, that are not lexicalized, are given by constructions which are also modelled as frames. For example, the frame for the DO construction is shown in Figure (15).[21]

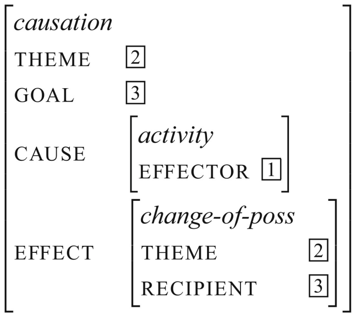


Fig. 15. Lexical frame for the DO construction

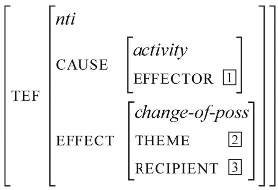


Fig. 16. Alternative DO construction

Let us relate Kallmeyer & Osswald's analysis to the approach developed in the preceding sections. The EFFECTOR, THEME, GOAL part of the frame representation corresponds to an SEF. An event of type *sending* is related to different objects participating in it. The attributes CAUSE and EFFECT describe a particular type of TEF: it is a non-type identical sequential decomposition of the sending event. The types of those decompositions describe particular kinds of programs and are therefore related to the second dimension at which a sequential decomposition can be described (the way a postcondition is brought about). Thus, adapting the Kallmeyer & Osswald analysis to our framework the relevant

[20] Thus, the lexical contribution of *send* already comprises that of the PO construction, except for aspects like those related to the influence of *to* on possible recipients (Rainer Osswald p.c.). For example, consider the difference between *John send the package to Mary/London* and *John gave the package to Mary/*London.*

[21] Thus, Kallmeyer & Osswald's analysis separates the contribution of the lexical meaning from the contribution of a construction in the sense of construction grammar.

part of the representation of the DO construction is as in Figure (16) (where *nti* is the type of non-type identical TEFs).

What is missing in this representation is the constraint on the way an object is simulated. Therefore, the values of the attributes CAUSE and EFFECT should themselves be TEFs in which the values of attributes like EFFECTOR are Petersen-frames capturing the required constraints.

The distinction between the lexical contribution of a verb and those of a construction is related to the fact that an event is in general related not to a single but to a set of TEFs. In the case of *send* at least the following temporal decompositions can be distinguished.

- type 1: caused change of location
- type 2: caused change of location plus a movement to the destination
- type 3: caused change of location plus a change of possession

The first type defines that part of a sequential decomposition which is common to all sending-events, thus reflecting the fact that it expresses the semantic contribution of the verb. The second and third type can be distinguished by making use of the fact that the two kinds of changes are related to different scales: a path-scale for the change of location to the destination and a (binary or simplex) 'possession'-scale for the change of possession (see Beavers 2011). These two types of changes are, however, not independent of each other: whenever the theme arrives at the destination (type 2), the recipient comes to possess it. One way of modelling this relationship consists in having a temporal decomposition which is of both types (or a common subtype of those two types). This combination yields a fourth type.

- type 4: caused change of location plus a movement to the destination plus a change of possession

The lexical contribution of *send* is a type 2 TEF whereas the DO construction contributes a type 3 TEF. In a sentence like *John sent Mary the book* both TEFs are combined to yields a type 4 TEF. For *send* this combination can be defined as follows. Since the 'possession'-scale, on which the type 3 TEF is built, is binary, i.e. there are only two values, the decomposition consists only of e_a and e_b related by the sending-event. At e_a $\neg Have(y, z)$ holds (precondition), whereas at e_b one has $Have(y, z)$ (postcondition). In the combined type 4 TEF the preconditions of both TEFs are combined at e_a and the same is done for the postconditions at e_b.

6 Conclusion and Directions for Future Work

In this article we presented a dynamic theory of frames in which (basic) frames are defined as Kripke-models. In order to model the dynamic dimension of concepts for actions and events, not only basic frames but also combinations of and transformations between such frames must be considered. The dynamics proper is modelled by an update construction between a simple frame and an update frame.

There are at least the following two important issues that haven't been addressed in this article:

1. How is the meaning of verbs built in terms of the different levels of frames? An answer to this question depends, at least in part, on results from psycholinguistics and brain science. Some preliminary results of how a dynamic theory of frames can be combined with recent results in the latter areas are presented in Naumann(2012a,b).
2. What are appropriate logics (or languages) for talking about the structures defined in section(3)? At present, we are using some form of extended modal logics like arrow logic or hybrid logic.

Let me close by mentioning some further questions: (i) How can the concept of a scale be integrated into the theory? (ii) How is the concept of causation modelled in the theory? The update construction used in the theory only changes the valuations of Kripke-models. However, other operations on Kripke-models are possible. Löbner (2011), for example, mentions: adding or deleting attributes (level of Petersen-frames), saturating arguments and relocating the central node. With respect to linguistic applications Löbner refers to metaphors: attributes with values are transferred from one frame to another one, forming a new concept. Other examples include shifts, for example from *play* to *player*, where the central node of a frame for a verb is shifted to the actor node. Finally, in order to model non-factual, for instance, epistemic changes triggered by communicative acts like announcements, further strategies from Dynamic Epistemic Logic must be incorporated into the theory.

References

1. Barsalou, L.W.: Frames, concepts, and conceptual fields. In: Lehrer, A., Kittay, E.F. (eds.) Frames, Fields, and Contrasts, pp. 21–74. Erlbaum, Hillsday (1992)
2. Barsalou, L.W.: Perceptual symbol systems. Behavioral and Brain Sciences 22, 577–660 (1999)
3. Beavers, J.: An aspectual analysis of ditransitive verbs of caused possession in English. Journal of Semantics 28, 1–54 (2011)
4. Blackburn, P.: Modal Logic and Attribute Value Structures. In: de Rijke, M. (ed.) Diamonds and Defaults, pp. 19–65. Kluwer Academic Publishers (1993)
5. Blackburn, P.: Structures, Languages and Translations: the Structural Approach to Feature Logic. In: Rupp, C., Rosner, M., Johnson, R. (eds.) Constraints, Language and Computation, pp. 1–27. Academic Press (1994)
6. Blackburn, P., de Rijke, M.: Zooming in, zooming out. Journal of Logic, Language, and Information 6, 5–31 (1997)
7. Carpenter, B.: The logic of typed feature structures. Cambridge UP, Cambridge (1992)
8. Fillmore, C.J.: Frame semantics. In: Linguistics in the Morning Calm, pp. 111–137. Hanshin Publishing Co., The Linguistic Society of Korea, Seoul (1982)

9. Finger, M., Gabbay, D.: Adding a temporal dimension to a logic system. J. of Logic Language and Information 1, 203–233 (1992)
10. Harel, D., Kozen, D., Tiuryn, J.: Dynamic logic, Cambridge. MIT Press, Cambridge (2000)
11. Kallmeyer, L., Osswald, R.: A frame-based semantics of the dative alternation in lexicalized tree adjoining grammars, Paper Presented at Empirical Issues in Syntax and Semantics 9, Paris (2012)
12. Levin, B., Rappaport-Hovav, M.: Argument realization. Cambridge University Press, Cambridge (2005)
13. Löbner, S.: General introduction. In: Proposal for the CRC 991. Institute for Language and Information, University of Düsseldorf (2011)
14. Moens, M., Steedman, M.: Temporal ontology and temporal reference. Computational Linguistics 14(2), 15–28 (1988)
15. Naumann, R.: Aspects of changes: a dynamic event semantics. Journal of Semantics 18, 27–81 (2001)
16. Naumann, R.: Relating ERP-Effects to Theories of Belief Update and Combining Systems. In: Aloni, M., Kimmelman, V., Roelofsen, F., Sassoon, G.W., Schulz, K., Westera, M., et al. (eds.) Logic, Language and Meaning. LNCS, vol. 7218, pp. 160–169. Springer, Heidelberg (2012A)
17. Naumann, R.: Dynamics in the brain and dynamic frame theory for action verbs. To appear in: Proceedings of SMCLC (2012b)
18. Naumann, R.: A dynamic update model of sentence processing, ms, University of Düsseldorf (2012c)
19. Osswald, R., Van Valin, R.: FrameNet, frame structure, and the syntax-semantics interface. To appear in: Proceedings CTF (2012)
20. Petersen, W.: Decomposing concepts with 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 (2007)
21. Petersen, W., Osswald, T.: A formal interpretation of concept types and type shifts. In: Kosecki, K., Badio, J. (eds.) Cognitive Processes in Language. Lodz Studies in Language, vol. 25, pp. 183–191. Peter Lang, Frankfurt (2012)
22. Pinon, C.: Achievements in an event semantics. In: Lawson, A. (ed.) Proceedings of Semantics and Linguistic Theory VII, pp. 276–293. CLC Publications, Cornell University, Ithaca, NY (1997)
23. Van Benthem, J., et al.: Logics of communication and change. Information and Computation 204, 1620–1662 (2006)
24. Wiggins, D.: Sameness and substance. Harvard UP, Harvard (1980)
25. Zwaan, R.A., Stanfield, R.A.: The effect of implied orientation derived from verbal context on picture recognition. Psychological Science 12(2), 153–156 (2001)
26. Zwaan, R.A., et al.: Language comprehenders mentally represent the shape of objects. Psychological Science 13(2), 168–171 (2002)

What Does It Mean for an Indefinite to Be Presuppositional?*

Umut Özge

University of Stuttgart

Abstract. The paper is concerned with the nature of the presuppositionality involved in "strong" (or presuppositional) indefinite noun phrases in general, and Turkish accusative marked indefinites in particular. It investigates the semantics of Turkish accusative indefinites with regard to the categories of existential import, contextual restrictedness (or D-linking) and semantic scope, within the DRT-based Binding Theory of presupposition justification. It argues that neither contextual restrictedness nor scope properties alone can account for the semantics of Turkish Acc-indefinites. It further argues that existential import, modeled as anaphoricity encoded in the semantics of Acc-indefinites, is fundamental to "strong" indefiniteness in Turkish and can be construed as the source of both contextual restrictedness and wide scope behavior.

1 Introduction

Following the development of Discourse Representation Theories (aka. dynamic semantics) of Kamp (1984) and Heim (1982), it has become almost standard to treat indefinite noun phrases (indefinites for short) as linguistic devices that introduce new referents into the discourse model, replacing their classical Russellian analysis as existential quantifiers. According to the dynamic model, two basic characteristics of an indefinite noun phrase are that its associated referent is novel in the discourse, and the meaning of the indefinite does not involve any presuppositions. These two aspects are usually taken to be the crucial difference between indefinites and definites.

This fundamental model of indefiniteness has been further developed in various ways in the face of the fact that in many languages noun phrases appearing at certain positions or bearing certain marked forms, while behaving like ordinary indefinites in introducing new referents, encode certain relations to their sentential and extra-sentential context.

For an instance, Diesing (1992) argues that indefinites come in two varieties as presuppositional and non-presuppositional, where the type of an indefinite is decided on the basis of its syntactic position. One such position is the subject slot of an "individual-level" predicate. Indefinites appearing at that position are argued to carry an existence

* I gratefully acknowledge the feedback I received from two anonymous reviewers, Varol Akman, Cem Bozşahin, Daniel Büring, Elizabeth Coppock, Jin Cui, Cornelia Ebert, Chiara Gianollo, Aslı Göksel, Klaus von Heusinger, Hans Kamp, Yiğit Karahanoğulları, Cem Keskin, Paul Kiparsky, Jaklin Kornfilt, Sebastian Löbner, Rick Nouwen, Duygu Özge, Arndt Riester, Maribel Romero, Ceyhan Temürcü, Ümit Turan, and Deniz Zeyrek, and the material support of the German Science Foundation by Project C2 "Case and Referential Context" under Sonderforschungsbereich 732 "Incremental Specification in Context" of University of Stuttgart.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 138–154, 2013.

presupposition (or existential import). In the following example from von Fintel (1998), the oddness of continuation (1b) is explained as the failure of the existence presupposition triggered by the subject of the "individual-level" predicate *are major*. The failure is due to the ignorance of the speaker regarding the existence of mistakes, which she declares in the opening section of the discourse.

(1) I'm not sure yet whether there are any mistakes at all in this book manuscript, but we can definitely not publish it...
   a. if some mistakes are found.
   b. #if some mistakes are major.

For another presuppositionality claim regarding indefinites, Enç (1991) claims that in certain languages, at certain positions indefinites are explicitly marked as Discourse-linked (or D-linked for short).[1] Turkish is one language that employs an overt morphological marker to indicate the category of D-linking. Enç (1991) claims that indefinite noun phrases at the immediately preverbal position are bound to get interpreted as connected to the previous discourse when they carry the accusative marker (henceforth Acc). For instance while (2a), which has no overt case marking on the direct object (henceforth ∅) can be a perfectly natural discourse opener, (2b), which differs from the former only in the Acc-marker on the direct object, is not interpretable unless the hearer accommodates the speaker by inserting a familiar set of books or things to his/her discourse model; a typical effect for presuppositional expressions uttered in contexts that are insufficient to justify those presuppositions.

(2) a. Dün gece bir kitap okudum.
   yesterday night a book read.1sg
   'Last night I read a book.'
   b. Dün gece bir kitab-ı okudum.
   yesterday night a book-Acc read.1sg
   'Last night I read one of the books.'
   'Last night I read a book.' (picked from a familiar set of items)

The literature is not conclusive on the relation between the notions of existential import (or presuppositionality) and D-linking.[2] In this regard it is important to get clear about the nature of presuppositionality involved in Turkish Acc-indefinites, which could provide answers to questions like: Is D-linking the same concept as existential import? If not, which one is fundamental to the behaviour of "strong" indefinites? Is there an empirical basis for D-linking in Turkish?

[1] The term Enç 1991 uses for her semantic category is "specificity". However, she explicitly identifies this notion with what Pesetsky (1987) calls Discourse-linking. I prefer to use the second term (and "contextual restrictedness" in the later parts) to guard ourselves from the confusion surrounding the term "specificity" as much as possible.

[2] Diesing 1992 entertains the possibility that they are the same concept, while van Geenhoven 1998 holds them distinct. Enç 1991 claims that her notion of "specificity" (D-linking) is what underlies Milsark's 1977 "strong"/"weak" distinction, which is the point of departure for Diesing 1992.

In the aim of shedding some light on these issues, the paper investigates the semantics of Turkish Acc-indefinites with regard to the categories of existential import, D-linking (or, later, contextual restrictedness) and semantic scope. In this investigation I adopt the DRT-based Binding Theory of presupposition justification (Geurts 1999; van der Sandt 1992) as the formal framework. I start in Section 2 with a closer look at D-linking as it is formalized by Enç (1991). In Section 3 I introduce the Binding Theory. In Section 4 I address the question whether D-linking and existential import are distinct, and if they are, which one is fundamental for Turkish Acc-indefinites. There I argue that Enç's (1991) proposal is not adequate in capturing the Turkish facts, and existential import is more fundamental than D-linking to the semantics of Turkish Acc-indefinites. In Section 5 I discuss and object to some potential and actual arguments in defense of Enç 1991, which give semantic scope a fundamental role in the semantics of "strong" indefiniteness. In Section 6 I present a proposal, and discuss how it handles the data left uncovered by D-linking and scope based proposals. Finally I conclude in Section 7.

2 Acc-Marking and D-Linking

Enç (1991) claims that there is a bidirectional implication between Acc-marking and D-linking. The notion of D-linking is best illustrated by an example from Enç (1991).

(3) Odam-a birkaç çocuk girdi.
my-room-dat several child entered
'Several children entered my room.'

(4) a. İki kız-ı tanıyordum.
two girl-Acc knew.1sg
'I knew two girls [among the children].' (D-linked)

b. İki kız tanıyordum.
two girl knew.1sg
'I knew two girls.' (non-D-linked)

Enç (1991) observes that in (4a) the girls are necessarily understood as belonging to the set of children mentioned in (3),[3] while those in (4b) necessarily introduce referents from a domain disjoint with the one in (3).

Enç (1991) adopts a dynamic framework, where the primary function of nominal expressions is to introduce discourse referents (modeled as variables) into the discourse

[3] Enç 1991 alludes to the notion "partitivity" in characterizing her notion of "specificity". It is crucial to note here that the Acc-indefinite in (4a) is not interchangeable with an explicit partitive like *kızlardan ikisini* ('two of the girls') (cf. Enç 1991:6). First, the explicit partitive implies that there are more than two girls in the group, while no such implication is present for the indefinite form. Second, the acceptability of the example as a continuation to (3) significantly degrades, presumably because the presupposition involved with the explicit partitive is much harder to accommodate than the one involved with the Acc-indefinite. This is the reason why the example should not be translated as *I knew two of the girls*. I am grateful to an anonymous reviewer for pointing out the need for clarification at this point.

model. In Enç's (1991) treatment every NP introduces two variables (instead of the customary one). Both of these variables can be either indefinite or definite. The first variable (x_i in 5) stands for the referent of the NP, while the second variable (x_j in 5) stands for the superset this variable is required to be a subset or an element of, depending on whether the NP in question is plural or singular.

(5) Every $[_{\mathrm{NP}}\ \alpha]_{\langle i,j\rangle}$ is interpreted as $\alpha(x_i)$ and
$x_i \subseteq x_j$ if $\mathrm{NP}_{\langle i,j\rangle}$ is plural;
$\{x_i\} \subseteq x_j$ if $\mathrm{NP}_{\langle i,j\rangle}$ is singular.

Ordinary definite descriptions are definite in their first variable, meaning that their referent must be given in the prior discourse. Non-D-linked indefinites are indefinite in both variables, meaning that both their referents and the supersets they come from are new to the discourse. D-linked indefinites are those which are indefinite in the first variable, but definite in the second. They introduce novel referents into the discourse model, but the superset from which this novel referent is picked from has to be given in the discourse model.

Enç (1991) extends her treatment to direct objects headed by "strong" quantifiers like *every* and *most*, which obligatorily receive Acc-marking in Turkish. Thereby, D-linking is offered as a unified concept underlying the "strong"/"weak" distinction.

In Section 4 I will have a closer look at Enç's (1991) proposal. Before that I introduce the formalism used in the rest of the paper.

3 Presuppositionality as Anaphoricity

Definite descriptions like *the errors*, together with other "strong" NPs, are usually argued to carry, among possibly others, the presupposition that their domain is not empty. One way to model this is to explicitly define a definedness relation between expressions and contexts, which says that a definite description *the N* is defined (i.e. has a semantic value) in a context c only if c entails that there is exactly one individual that satisfies N.

The DRT-based Binding Theory of presuppositions (Geurts 1999; van der Sandt 1992) offers an alternative way to think about presuppositionality.[4] The basic idea behind the Binding Theory is "presuppositionality as anaphoricity". Geurts (2007:253) explains how this idea can be applied to existence presuppositions triggered by "strong" NPs in the following quote: "A strong quantifier does not merely presuppose that its domain is non-empty; rather, the purpose of its presupposition is to *recover* a suitable domain from the context."

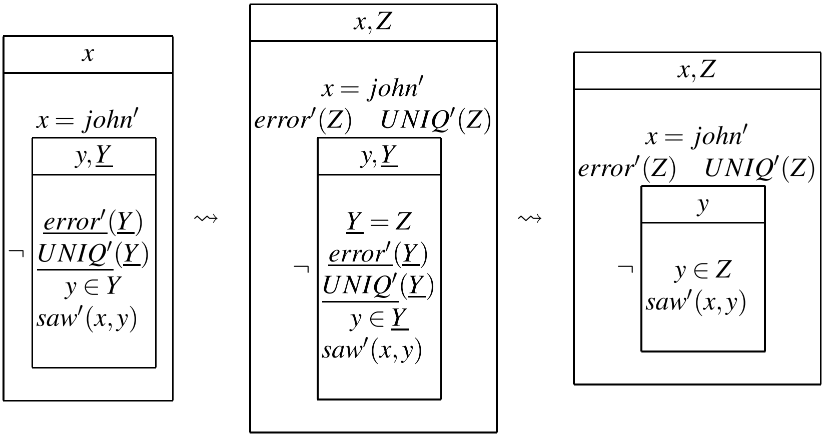
In the rest of the paper I use a simplified version of the Binding Theory. Here is an example of how a sentence involving a presupposition trigger is handled within Binding Theory.[5]

(6) John saw none of the errors (in the article).

[4] The present Binding Theory is totally distinct from the module of the Government Binding Theory with the same name.

[5] For space concerns I assume basic familiarity with DRT.

(7)



The first step of (7) consists of building a preliminary representation from (6).[6] The crucial point here is how the presupposition triggered by the definite description *the errors* is represented. The presuppositional content of the definite description is added to the same discourse representation structure (DRS) together with its assertive content (i.e. the negated box in the left most representation in 7). The presuppositional content is notationally distinguished from the assertive content by underlining. The meaning of this convention is that underlined discourse referents and conditions should get bound by antecedents in the same or a higher and accessible DRS in order for the whole representation to be interpretable. The predicate constant '*UNIQ′*' stands for "unique identifiability", which I employ as a placeholder for a more thorough formulation of definiteness.[7] What all this mechanism amounts to say is that for the representation constructed on the basis of lexical content and compositional derivation of (6) to be interpretable, the hearer needs to find a uniquely identifiable set of errors in the discourse model. Now I turn to what happens after this preliminary representation is constructed.

Two things happen in the second step of (7). One, given that (6) is a discourse opener, and hence there were no suitable antecedent for Y and its associated conditions to get bound to, an antecedent is accommodated into the main DRS.[8] In other words, the hearer acts as if there were a uniquely identifiable set of errors in the discourse context to which (6) is contributed. It is crucial at this point to note that, as a general principle, presuppositions tend to get accommodated at the highest possible position (aka "global accommodation"), especially in the absence of contextual factors that force "non-global accommodation" (see below and Geurts 1999 for more on this). Two, Y is bound to the accommodated antecedent Z by an equivative condition.

Finally, in the third step we get rid of underlined referents and conditions and arrive at the final (and interpretable) representation. The resulting representation is verified in

---

[6] Notational conventions: Primes ('′') distinguish constants from variables. Upper case variables 'X, Y, Z' stand for sets, lower ones stand for atomic individuals. I assume that same predicate can apply both to atomic and set arguments.

[7] There also needs to be a plurality constraint, which I gloss over for simplicity.

[8] The communication fails if such an accommodation is not possible for some reason.

contexts where there is an individual John and a (uniquely identifiable) set of errors, and it is not the case that there is at least one error from this set such that John saw it.

4 Presuppositionality of Acc-Indefinites

Having seen how the Binding Theory works, now it can be used to clarify the nature of the presuppositionality involved in Turkish Acc-indefinites. An obvious way to reconstruct Enç's (1991) formulation of D-linking given in (5) above is to make the contextual restriction requirement a presupposition triggered by the Acc-marker. To illustrate, let us return to her example:[9]

(8) Odam-a birkaç çocuk girdi.
my-room-dat several child entered
'Several children entered my room.'

(9) a. Bir kız-ı tanıyordum.
a girl-Acc knew.1sg
'I knew a/one girl.' (Acc-marked: D-linked)

b. Bir kız tanıyordum.
a girl knew.1sg
'I knew a girl.' (∅-marked: non-D-linked)

The discourse opener (8) gets the following simple DRS:

(10)

X
$child'(X) \quad entered'(X)$

The representation for (9a), which has an Acc-marked indefinite object, is as follows:

(11)

$x, y, \underline{Z}$
$x = spkr' \quad girl'(y) \quad y \in \underline{Z} \quad \underline{UNIQ'(\underline{Z})} \quad know'(x,y)$

Here the definiteness requirement Enç (1991) puts on the superset variable of D-linked indefinites is modeled by making the superset variable Z presuppositional by underlining it, and introducing the unique identifiability requirement. (11) states that the girl in question must be part of a contextually given set. When the representation in (11) is added to the established discourse given in (10), one gets:

[9] The example is slightly altered by switching to a singular indefinite in order not to deal with plurality. I will continue to use singular indefinites in the rest of the paper. To the best of my judgment, nothing important hinges on this alteration.

(12)
$$\begin{array}{|c|}\hline X, x, y, \underline{Z} \\ \hline child'(X) \quad entered'(X) \\ x = spkr' \quad girl'(y) \quad y \in \underline{Z} \quad \underline{UNIQ'}(\underline{Z}) \quad know'(x,y) \\ \hline \end{array}$$

The obvious justification of the presupposition triggered by the Acc-marker is to bind Z to the set of children introduced before. This gives us the final representation below.[10]

(13)
$$\begin{array}{|c|}\hline X, x, y \\ \hline child'(X) \quad entered'(X) \\ x = spkr' \quad girl'(y) \quad y \in X \quad know'(x,y) \\ \hline \end{array}$$

In the derivation for the ∅-marked version (9b) the superset variable Z in (11) would be novel. Enç (1991:note 11) suggests that for non-D-linked indefinites the superset is simply identified with the restrictor predicate of the head noun of the NP. This means that anything related to the superset would simply get eliminated resulting in the following.

(14)
$$\begin{array}{|c|}\hline X, x, y \\ \hline child'(X) \quad entered'(X) \\ x = spkr' \quad girl'(y) \quad know'(x,y) \\ \hline \end{array}$$

As far as I can see, my reconstruction of Enç's (1991) proposal in DRT terms does full justice to its original, apart from dealing with plurality.

Now I will argue that Enç's (1991) formulation is not fully adequate in capturing certain empirical facts. In this argument I make use of the behavior of Acc-indefinites under negation. Let us start by considering the negated version of (9a).

(15) Odam-a birkaç çocuk girdi.
my-room-dat several child entered
'Several children entered my room.'

(16) a. Bir kız-ı tanımıyordum.
a girl-Acc knew-Neg.1sg
'I didn't know a girl. (=There was a girl [among them] I didn't know.)'

b. Bir kız tanımıyordum.
a girl knew-Neg.1sg
'I didn't know any girl.' (non-D-linked)

[10] Note that the unique identifiability of the set of children is assumed and left implicit in the DRSs.

The representation of the Acc-marked (16a) is given in two steps, where the presuppositional superset is again bound by the children introduced in the opening sentence (15):

(17)
$$
\begin{array}{|c|}
\hline
X,x \\
\hline
child'(X) \quad entered'(X) \\
x = spkr' \\
\neg \begin{array}{|c|}
\hline
y,\underline{Z} \\
\hline
girl'(y) \quad y \in \underline{Z} \quad \underline{UNIQ'}(\underline{Z}) \\
know'(x,y) \\
\hline
\end{array} \\
\hline
\end{array}
\rightsquigarrow
\begin{array}{|c|}
\hline
X,x \\
\hline
child'(X) \quad entered'(X) \\
x = spkr' \\
\neg \begin{array}{|c|}
\hline
y \\
\hline
girl'(y) \quad y \in X \quad know'(x,y) \\
\hline
\end{array} \\
\hline
\end{array}
$$

Informally, the end result of (17) reads:

(18) There is no individual known by the speaker such that she is a girl and she is one of the children who entered the room of the speaker.

The crucial observation is that this interpretation is verified in a context where there are no girls among the children who entered the room. However the most immediately available interpretation of (16a) is:

(19) There is a girl among the children who entered the room such that the speaker doesn't know her.

More strongly, to the best of my and my informants' judgement, there is no interpretation of (16a) such that a speaker can use this sentence without committing to the existence of girls among the children. Therefore (18), which is predicted by Enç's (1991) formulation, cannot be among the interpretations of (16a).[11]

I argue that the inadequacy of Enç's (1991) model lies in its ignoring the existential import as a presuppositional component of Acc-indefinites, while concentrating solely

---

[11] One might attribute the unavailability of (18) as an interpretation of (16a) due to pragmatic reasons, along the following lines. In the general case, the absence or presence of girls among a bunch of children is a visually decidable matter. By this token it is reasonable to assume that the speaker of (16a) knows whether there are girls in the group. Then one continues to reason as follows: If someone wants to claim (18) on the basis of the proposition *There are no girls in the group*, then she is expected to assert this stronger proposition rather than the weaker (18). Therefore, one might argue, we infer that the speaker knows (or sees) that there are girls in the group, and this inference is the source of the existential import. However this pragmatic explanation cannot be valid: My argument can be replicated by changing the example by replacing *children* and *girls* respectively with *academicians* and *professors*, where visual identifiability is presumably not at issue. Still, the sentence *Bir profesörü tanımıyordum.* ("I didn't know a professor."), which has an accusative marker on the object, cannot get the type of interpretation given in (18), because again the speaker commits herself to the existence of professors among the academicians.

on the contextual restrictedness requirement.[12] If the semantics of the Acc-marker were that of contextual restriction, then there would be no reason why (18) cannot be expressed with (16a).[13]

Before giving the present proposal in Section 6, I look in some detail at some arguments that might be put in defense of Enç's (1991) modelling of Acc-indefinites.

5 Acc-Indefinites and Scope

It might be argued that the most prominent reading of (16a) given in (19), which cannot be captured by (my reconstruction of) Enç's (1991) proposal, is a so called "specific" or "existential taking scope over negation" reading. In this case the existence of at least one girl among the students would be an assertion rather than a presupposition of (16a). If one could add an independent mechanism that forces the Acc-indefinite to take a wider scope than negation, then one could defend Enç's (1991) proposal. I look at two forms such an argument can take.

Argument 1. The Acc-marker has a scopal semantics which forces its host NP to raise in some level of logical form.[14] This mechanism gives the Acc-marked indefinite in (16a) wide scope over negation, thereby resulting in an assertion of existence.

Objection. Acc-marked indefinites do not necessarily take wide-scope over commanding operators (see Enç 1991 and Özge 2011 on the interaction of Acc-indefinites with various intensional and nominal operators).[15] The possibility of narrow scope Acc-indefinites does not in itself refute the general argument from scope, however. There may be other mechanisms at work that force Acc-indefinites to take wide scope in certain occasions. Now I turn to an argument alluding to such a mechanism.

Argument 2. It is common for marked (or "strong") indefinites to tend to take wide scope with respect to commanding operators (see Farkas 2002 for a review).

[12] See Kelepir 2001 for an earlier claim that the essential interpretative aspect of Acc-marking is the presupposition that the domain of the indefinite is not empty. Her claim is backed by the observation that in the object position of referentially opaque verbs Acc-marked indefinites get a referential reading without being D-linked (or partitive). Therefore, she concludes, D-linking cannot be the underlying semantics of Acc-marking.

[13] It should also be noted that (18) is not available with (16b) either, which has the following interpretation:

(i) The speaker does not know any girl (whomsoever).

If (18) were available with the ∅-marked (16b) then it would have been possible to make an argument from Gricean inference to the effect that the existence committing reading (19) is a by-product of using the marked (Acc) form, in a situation where the unmarked (∅) form could have been used as well. We will have more on this below.

[14] See Aygen-Tosun 1999 for such a proposal on Acc-marked indefinites in Turkish; and Farkas 2002 for various cross-linguistic examples of scopal semantics of special indefinites.

[15] This may not be so obvious for example (16a) in particular, and negation in general, due to certain contextual factors that interact with negation. Below we will see that once these contextual factors are adjusted, the so called "specific" or wide-scope reading in (19) is not the only reading one could get.

Furthermore, it is also common, and true for Turkish, that unmarked forms are obligatorily narrow scope. Given these facts, Enç (1991:23) suggests that a Gricean inferential mechanism might be responsible for giving Acc-marked indefinites wider scope than commanding operators. As far as I can see, Enç's (1991) Gricean argument hinges on flouting the maxim of quantity and can be stated as follows:

(20) If the speaker wants to convey a narrow-scope reading, it is enough that he uses the unmarked ∅-form. But he uses the stronger (more informative) Acc-marked form, therefore he either flouts the maxim of quantity by saying more than needed, or his aim is not to convey a narrow scope reading. Since I assume the speaker is a cooperative one, I go for the second possibility.

Therefore, the argument would go, it is this wide scope implicature, rather than the presuppositional properties of the Acc-marked indefinite, that is the source of existential import.

Objection. In order for the argument in (20) to go through two conditions need to be met: (i) The Acc-marking should not be already motivated by something other than scope; (ii) The Acc-marked form should indeed be stronger (more informative) than the ∅-marked form. Remember from note 13 that (16b) is interpreted as claiming that the speaker does not know any girl whomsoever. It crucially lacks an interpretation which states that the speaker does not know any girl from among the given set of children. This corroborates Enç's (1991) intuition that the ∅-marked indefinites are context independent, giving the classical narrow reading of so called "non-specific" indefinites. This also shows that the first necessary condition of a Gricean inference is not met in our example, since the Acc-marker is motivated to "link" the girl to the given set of children.[16] The second requirement is violated as well. Thanks to the downward entailing context of (16a), the Acc-marked version, which the argument in (20) requires to be stronger than the ∅-marked version, is indeed weaker than it: If I do not know any girls, then I do not know a girl among the children who entered my room; but even if I do not know a girl among the children who entered my room, it may still be the case that I know some other girls. I conclude that the Gricean inference argument in (20) does not go through. Therefore we still lack an explanation regarding the source of the existential import of (16a).

I have another objection directed towards the argument from scope in general. This objection is based on the independence of existential import from wide-scope. In note 15 above, I claimed that the so called "specific" reading (19) is not the only reading one can get from (16a). I claim that there is also the reading represented as (21), which says that the speaker does not know any of the girls among the given set of children. I argue that this reading is inhibited in the context of (8) due to pragmatic concerns. Specifically, in order for (21) to be available, the context should attribute a "significance" to the state of affairs represented in (21); and (8) does not meet this requirement.

[16] Enç (1991) herself states that her Gricean argument works only for out-of-the-blue Acc-marked indefinites, where, she argues, this linking function does not apply.

(21)
$$
\begin{array}{|c|}
\hline
X, x \\
\hline
child'(X) \quad entered'(X) \\
x = spkr' \\
girl'(Z) \quad Z \subseteq X \\
\neg \begin{array}{|c|}
\hline
y \\
\hline
y \in Z \quad know'(x,y) \\
\\
\hline
\end{array} \\
\\
\hline
\end{array}
$$

Now I will provide a context where the type of reading depicted in (21) is available.[17] Assume a camping context where there is a given set of campers. The individual John took an excursion. On his return he told about some dangers he faced with, and how hard it was for him to be able to get back to the camp. Someone utters the following:[18]

(22) Neyse ki John giderken yanına bir çocuğ-u almamış.
fortunately J. while going with-him a child-Acc take-Neg-Past.4sg
'Fortunately, John hasn't taken a child with him to the excursion.'

which can naturally get the following interpretation, ignoring the contribution of *fortunately*.[19]

---

[17] The reading in (21) is readily available with a free-choice determiner *herhangi bir* ('any') or the negative polarity item *hiç bir* ('lit. 'none a'') as the determiner and with an accusative marker on the indefinite:

(i) a. Hiç bir kız-ı tanımıyordum.
none a girl-**Acc** knew-Neg.1sg
'I knew none of the girls.' (D-linked)

b. Hiç bir kız tanımıyordum.
none a girl knew-Neg.1sg
'I didn't know any girls (whomsoever).' (non-D-linked)

As the glosses make clear, the function of the Acc-marker cannot be reduced to a scope difference, as in both variants the indefinite takes scope under negation.

[18] It should be noted that (22) is not the most natural way to put the meaning represented in (23). The explicit partitive *çocuklardan birini* ("one of the children") would be more natural. I nevertheless think that (22) is an acceptable Turkish sentence which gives the meaning contrast I build the argument over. In any case, the scope versus existential import distinction is empirically quite reliable for free choice and negative polarity items of note 17.

[19] The reason why I think we do not need to pay attention to the contribution of *fortunately* is that whatever analysis one adopts for this (presumably focus sensitive) operator, one needs to "feed" it with a representation like (23), with a focus marking on the "most interesting" constituent *çocuğ-u* ('child-Acc'). Under the assumption of compositionality, this means that one should have the representation in (23) *at some point* in the computation of the meaning of (22).

(23)
$$\begin{array}{|c|}\hline X,x \\ \hline camper'(X) \quad x = john' \\ child'(Z) \quad Z \subseteq X \\ \neg \begin{array}{|c|}\hline y \\ \hline y \in Z \quad take'(x,y) \\ \hline \end{array} \\ \hline \end{array}$$

It is important to observe that (22) is not an "emphatic denial". We have the following reasons behind this judgment. One, it is not necessary that John took some non-child individual with him—he may well have gone alone; or it is likewise not necessary that whether John should take a child with him or not be a question under discussion previously. Two, the example passes the test of "*why*-question contextualization" proposed in Szabolcsi 2004: For instance, (22) might be a response to the question *Why do you feel so relieved?*, asked when John is reported to be currently on a dangerous excursion. Such type of contextualization is not available for "emphatic denials".

The most relevant point concerning (22) is its triggering the inference that there were children among the campers. The speaker could not be talking about any child, which would be possible without the accusative marking on the indefinite. She necessarily commits to the existence of children among the campers. This in turn shows that her utterance carries the existential import associated with the indefinite *bir çocuğ-u* ('a child-Acc') without giving the indefinite a wider scope than negation. I think this is a clear illustration of why existential import should be kept distinct from scope.

Let us sum up what we had so far. We translated Enç's (1991) proposal for Turkish Acc-marked indefinites (and similar constructions in other languages) into a DRT-based Binding Theory of presupposition justification. Then, with the aid of negation, we showed that Enç's (1991) model is too weak to capture the relevant empirical facts. Specifically, her model fails to account for existential import while concentrating on contextual restrictedness. Next, we considered some potential arguments that can be put in defense of Enç (1991). These arguments proposed that the source of the effects of the Acc-marker under negation might be due to scope properties of the marker rather than any presupposition triggered by it. I provided various objections to these arguments, and established that existential import should be kept distinct from scope as well. The upshot of the discussion so far is that existential import should be kept apart from both contextual restriction/dependence and semantic scope; its effects cannot be reduced to either of them. In the next section I will propose a model that aims to do justice to these observations.

6 A Proposal

I claim that the basic distinction between an Acc-marked indefinite and a ∅-marked one is that the restrictor of the former is an anaphoric expression where the restrictor of the latter is an ordinary predicate. In an Acc-indefinite like *bir çocuğu* ('a child-Acc'), there is an anaphoric component that is slightly different from an ordinary plural pronoun like *they*. The difference is that while *they* does not have any lexical content apart from plurality, the anaphoric component of the Acc-indefinite *bir çocuğu* ('a child-Acc') "seeks" an antecedent that satisfies the predicate $child'$. To illustrate let us revisit (9a) considered again in the context of (8). The discourse opener (8) has the same interpretation as above:

(24)
$$\begin{array}{|c|}\hline X \\ \hline child'(X) \quad entered'(X) \\ \hline \end{array}$$

(9a) gets the following preliminary representation under the present proposal:

(25)
$$\begin{array}{|c|}\hline x, y, \underline{Z} \\ \hline x = spkr' \quad \underline{girl'}(\underline{Z}) \quad y \in \underline{Z} \quad \underline{UNIQ'}(\underline{Z}) \quad know'(x,y) \\ \hline \end{array}$$

Merging these two representations gives the following:

(26)
$$\begin{array}{|c|}\hline X, \underline{Z}, x, y \\ \hline child'(X) \quad entered'(X) \\ x = spkr' \quad \underline{girl'}(\underline{Z}) \quad y \in \underline{Z} \quad \underline{UNIQ'}(\underline{Z}) \quad know'(x,y) \\ \hline \end{array}$$

In order to arrive at an interpretable representation from here we need to resolve the underlined anaphora, either by binding them to some elements already present in the discourse model, or we need to first adjust our model by introducing some suitable referents, and then bind our anaphora to these accommodated referents. The first option is not available. Binding Z to the given set of children X is no more an option, thanks to the anaphoric condition '$girl'(Z)$'. What is left as an option is to accommodate a set of girls (Y below). This computation is depicted as follows:

(27)
$$\begin{array}{|c|}\hline X, Y, \underline{Z}, x, y \\ \hline child'(X) \quad entered'(X) \\ girl'(Y) \\ x = spkr' \quad \underline{girl'}(\underline{Z}) \quad y \in \underline{Z} \quad \underline{UNIQ'}(\underline{Z}) \quad know'(x,y) \\ \hline \end{array}$$

At this point performing the binding $Y = Z$ still cannot give a fully satisfactory representation. What is missing is the information that the accommodated girls belong among the children. One option here is to introduce a contextual restriction predicate into the semantics of the Acc-indefinite which needs to get bound in the discourse context. For the sake of homogeneity of the representation we can model this as a set, and add an extra condition that our anaphoric restrictor is a subset of this contextual restrictor set. In this setting, the most natural binder of this contextual restriction set would be the set of children, eventually giving us the result we desire.

For all its technical clarity, we are rather sceptic about the necessity of a semantically coded contextual restriction mechanism. After all we are dealing with anaphora resolution, which is a highly "intelligent" process that trades on various factors like recency, salience, and so on. As the proposal goes, the task of the interpretation process at the point the structure in (25) is built is to find a uniquely identifiable set of girls as an antecedent for Z, thanks to the $UNIQ'$ condition on Z. It appears reasonable at this point for an inference to occur, which "carves out" a set of girls from the set of children and makes this set available for binding. Therefore, I suggest that contextual restrictedness involved in Acc-indefinites in Turkish may simply be a by-product of the inferential anaphora resolution process which is responsible for taking care of the anaphoric (or, equivalently, presuppositional) constraints semantically encoded into Acc-indefinites.

In its final form, the present model handles the example under discussion as in (28), where the contextual restriction information (i.e. $Y \subseteq X$) is part of the accommodated information.[20]

(28)
$$\begin{array}{|c|}\hline X, Y, \underline{Z}, x, y \\ \hline child'(X) \quad entered'(X) \\ girl'(Y) \quad Y \subseteq X \quad Z = Y \\ x = spkr' \quad \underline{girl'(\underline{Z})} \\ y \in \underline{Z} \quad \underline{UNIQ'(\underline{Z})} \\ know'(x, y) \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|}\hline X, Y, x, y \\ \hline child'(X) \quad entered'(X) \\ girl'(Y) \quad Y \subseteq X \\ x = spkr' \\ y \in Y \\ know'(x, y) \\ \hline \end{array}$$

Before moving on to how our proposal handles more complicated examples discussed above, a remark concerning number is in order. (28) is not fully specified with regard to the number of girls in the children set; it may be something between one and the size of the children set. I think this is a desirable situation because apart from the uses of stressed *bir* ('one'), which gives a plurality implicature for the restrictor, indefinites are underspecified as to the plurality of their restrictor.

Now we can return to the negative example (22), repeated here:

(29) Neyse ki John giderken yanına bir çocuğ-u almamış.
fortunately J. while going with-him a child-Acc take-Neg-Past.4sg
'Fortunately, John hasn't taken a child with him to the excursion.'

[20] Also note that the unique identifiability of the accommodated set is left implicit in the final representation and the accommodation and binding operations are given together in the first representation.

The computation of an interpretable representation is given as follows:

(30)
$$
\begin{array}{|c|}\hline X,x \\ \hline camper'(X) \\ x = john' \\ \neg\begin{array}{|c|}\hline y,\underline{Z} \\ \hline \underline{child'}(\underline{Z}) \\ \underline{UNIQ'}(\underline{Z}) \\ y \in \underline{Z} \\ take'(x,y) \\ \hline \end{array} \\ \hline \end{array}
\rightsquigarrow
\begin{array}{|c|}\hline X,Y,x \\ \hline camper'(X) \\ x = john' \\ child'(Y) \quad Y \subseteq X \\ Y = Z \\ \neg\begin{array}{|c|}\hline y,\underline{Z} \\ \hline \underline{child'}(\underline{Z}) \\ \underline{UNIQ'}(\underline{Z}) \\ y \in \underline{Z} \\ take'(x,y) \\ \hline \end{array} \\ \hline \end{array}
\rightsquigarrow
\begin{array}{|c|}\hline X,Y,x \\ \hline camper'(X) \\ x = john' \\ child'(Y) \quad Y \subseteq X \\ \neg\begin{array}{|c|}\hline y \\ \hline y \in Y \\ take'(x,y) \\ \hline \end{array} \\ \hline \end{array}
$$

Once again I assume that the contextual link, namely that the child set is a subset of the camper set, is an explicit part of the accommodation step, obviating the need for an explicit contextual restriction mechanism. The antecedent child set is accommodated in the top most discourse representation structure in line with the interpretive principles of the Binging Theory.

An important question that is brought to my attention by Rick Nouwen (p.c.) is how do we know that the "right" set of children is accommodated in (30)? Assume that there are 5 children in the camp, and John took one of them to the excursion. The DRS in (30) gets verified, if the hearer accommodates a set of 4 children, excluding the one who went with John. In order to avoid this problem, we need to assume that the accommodated set is maximal in the sense that it includes all the children in the camp. This kind of constraints on accommodation of antecedent sets should in the end be articulated in an explicit account of speaker intentions as represented by hearers.

One remaining question before I conclude is what happens to so called "specific" readings. For the above example, such a reading would state that there is one particular child that John didn't take to the excursion. Although this reading is not contextually well-supported in this example, we know from (19) that such readings are quite readily available in the absence of contextual factors that foreground readings like those in (30).

The proposed mechanism is, at least technically, capable of capturing "specific" readings. All we need to do to arrive at a "specific" reading for (30) is to accommodate the additional information that the accommodated set of children is a singleton.[21] At this point care should be taken not to think that accommodating a singleton set of children amounts to committing oneself to the claim that there is only one children among the campers. The hearer could be totally ignorant about the actual number of children

[21] Such a move could be thought of as a model of the notion "epistemic specificity" or "speaker having an individual in mind' (Farkas 2002). I think that a fruitful way of implementing this type of "specificity" in the present system would be to use anchoring relations of the type defended in Kamp and Bende-Farkas submitted. I leave this as future work.

among the campers, both when accommodating *some* set of children and when accommodating a singleton set of children. All he or she needs is to find (or create) an antecedent in the discourse model.[22]

Having proposed a technical solution, I leave it open whether or not an independent mechanism is needed for the so called "specific" readings. The decision on this matter needs to be based partly on an investigation of whether the present proposal on its own is adequate in capturing various scope phenomena, which usually motivates the existence of the so called "specific" readings. I do not have room here for such an investigation.

7 Conclusion

The general question this paper was concerned with is the nature of the presuppositionality involved in special types of indefinites, sometimes called "strong" or presuppositional indefinites. We concentrated on Turkish Acc-marked indefinites. I tried to clarify the semantics of Acc-indefinites with respect to three semantic properties: existential import, contextual restrictedness (or D-linking) and operator scope. I argued that neither contextual restrictedness nor scope properties alone can account for the semantics of Turkish Acc-indefinites. I also argued that existential import, modeled as anaphoricity encoded in the semantics of Acc-indefinites, is not only fundamental to "strong" indefiniteness in Turkish, but also can be construed as the source of both contextual restrictedness and wide scope behavior. Admittedly our position is more tentative on scope than it is on contextual restrictedness.

Although this paper concentrated on Turkish Acc-indefinites, I believe that our discussion has implications concerning "strong" indefiniteness in general and the relation between "strong" indefinites and determiner phrases headed by "strong" determiners like *each*, *every*, *most*, and so on. I leave the investigation of these issues to a forthcoming paper.

References

Aygen-Tosun, G.: Specificity and subject-object positions/scope interactions in Turkish. In: Proceedings of the Conference on Turkic Linguistics at Manchester University (1999)

Diesing, M.: Indefinites. MIT Press, Cambridge (1992)

Enç, M.: The semantics of specificity. Linguistic Inquiry 22, 1–25 (1991)

Farkas, D.: Specificity distinctions. Journal of Semantics 19, 213–243 (2002)

von Fintel, K.: Evidence for presuppositional indefinites. MIT, Ms (1998)

van Geenhoven, V.: Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic. CSLI, Stanford (1998)

Geurts, B.: Presuppositions and Pronouns. CRiSPI, vol. 3, Elsevier, Oxford (1999)

Geurts, B.: Existential import. In: Comorovski, I., von Heusinger, K. (eds.) Existence: Semantics and Syntax, pp. 253–271. Springer, Dordrecht (2007)

Heim, I.: The Semantics of Definite and Indefinite Noun Phrases in English. Ph.D. thesis, University of Massachussetts, Amherst (1982)

[22] The solution I offer here for "specific" readings is closely related, but not identical, to proposals like Schwarzschild 2002. The most major point I diverge from such proposals is that I do not treat indefinites as existential quantifiers.

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 (1984)

Kamp, H., Bende-Farkas, A.: Epistemic specificity from a communication-theoretical perspective. Journal of Semantics (submitted)

Kelepir, M.: Topics in Turkish Syntax: Clausal Structure and Scope. Ph.D. thesis. MIT (2001)

Milsark, G.: Toward an explanation of certain peculiarities of the existential construction in English. Linguistic Analysis 3, 1–29 (1977)

Özge, U.: Turkish indefinites and accusative marking. In: Simpson, A. (ed.) Proceedings of the 7th Workshop on Altaic Formal Linguistics, MITWPL #63, pp. 253–267. MITWPL, Cambridge (2011)

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)

van der Sandt, R.: Presupposition projection as anaphora resolution. Journal of Semantics 9, 333–377 (1992)

Schwarzschild, R.: Singleton indefinites. Journal of Semantics 19, 289–314 (2002)

Szabolcsi, A.: Positive polarity–negative polarity. Natural Language and Linguistic Theory 22, 409–452 (2004)

Dynamics of Defeasible and Tentative Inference

Patrick Allo

Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel
patrick.allo@vub.ac.be
http://homepages.vub.ac.be/~pallo

Abstract. Standard refinements of epistemic and doxastic logics that avoid the problems of logical and deductive omniscience cannot easily be generalised to default reasoning. This is even more so when defeasible reasoning is understood as tentative reasoning; an understanding that is inspired by the dynamic proofs of adaptive logic. In the present paper we extend the abnormality (preference) models for adaptive consequence with a set of open worlds to account for this type of inferential dynamics. In doing so, we argue that unlike for mere deductive reasoning, tentative inference cannot be modelled without such open worlds.

Keywords: adaptive logic, awareness, defeasible belief, modal logic, omniscience, open worlds.

1 Omniscience and Tentative Inference

The problem of logical and deductive omniscience can be summarised as follows: Whenever we model the beliefs of some agent a in terms of what is true at all possible worlds that agent considers possible relative to the actual world, the totality of a's beliefs will contain all logical truths, and will be deductively closed. A standard way out of this conundrum is to separate that agent's *explicit* beliefs from her merely *implicit* beliefs, and requiring that the latter but not the former be deductively closed [9]. This can be done by either stipulating that an agent's explicit beliefs are just those implicit beliefs that satisfy an additional condition, like being included in that agent's *awareness*-set, or by adding new, non-standard, states to the range of entities we need to quantify over to obtain an agent's explicit beliefs [9,7]. In the latter case, explicit belief requires one to consider more states than implicit belief, but the bottom-result is still the same: Explicit belief in φ is just the implicit belief in φ together with the additional condition that no non-standard case where φ is true should be considered a live possibility. As a result, this general strategy treats an agent's explicit beliefs as a sub-set of her implicit beliefs. Put differently, while an agent's merely implicit beliefs (implicit beliefs that are not explicit beliefs) and her explicit beliefs are disjoint sets, they can never be in conflict. According to this picture, the dynamics of inference is just the process of moving formulae (or whatever entity we use to represent beliefs) from one set to another.

As long as we're only considering monotonic inferences, the above description of the dynamics of inference agrees with our intuitions. Inference is monotonic,

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 155–165, 2013.

and (except for forgetting) so is the process of moving formulae from one set to another. This is no longer obvious when non-monotonic or *defeasible* inferences are considered as well.

To a first approximation, a defeasible inference is an inference that is warranted by a default-rule [11]:

$$\frac{\psi : \varphi_1, \ldots, \varphi_n}{\chi}$$

that entitles one to derive χ from the *prerequisite* ψ on the condition that the *justifications* $\varphi_1, \ldots, \varphi_n$ are consistent with one's information.

Two related aspects of default-rules are particularly relevant here. First, unlike standard rules, default rules impose both a positive and a negative condition. The prerequisite is a positive condition on our information, but the demand that the justifications should be consistent with our information is a negative condition. While ψ should follow form our information, no contradiction should follow from the union of our information with the prerequisites. Second, because a negative condition can be satisfied relative to our present information but not necessarily relative to any extension of that information, the conclusion of a default-rule may have to be retracted in view of some newly acquired information. In short: default-reasoning is non-monotonic.

Modelling defeasible reasoning by deductively omniscient agents is a rather straightforward thing to do. Given a default-rule as the one mentioned above, an agent a can apply that rule whenever she doesn't believe the negation of some justification. Since our agent will believe the negation of some φ_i whenever the justifications are inconsistent with her beliefs, she will make no mistakes in the application of such default-rules. The only real challenge lies in the representation of how χ should be retracted in view of some novel information. This is at bottom the problem of belief-revision, which is not the primary aim of this paper.

Once we separate an agent's explicit beliefs from her merely implicit beliefs, it becomes less straightforward to represent the application of default-rules. Presumably, the prerequisite should be explicitly believed (or known), but what about the justifications? One option, favoured in [14, §6.2], is to model the conditions as the implicit-beliefs that $\neg\varphi_i \rightarrow \neg\chi$ for each φ_i, and thus prevent the agent to believe the default-rule (and hence to apply it) whenever the negation of some condition is believed. Even without spelling out the details of this proposal we readily see that there is a gap between how prerequisites and conditions are handled. Prerequisites need to be explicitly believed (or known), but side-conditions only need to be implicitly believed. This leads to a picture of defeasible inference where the conclusion χ cannot be retracted in view of some newly derived explicit belief, but only in virtue of some truly novel information.

The alternative approach I want to pursue in the present paper ties the satisfaction of the side-conditions to the absence of some explicit belief. As a result, the conclusion χ of a default rule can then be retracted in view of a newly derived explicit belief (which may have been an implicit belief from the start). Crucially, this means that agents can in a sense mistakenly apply default-rules. The failure to realise that in view of one's implicit beliefs one or more of the conditions aren't satisfied can suffice to trigger the default-rule. The resulting picture is one of *tentative* defeasible inference, and is inspired by the dynamic proof-format of adaptive logics. It is also within the framework of adaptive logics, and more specifically in terms of a modal reconstruction of their consequence-relations, that I will provide a model of tentative and defeasible inference.

2 Adaptive Logics and Adaptive Preference Models

Adaptive logics are logics for defeasible inference that are characterised by, on the one hand, a formula preferential semantics, and, on the other hand, a dynamic proof-theory [5]. From the standpoint of its semantics, adaptive consequence relations are fairly natural: A stronger, but non-monotonic consequence relation is obtained by stipulating that φ follows from a premise-set Γ iff φ is true in some selection of the models of Γ. This approach is formula-preferential in the sense that the selection of models is based on the formulae (of a given form) that are true in these models. Dynamic proofs, by contrast, are less straightforward, as their dynamic nature stems from the presence of conditional proof-rules. In terms of the previous section: The dynamic proof-format is a proof-format that allows for tentative inferences. Lines can be added to a proof on a provisional basis, based on (the positive condition that) some formulae have already been derived at that stage of the proof, and (the negative condition that) some other formulae have not (yet) been derived at that stage. The connection with the description of tentative inference in the previous section is immediate.

The traditional characterisation of an adaptive logic is in terms of a triple consisting of (1) A Tarski-logic referred to as the *lower limit logic* (**LLL**), (2) a set of formulae Ω characterised by a logical form and referred to as the set of *abnormalities*, and (3) a criterion, referred to as the *adaptive strategy*, which in its model-theoretic form selects models of premise-sets that are no more abnormal than what is actually required by that premise-set, and in its proof-theoretical form specifies when the outcome of a conditional proof-rule should be retracted. When formulated in the same format as default rules, we have a default-rule

$$\frac{\psi : \neg\varphi_1, \ldots, \neg\varphi_n}{\chi}$$

whenever $\psi \models_{\mathbf{L}} \chi \vee (\varphi_1 \vee \ldots \vee \varphi_n)$, **L** is the lower-limit-logic, $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Omega$ and *reliability* is the adaptive strategy that is being used.[1]

Rather than to entirely describe adaptive logics in their traditional form, I will immediately introduce them in a modal reconstruction that was first given in [2]. Given an adaptive logic with a lower limit logic **L** (and consequence relation $\models_{\mathbf{L}}$) that is defined over a language $\mathcal{L}_0$, we extend that language with

[1] A similar straightforward reformulation isn't readily available for the minimal abnormality strategy.

the classical connectives as well as with two modal operators such as to obtain a basic preference language. The resulting language $\mathcal{L}_1$ is defined by:

$$\varphi ::= \psi \mid \bot \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \mathsf{U}\varphi \mid \Diamond^{\preceq}\varphi$$

with ψ ranging over the formulae of the base-language $\mathcal{L}_0$. The dual modal operators E and $\Box^{\preceq}$ are defined as usual.

Standardly, such languages are interpreted in preference structures $\mathfrak{M} = (S, \preceq, \|\cdot\|_{\mathfrak{M}})$, where $\preceq$ is a pre-order over S. For our purposes, we need to consider a restricted class of valuations over such structures. We refer to the resulting *general model* as the class of abnormality models.

Definition 1 (Abnormality Models). *An abnormality model is a 3-tuple $\mathfrak{M} = (S, \Omega, \|\cdot\|_{\mathfrak{M}})$ where S is a set of states, Ω the set of abnormalities, and $\|\cdot\|_{\mathfrak{M}}$ a valuation-function. We define a function $Ab_{\mathfrak{M}} : S \mapsto \mathcal{P}(\Omega)$, and a binary relation $\preceq$ over S in accordance with the following clauses:*

1. $Ab_{\mathfrak{M}}(s) = \{\omega \in \Omega : s \in \|\omega\|_{\mathfrak{M}}\}$
2. $s \preceq s' \Leftrightarrow Ab_{\mathfrak{M}}(s) \subseteq Ab_{\mathfrak{M}}(s')$

and restrict the valuation such as to comply with a last clause:

3. *For every proposition $\|\Gamma\|_{\mathfrak{M}} \subseteq S$ and every $s \in \|\Gamma\|_{\mathfrak{M}}$, if for some $\Delta \subset Ab_{\mathfrak{M}}(s)$, we have $\Gamma \cup \{\neg\varphi : \varphi \in \Omega \setminus \Delta\} \nvDash_{\boldsymbol{LLL}} \bot$, then there is an $s' \in \|\Gamma\|_{\mathfrak{M}}$ such that $Ab_{\mathfrak{M}}(s') = \Delta$.*

In this definition the Clauses (1) and (2) ensure that states are exclusively ordered on the basis of the abnormalities they satisfy. Clause (3) forces the presence of "sufficiently normal states" in every semantic proposition. As a result, abnormality models are a special kind of preference models: clauses (1) and (2) are sufficient, but not necessary conditions for $\preceq$ being a pre-order.

Satisfaction is standardly defined. We only mention the clauses for the modalities.

- $\mathfrak{M}, s \Vdash \mathsf{U}\varphi$ iff $\mathfrak{M}, s' \Vdash \varphi$ for all $s' \in S$,
- $\mathfrak{M}, s \Vdash \Diamond^{\preceq}\varphi$ iff $\mathfrak{M}, s' \Vdash \varphi$ for some $s' \preceq s \in S$.

We can now define two subsets of the truth-set of a given premise-set Γ that give a precise meaning to the loose expression "is no more abnormal than required for the satisfaction of Γ." Both these selections correspond to existing adaptive strategies. According to a first selection strategy, only those states that are $\preceq$-minimal in $\|\Gamma\|_{\mathfrak{M}}$ need to be retained. Accordingly, we refer to $\|\Gamma\|^m_{\mathfrak{M}}$ as the *minimally abnormal* states in $\|\Gamma\|_{\mathfrak{M}}$.

$$\|\Gamma\|^m_{\mathfrak{M}} = \{s \in \|\Gamma\|_{\mathfrak{M}} : \forall s'((s' \in \|\Gamma\|_{\mathfrak{M}} \;\&\; s \sim s') \Rightarrow s \preceq s'))\}$$

with $\sim \;=\; \preceq \cup \succeq$.

According to a second strategy, we should also retain those states that do not satisfy any abnormality that isn't also satisfied by some minimally abnormal state. Because this selection strategy is more cautious, we refer to $\|\Gamma\|^r_{\mathfrak{M}}$ as the *reliable* states in $\|\Gamma\|_{\mathfrak{M}}$.

$$\|\Gamma\|^r_{\mathfrak{M}} = \{s \in \|\Gamma\|_{\mathfrak{M}} : \forall \omega \in \Omega \ (s \in \|\omega\|_{\mathfrak{M}} \Rightarrow \exists s' \in \|\Gamma\|^m_{\mathfrak{M}} \ \&\ s' \in \|\omega\|_{\mathfrak{M}})\}$$

The main result from [2] is summarised by the following theorem:

Theorem 1. *Let* $\mid\!\sim^{(\mathbf{L},\Omega)}_m$ *and* $\mid\!\sim^{(\mathbf{L},\Omega)}_r$ *be two adaptive consequence relations based on a lower limit logic* $\mathbf{L}$*, a set of abnormalities* Ω *and, respectively, the minimal abnormality and reliability strategy. If a class* $\mathbb{M}$ *of adaptive preference models satisfies condition* $(*)$

$$\|\Gamma\|_{\mathfrak{M}} \subseteq \|\varphi\|_{\mathfrak{M}} \text{ for all } \mathfrak{M} \in \mathbb{M} \text{ iff } \Gamma \models_{\mathbf{L}} \varphi \tag{$*$}$$

for all $\Gamma \cup \{\varphi\}$ *that only contain* $\mathcal{L}_0$*-formulae, then the equivalences* (M) *and* (R) *stated below hold as well.*

$$\Gamma \mid\!\sim^{(\mathbf{L},\Omega)}_m \varphi \text{ iff } \|\Gamma\|^m_{\mathfrak{M}} \subseteq \|\varphi\|_{\mathfrak{M}} \text{ for all } \mathfrak{M} \in \mathbb{M} \tag{M}$$

$$\Gamma \mid\!\sim^{(\mathbf{L},\Omega)}_r \varphi \text{ iff } \|\Gamma\|^r_{\mathfrak{M}} \subseteq \|\varphi\|_{\mathfrak{M}} \text{ for all } \mathfrak{M} \in \mathbb{M} \tag{R}$$

Since in the remainder of this paper we shall use the two just introduced selection-strategies to define some new belief-operators, we do not yet extend the language $\mathcal{L}_1$ with new operators that express $\|\Gamma\|^m_{\mathfrak{M}} \subseteq \|\varphi\|_{\mathfrak{M}}$ and $\|\Gamma\|^r_{\mathfrak{M}} \subseteq \|\varphi\|_{\mathfrak{M}}$. The reader is referred to [2] for further details.

3 Implicit and Explicit Belief

A by now fairly common way of modelling the connection between knowledge and belief identifies the beliefs of an agent with what is true in all "most plausible" states that agent cannot distinguish from the actual state. That is, it equates the doxastic possibilities with the most plausible epistemic possibilities. In this section we follow the same strategy, but use abnormality-orderings instead plausibility-orderings to decide which epistemic possibilities are also doxastic possibilities. Because inferential processes are not typical interactive or social processes, we restrict ourselves to the single-agent case.

Given an indistinguishability-relation $\approx$, we shall, rather than to assign a set of premises Γ to each agent, treat the set of states that cannot be distinguished from the actual state s as the semantic proposition that is identical to $\|\Gamma\|$ for some Γ we might have used as a premise-set or belief or knowledge-base associated with some agent.[2] In doing so, we deliberately leave behind the *deduction model of belief* [8] that is often associated with models of belief based on a non-monotonic consequence relation. We also, but this is just a matter of convenience, make the simplifying assumption that all premises are true, and hence that an agent can only have false beliefs by making a defeasible inference from true premises.

[2] Because there need not be a unique such Γ, this move makes the third clause of Definition 1 superfluous.

Definition 2 (Doxastic Abnormality Model). *A Doxastic Abnormality Model is a 4-tuple* $\mathfrak{M} = (S, \approx, \Omega, \|\cdot\|_{\mathfrak{M}})$*, where* S *is a set of states,* $\approx$ *an equivalence-relation over* S*,* Ω *a set of abnormalities, and* $\|\cdot\|_{\mathfrak{M}}$ *a valuation-function.*

We define a function $Ab_{\mathfrak{M}} : S \mapsto \mathcal{P}(\Omega)$*, and a binary relation* $\preceq$ *over* S *in accordance with the following clauses:*

1. $Ab_{\mathfrak{M}}(s) = \{\omega \in \Omega : s \in \|\omega\|_{\mathfrak{M}}\}$
2. $s \preceq s' \Leftrightarrow Ab_{\mathfrak{M}}(s) \subseteq Ab_{\mathfrak{M}}(s')$

Remark 1. Unlike for the systems defined in [3], the indistinguishability-relation $\approx$ cannot as straightforwardly be defined in terms of the ordering $\preceq$. This is so because $\preceq$ is a global abnormality-ordering that is not associated with the partition induced by $\approx$. Given a set of states S' that cannot be distinguished from the actual state, we can consider the restriction $\preceq'$ of $\preceq$ to S'. Yet, because abnormality-orderings need not be connected, $\approx$ is not the union of $\preceq'$ with its converse. It is the transitive closure of that union.

The knowledge-operator $[\approx]$ is defined along standard lines:

$$\|[\approx]\varphi\| = \{s \in S : \forall t(s \approx t \Rightarrow t \in \|\varphi\|)\},$$

but because $\preceq$ is only a pre-order we have two non-equivalent ways of defining doxastic possibilities;[3] one for the standard minimal abnormality selection-strategy, and one for the more cautious reliability selection-strategy:

$$s \xrightarrow{m} t \Leftrightarrow s \approx t \ \&\ \forall u((t \approx u \ \&\ t \sim u) \Rightarrow t \preceq u) \tag{Min}$$

$$s \xrightarrow{r} t \Leftrightarrow s \approx t \ \&\ \forall \omega \in \Omega \ (t \in \|\omega\| \Rightarrow \exists u(s \xrightarrow{m} u \ \&\ u \in \|\omega\|)) \tag{Rel}$$

The clauses for the corresponding belief-operators $[\xrightarrow{m}]$ and $[\xrightarrow{r}]$ are again standard:

$$\|[\xrightarrow{m}]\varphi\| = \{s \in S : \forall t(s \xrightarrow{m} t \Rightarrow t \in \|\varphi\|)\},$$

$$\|[\xrightarrow{r}]\varphi\| = \{s \in S : \forall t(s \xrightarrow{r} t \Rightarrow t \in \|\varphi\|)\}.$$

Because (Min) and (Rel) can be shown to define serial, transitive, and euclidean relations, the doxastic operators $[\xrightarrow{m}]$ and $[\xrightarrow{r}]$ are entirely standard **KD45**-operators.

In analogy with how standard epistemic and doxastic logics fall prey to the standard problems of logical and deductive omniscience, the doxastic model of Definition 2 fails to account for the proof-dynamics that are typical for adaptive logics. In that respect, doxastic adaptive logics are exactly like the abnormality preference models introduced in the previous section. A further modification of these models will overcome this problem. As hinted at in the introduction, a proper account of tentative inference cannot be achieved by separating explicit from merely implicit beliefs in such a way that an agent's explicit beliefs are a

[3] Both strategies coincide when $\preceq$ is connected within a set of indistinguishable states.

subset of her implicit beliefs. We need an account of explicit beliefs that leaves room for actual mistakes. An appeal to *open worlds* in the sense of [10]—worlds where there need not be any logical connection between typographically distinct formulae—will allow for that.

The inspiration for our solution is the idea that by making inferences we gain *insight* in our premises. That is, while initially we need not see beyond the surface structure of a formula, using that formula in an inference will often require us to recognise its logical structure. This idea was first applied in one of the earliest attempts to give a semantic account of the proof-dynamics of adaptive logics by making the insight one needs to have in one's premises at a given proof-stage explicit at the syntactical level. A 'block-language' was introduced to that effect in [4]. Here, we implement the same idea without having to appeal to a dedicated block-language, but by formalising insight in terms of the ability to discriminate open worlds from closed worlds. Discriminability, in that sense, can be formalised with the truth-value equivalence relations introduced in [13].

Definition 3 (Truth-Value Equivalence). *Given a set of formulae Θ, we say that $s, t \in S$ are equivalent with respect to Θ iff for all $\theta \in \Theta$*

$$s \in \|\theta\| \Leftrightarrow t \in \|\theta\|$$

When s and t are equivalent with respect to Θ, we write $s \equiv_\Theta t$.

Using that definition, $o \equiv_{\{p \to q\}} s$ shall mean that if s is a closed world where $p \to q$, but also p and q are true, and o is an open world where $p \to q$ is still true, but where p can be true without q also being true, these two worlds still cannot discriminated.

Definition 4 (i-Doxastic Abnormality Model). *An i-Doxastic Adaptive Preference Model is a 6-tuple $\mathfrak{M} = (S, O, \approx, I, \Omega, \|\cdot\|_\mathfrak{M})$, where S is a set of closed worlds, O a set of open worlds, $\approx$ an equivalence-relation over S, I a set of $\mathcal{L}_0$-formulae that is closed under formation-rules that do not decrease the complexity of a formula, Ω the set of abnormalities, and $\|\cdot\|_\mathfrak{M}$ a valuation-function. $Ab_\mathfrak{M}$ a map: $S \cup O \mapsto \mathcal{P}(\Omega)$ and $\preceq \subseteq (S \cup O) \times (S \cup O)$ are defined as in Definition 2.*

We additionally define the binary relation $\simeq$ in accordance with the following clause:

1. *$s \simeq t$ iff either $s \equiv_I t$ or there is a $u \in S$ such that $s \approx u$ and $u \equiv_I t$.*

Since $\approx$ is only defined over S, it follows that each $\xrightarrow{m}$ and each $\xrightarrow{r}$ is also only defined over S. Consequently, the prior definitions of knowledge and belief remain applicable, but are now overtly accounts of implicit knowledge and implicit belief. Likewise, since $\preceq$ is defined over $S \cup O$, we can still use the abnormality-ordering to define a new notion of explicit belief. Note that the fact that $\approx$ is undefined over O does not mean that $[\approx] \perp$ will hold at each $o \in O$. Because satisfaction is entirely arbitrary at open worlds, the usual satisfaction-clauses do not apply and accessibility-relations have no effect.

Since $\simeq$ is defined in terms of $\approx$ and $\equiv_I$, it extends the set of indistinguishable states with those that cannot be distinguished on the basis of the insight one has relative to I. Unlike $\approx$, $\simeq$ isn't an equivalence-relation. The disjunctive condition in Clause 1 merely ensures that $\simeq$ has a non-empty extension over O, and thus that $\simeq$ can be used to express facts about which open worlds cannot be discriminated.

Before we move on to the definition of explicit belief, consider first the following two alternative definitions for explicit knowledge.

Version 1 φ is explicitly known at s iff $s \in \|[\approx]\varphi\|$ and $\varphi \in I$.
Version 2 φ is explicitly known at s iff $s \simeq t$ implies $t \in \|\varphi\|$.

Unless we impose further conditions on the set of open worlds, the above two versions need not be equivalent. Indeed, when the set of open worlds is empty, the second, but not the first version will, irrespective of I still result in deductive omniscience.[4]

Despite this divergence, both versions yield at least minimally acceptable accounts of explicit knowledge. A similar choice is, in view of our characterisation of *tentative inferences* as inferences that may have to be retracted in view of an enhanced insight in one's premises, not available for explicit belief. We have no other choice than to generalise (Min) and (Rel):

$$s \xrightarrow{mi} t \Leftrightarrow s \simeq t \;\&\; \forall u((t \simeq u \;\&\; t \sim u) \Rightarrow t \preceq u) \tag{i-Min}$$

$$s \xrightarrow{ri} t \Leftrightarrow s \simeq t \;\&\; \forall \omega \in \Omega \; (t \in \|\omega\| \Rightarrow \exists u(s \xrightarrow{mi} u \;\&\; u \in \|\omega\|)) \tag{i-Rel}$$

As it stands, the insight-set I and the indistinguishability-relation $\approx$ do not constrain each other, and can even be totally unrelated. This is undesirable in the present context, as it allows for a situation where sheer reasoning does not suffice to make all implicit knowledge explicit. Hence, it makes sense to require that the totality of one's implicit knowledge should not exceed the deductive closure[5] of one's explicit knowledge. When formalised as (†), this requirement implies, but is not implied by (‡).[6]

$$S \cap \|\{\varphi : t \in \|[\approx]\varphi\| \;\&\; \varphi \in I\}\| = \{s \in S : t \approx s\} \tag{†}$$

$$\simeq \cap \; (S \times S) = \approx \tag{‡}$$

In view of the role of $\approx$ and $\simeq$ in, respectively, (Min) and (Rel), and (i-Min) and (i-Rel), the identity expressed by (‡) also constrains the relation between $\xrightarrow{m}$ and $\xrightarrow{mi}$ as well as between $\xrightarrow{r}$ and $\xrightarrow{ri}$. Since in view of their definitions we already have $\xrightarrow{mi} = \xrightarrow{m}$ and $\xrightarrow{ri} = \xrightarrow{r}$ whenever $\simeq$ and $\approx$ coincide, (‡) entails that when all implicit knowledge has been made explicit (or, equivalently, whenever I is closed under sub-formulae), explicit and implicit beliefs will coincide as well.

[4] Note that even in that case implicit and explicit knowledge need not coincide: It can still be the case that $s \simeq t$ while $s \not\approx t$. We shall come back to this issue below.

[5] **L**-closure, *not* closure under some adaptive consequence relation.

[6] As a counterexample, just take a case where $[\approx]p$ holds at t, but $I = \{\neg p\}$.

Now that we've imposed some further constraints on our model, let us take a brief look at how it can be used to model tentative inference. The basic idea is this: When the insight-set isn't closed under sub-formulae, there can be states $s \in S$ and $t \in O$ such that $s \simeq t$. Moreover, it could (to give an extreme example) even be the case that every u such that $s \xrightarrow{m} u$ is in fact an open world. In that case, no such u is really possible relative to one's implicit knowledge, but one simply hasn't found out yet. As a result, each such u could be excluded by sheer reasoning, and thereafter only closed worlds (that were previously deemed *too abnormal*) would be considered possible. In summary: We have a model of tentative inference because (a) we can have open worlds such that $s \simeq t$, and (b) there is no closed world v such that $s \simeq v$ and $v \prec t$.

To be sure, this model is far from being perfect. Its main drawback is that whenever some sub-formula is in the insight-set I, this will have an impact on the insight we have in every formula that has that formula as a sub-formula. Nevertheless, the present coarse version is sufficient to illustrate the basic idea of insight (but see the remarks in the final section) as well as the resulting inferential dynamics.

4 Inferential Dynamics

A model for tentative inference is only of interest if we can also account for the inferences themselves, rather than just for their outcomes. In the present section we provide a generic model for inferential dynamics that is based on two basic types of analytic inferences (in a non-modal, purely propositional logic): gaining insight in a formula of type α and gaining insight in a formula of type β. This terminology, which is taken from the literature on tableau-systems [6,12], allows us to focus only on those inferences that lead to changes in I, and to do so without having to refer to a specific underlying logic.

The distinction between α and β formulae is related to a distinction between *branching* and *linear* rules; a distinction that is familiar from tableau-systems and sequent-calculi, but that is typically absent from natural deduction systems. To gain insight in a formula of type α, one must apply a linear rule to that formula; to gain insight in a formula of type β, one must apply a branching rule to that formula. While linear rules merely extend the present alternative (e.g. adding p and q to analyse $p \wedge q$), branching rules create two new alternatives and extend each of these differently (e.g. creating two copies of the present alternative, and extend one with p and the other with q, to analyse $p \vee q$). Depending on the actual formalism one uses, alternatives are different entities: branches of a tableau or a proof, tableaux and copies of tableaux, etc. In our formalism the alternatives are only indirectly modelled by the insight-set I.

In the description of the actions we shall adopt the following notational conventions. When φ is of type α, we analyse φ in its α_1 and α_2 components (which need not be the immediate sub-formulae of φ). Similarly, we analyse formulae of type β in its β_1 and β_2 components. Finally, we stipulate that atoms as well as Boolean negations of such atoms are of neither type.

We model the action of analysing φ as a relation $\xrightarrow{\varphi}$ between pointed models:

Definition 5 (Analysing φ). *Where $\mathfrak{M}, s = (S, O, \approx, I, \Omega, \|\cdot\|_{\mathfrak{M}}), s$ and $\mathfrak{M}', s' = (S', O', \approx', I', \Omega', \|\cdot\|_{\mathfrak{M}'}), s'$ are two pointed i-Doxastic Abnormality Models, we say that $\mathfrak{M}, s \xrightarrow{\varphi} \mathfrak{M}', s'$ iff:*

1. $S = S'$, $O = O'$, $\approx = \approx'$, $\Omega = \Omega'$, $\|\cdot\|_{\mathfrak{M}} = \|\cdot\|_{\mathfrak{M}'}$, *and* $s = s'$.
2. $\varphi \in I$.
3. $s \in \|\langle\approx\rangle\varphi\|$.
4. *If φ is of type α, φ_1 and φ_2 are its α_1 and α_2 components, then* $I' = I \cup \{\varphi_1, \varphi_2\}$.
5. *If φ is of type β, φ_1 and φ_2 are its β_1 and β_2 components, then* $I' = I \cup \{\varphi_1, \varphi_2\}$.
6. *If φ is a literal (atom or* Boolean *negation of an atom),* $I' = I$.
7. $\simeq' \subseteq \simeq$ *(as required by the definition of $\simeq$ in Definition 4).*

In this definition, the Clauses 2 and 3 are the pre-conditions for the inferential action, and the Clauses 4–6 its post-conditions. One note-worthy feature of this type of action is that the formulae that are analysed by such actions do not need to be explicitly known formulae. Instead, they only need to be included in the insight-set and be true at some closed world that is indistinguishable from the actual world. This weaker pre-condition is entirely natural from the perspective of analytic calculi, but can also be related to hypothetical reasoning (gaining insight in a hypothesis) in a natural deduction system. Another remarkable feature is that branching and linear rules do not require a different treatment. This is because the conjunctive and disjunctive behaviour of α and β formulae is already enforced by the closed worlds; the only thing I needs to take care of is the exclusion of open worlds that do not respect this behaviour.

The effect of an analysing-action can be described as follows: By gaining insight in φ, one will be able to distinguish the actual world s where $\langle\approx\rangle\varphi$ holds from any open world where φ is true, but neither of its α (or β) components are true. This is because (a) its α (or one of its β) components are already true at some t such that $s \approx t$ (because t is a closed world), but for no open world u where φ is true but neither of its α (or β) components are it will be that case that $t \equiv_{I'} u$.

5 Concluding Remarks

In [7] it is argued that impossible worlds are both a liability for models for non-ideal belief, as well as an unnecessary modification of the possible-world framework. In this paper we showed that this isn't always so. Here, open worlds are indispensable as well as unobjectionable. This is because, contrary to what is the case with plain deductive inference, non-ideal agents will unavoidably make mistakes. Even more so, in order to be rational at all, non-ideal agents have no other choice than to make mistakes they will only be able to correct later on. This is an important lesson from defeasible inference.

As mentioned, the present model remains too coarse to provide a realistic model of insight-driven inferential semantics. This is because the insight-set I imposes a global type of insight in one's premise: either we can discriminate between the truth and falsity of some sub-formula, or we can't. This problem can be avoided by moving to a more fragmented approach, like the one described in [1], but where the fragmentation is premise-based, and each fragment comes with its own insight-set. This refinement does not affect the basic ideas and concepts introduced in this paper, but still requires some modifications to deal with the move to multiple indistinguishability-relations, and especially with the problem of higher-order beliefs. We leave this for another occasion.

References

1. Allo, P.: The many faces of closure and introspection. Journal of Philosophical Logic, 1–34 (2011)
2. Allo, P.: Adaptive logic as a modal logic. Studia Logica (forthcoming)
3. Baltag, A., Smets, S.: A Qualitative Theory of Dynamic Interactive Belief Revision, pp. 11–58. Amsterdam University Press, Amsterdam (2008)
4. Batens, D.: Blocks. the clue to dynamic aspects of logic. Logique & Analyse 38(150-152), 285–328 (1995)
5. Batens, D.: A universal logic approach to adaptive logics. Logica Universalis 1(1), 221–242 (2007)
6. D'Agostino, M.: Tableau Methods for Classical Propositional Logic, pp. 45–123. Kluwer Academic Publishers, Dordrecht (1999)
7. Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence 34, 39–76 (1988)
8. Konolige, K.: A deductive model of belief, pp. 377–381. Morgan Kaufmann Publishers Inc., Karlsruhe (1983)
9. Levesque, H.J.: A logic of implicit and explicit belief. In: National Conference on Artificial Intelligence, Houston, Texas, pp. 198–202 (1984)
10. Priest, G.: Towards Non-Being. The logic and metaphysics of intentionality. Oxford University Press, Oxford (2005)
11. Reiter, R.: A logic for default reasoning. Artificial Intelligence 13(1-2), 81–132 (1980)
12. Smullyan, R.M.: First-Order Logic. Dover Publications, New York (1995)
13. van Benthem, J., Girard, P., Roy, O.: Everything else being equal: A modal logic for ceteris paribus preferences. Journal of Philosophical Logic 38(1), 83–125 (2009)
14. Velzquez-Quesada, F.R.V.: Small steps in dynamics of information. PhD thesis (2011)

Decidability for Justification Logics Revisited

Samuel Bucheli*, Roman Kuznets**, and Thomas Studer

Institut für Informatik und angewandte Mathematik
Universität Bern
{bucheli,kuznets,tstuder}@iam.unibe.ch
http://www.iam.unibe.ch/ltg

Abstract. Justification logics are propositional modal-like logics that instead of statements A *is known* include statements of the form A *is known for reason* t where the term t can represent an informal justification for A or a formal proof of A. In our present work, we introduce model-theoretic tools, namely: filtrations and a certain form of generated submodels, in the context of justification logic in order to obtain decidability results. Apart from reproving already known results in a uniform way, we also prove new results. In particular, we use our submodel construction to establish decidability for a justification logic with common knowledge for which so far no decidability proof was available.

Keywords: Justification logic, decidability, filtration.

1 Introduction

Justification logics are epistemic logics that explicitly include justifications for the agents' knowledge [3,4]. The first logic of this kind, the *Logic of Proofs* LP, was developed by Artemov to provide the modal logic S4 with provability semantics [1,2]. The language of justification logics has also been used to create a new approach to the logical omniscience problem [5], to study self-referential proofs [14], and to explore the evidential dynamics of public announcements [8,10].

Instead of statements A *is known*, denoted $\Box A$, justification logics reason about justifications for knowledge by using the construct $t : A$ to formalize statements t *is a justification for* A, where, dependent on the application, the *evidence term* t can be viewed as an informal justification or a formal mathematical proof. For an example see Fig. 1 where the axioms of the justification logic LP are listed alongside the axioms of S4 to point out the correspondence of the operations on evidence terms to standard modal axioms. This correspondence (as well as many other such correspondences between certain modal logics and justification logics) can be shown in a formal way. While it is easy to see, that replacing all justification terms by $\Box$ in a theorem of LP yields a theorem of S4, the other direction is much more involved and known as the realization theorem. See [7] for a uniform proof and survey of realization theorems for all logics in the modal cube.

* Supported by Swiss National Science Foundation grant 200021–117699.
** Supported by Swiss National Science Foundation grant PZ00P2–131706.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 166–181, 2013.

S4 axioms	LP axioms	
$\Box(A \to B) \to (\Box A \to \Box B)$	$t : (A \to B) \to (s : A \to t \cdot s : B)$	(application)
$\Box A \to A$	$t : A \to A$	(reflexivity)
$\Box A \to \Box\Box A$	$t : A \to !t : t : A$	(inspection)
	$t : A \vee s : A \to t + s : A$	(sum)

Fig. 1. Non-propositional axioms of S4 and LP

Fitting [11] introduced epistemic semantics for justification logics. The so-called Fitting models are Kripke models (W, R, ν) augmented by an evidence relation $\mathcal{E}$ that states which terms are admissible evidence for which formulae.

Filtrations are a tool in modal logic for obtaining from a given, usually infinite, model a smaller, usually finite, model by factoring the set of worlds with respect to a certain equivalence relation. As noted in [6], filtrations were first introduced in [19] and given their name in [15]. Given the close relationship between Fitting models and Kripke models, it is a natural task to adopt filtrations for justification logics. The crucial step is of course to take into account the evidence relation when identifying states.

Filtrations are often used to prove a finite model property and thereby establish decidability of a given modal logic, see e.g. [6]. Decidability for the justification logics presented here was originally shown in [12,13,16]. We adapt the filtration technique from modal logic to obtain an alternative uniform proof of decidability for these justification logics. We then apply the newly developed technique to establish the decidability of the multi-agent justification logic with common knowledge presented in [9].

In Section 2, we introduce the syntax and semantics of the justification logics we are using. In Section 3, we define filtrations for justification logics and prove their basic properties. We treat two specific examples of filtrations in Sections 4 and 5. In Section 6, we use these two examples to prove the decidability of the defined justification logics. This also leads us to investigate general properties necessary for the decidability of justification logics and enables us to prove the decidability of a multi-agent justification logic with common knowledge in Section 7.

2 Justification Logics

Justification terms are built from constants c_i and variables x_i according to the following grammar:

$$t ::= c_i \mid x_i \mid (t \cdot t) \mid (t + t) \mid !t \quad .$$

We denote the set of terms by Tm. Formulae are built from atomic propositions p_i according to the following grammar:

$$F ::= p_i \mid \neg F \mid (F \to F) \mid t : F \quad .$$

Prop denotes the set of atomic propositions and Fm denotes the set of formulae.

The set Sub(F) of subformulae of a given formula F is defined inductively as follows

- $\mathrm{Sub}(p_i) := \{p_i\}$
- $\mathrm{Sub}(\neg F) := \{\neg F\} \cup \mathrm{Sub}(F)$
- $\mathrm{Sub}(F_1 \to F_2) := \{F_1 \to F_2\} \cup \mathrm{Sub}(F_1) \cup \mathrm{Sub}(F_2)$
- $\mathrm{Sub}(t : F) := \{t : F\} \cup \mathrm{Sub}(F)$

A set of formulae $\Phi \subseteq \mathrm{Fm}$ is *closed under subformulae* if $\bigcup_{F\in\Phi} \mathrm{Sub}(F) \subseteq \Phi$.

The axioms of $\mathsf{J_{CS}}$ consist of all instances of the following schemes:

A1 finitely many schemes axiomatizing classical propositional logic
A2 $t : (A \to B) \to (s : A \to t \cdot s : B)$
A3 $t : A \vee s : A \to t + s : A$

We will consider extension of $\mathsf{J_{CS}}$ by the following axioms schemes.

(jd) $t : \bot \to \bot$
(jt) $t : A \to A$
(j4) $t : A \to !t : t : A$

A *constant specification* CS for a logic L is any subset

$$\mathsf{CS} \subseteq \{c : A \mid c \text{ is a constant and } A \text{ is an axiom of } \mathsf{L}\}.$$

A constant specification CS for a logic L is called

1. *axiomatically appropriate* if for each axiom A of $\mathsf{L_{CS}}$ there is a constant c such that $c : A \in \mathsf{CS}$
2. *schematic* if for each constant c the set $\{A \mid c : A \in \mathsf{CS}\}$ consists of one or several (possibly zero) axiom schemes, i.e., every constant justifies certain axiom schemes.

For a constant specification CS the deductive system $\mathsf{J_{CS}}$ is the Hilbert system given by the axioms A1–A3 and by the rules modus ponens and axiom necessitation:

$$\frac{A \quad A \to B}{B}\,(\mathrm{MP}) \ , \qquad \frac{c : A \in \mathsf{CS}}{\underbrace{!!\cdots!}_{n}c : \underbrace{!\cdots!}_{n-1}c : \cdots :!!c :!c : c : A}\,(\mathrm{AN!}) \ ,$$

where $n \geq 0$. In the presence of the j4 axiom a simplified axiom necessitation rule can be used:

$$\frac{c : A \in \mathsf{CS}}{c : A}\,(\mathrm{AN}) \ .$$

Table 1 defines the various logics we consider.

We now present the semantics for these logics

Definition 1 (Evidence Relation). *Let (W, R) be a Kripke frame, i.e., $W \neq \varnothing$ and $R \subseteq W \times W$, and CS be a constant specification. An admissible evidence relation $\mathcal{E}$ for a logic $\mathsf{L_{CS}}$ is a subset of $Tm \times Fm \times W$ that satisfies the closure conditions:*

Table 1. Deductive Systems

	A1	A2	A3	jd	jt	j4	MP	AN!	AN
J_{CS}	✓	✓	✓				✓	✓	
$\mathsf{JD}_{\mathsf{CS}}$	✓	✓	✓	✓			✓	✓	
$\mathsf{JT}_{\mathsf{CS}}$	✓	✓	✓		✓		✓	✓	
$\mathsf{JD4}_{\mathsf{CS}}$	✓	✓	✓	✓		✓	✓		✓
$\mathsf{J4}_{\mathsf{CS}}$	✓	✓	✓			✓	✓		✓
$\mathsf{LP}_{\mathsf{CS}}$	✓	✓	✓		✓	✓	✓		✓

1. *if* $(s, A, w) \in \mathcal{E}$ *or* $(t, A, w) \in \mathcal{E}$*, then* $(s + t, A, w) \in \mathcal{E}$
2. *if* $(s, A \to B, w) \in \mathcal{E}$ *and* $(t, A, w) \in \mathcal{E}$*, then* $(s \cdot t, B, w) \in \mathcal{E}$

Depending on whether or not the logic L_{CS} *contains the j4 axiom, the evidence function has to satisfy one of the following two sets of closure conditions. If* L_{CS} *does not include the j4 axiom, then the additional requirement is:*

3. *if* $c : A \in \mathsf{CS}$ *and* $w \in W$*, then* $(\underbrace{!! \cdots !}_{n} c, \underbrace{! \cdots !}_{n-1} c : \cdots :!!c :!c : c : A, w) \in \mathcal{E}$

If L_{CS} *includes the j4 axiom, then the additional requirement is:*

4. *if* $c : A \in \mathsf{CS}$ *and* $w \in W$*, then* $(c, A, w) \in \mathcal{E}$
5. *if* $(t, A, w) \in \mathcal{E}$*, then* $(!t, t : A, w) \in \mathcal{E}$
6. *if* $(t, A, w) \in \mathcal{E}$ *and* wRv*, then* $(t, A, v) \in \mathcal{E}$

If we drop condition 6, then we say $\mathcal{E}$ *is a* t-evidence relation*.*
Sometimes we use $\mathcal{E}(s, A, w)$ *for* $(s, A, w) \in \mathcal{E}$*.*

Definition 2 (Evidence Bases).

1. *An* evidence base $\mathcal{B}$ *is a subset of* $Tm \times Fm \times W$*.*
2. *An evidence relation* $\mathcal{E}$ *is* based on $\mathcal{B}$*, if* $\mathcal{B} \subseteq \mathcal{E}$*.*

The closure conditions in the definition of admissible evidence function give rise to a monotone operator. The minimal evidence relation based on $\mathcal{B}$ is the least fixed point of that operator and thus always exists.

Definition 3 (Model). *Let* CS *be a constant specification. A* Fitting model *for a logic* L_{CS} *is a quadruple* $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ *where*

- (W, R) *is a Kripke frame such that*
  - *if* L_{CS} *includes the j4 axiom, then R is transitive;*
  - *if* L_{CS} *includes the jt axiom, then R is reflexive;*
  - *if* L_{CS} *includes the jd axiom, then R is serial.*
- $\mathcal{E}$ *is an admissible evidence relation for* L_{CS} *over the frame* (W, R)*,*
- $\nu : Prop \to \mathcal{P}(W)$*, called a valuation function.*

Definition 4 (Satisfaction Relation). *The relation of formula A being satisfied in a model* $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ *at a world* $w \in W$ *is defined by induction on the structure of A by*

– *$\mathcal{M}, w \Vdash p_i$ if and only if $w \in \nu(p_i)$*
– *$\Vdash$ commutes with Boolean connectives*
– *$\mathcal{M}, w \Vdash t : B$ if and only if*
  1) *$\mathcal{M}, v \Vdash B$ for all $v \in W$ with wRv and*
  2) *$(t, B, w) \in \mathcal{E}$*

We say a formula A is valid in a model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ if for all $w \in W$ we have $\mathcal{M}, w \Vdash A$. We say a formula A is valid for a logic L_{CS} if for all models $\mathcal{M}$ for L_{CS} we have that A is valid in $\mathcal{M}$.

The logics defined above are sound and complete (with a restriction in case of the logics containing the jd axiom). See [3,11,17] for the full proofs of the following results.

Soundness can be obtained by an easy induction on the derivation of the formula.

Theorem 5 (Soundness). *Let CS be a constant specification. If a formula A is derivable in a logic L_{CS}, then A is valid for L_{CS}.*

For completeness a canonical model construction is used. The axiomatical appropriateness of the constant specification in case the logic contains the jd axiom is necessary to show the seriality condition on the accessibility relation.

Theorem 6 (Completeness)

1. *Let CS be a constant specification. If a formula A is not derivable in $\mathsf{L}_{\mathsf{CS}} \in \{\mathsf{J}_{\mathsf{CS}}, \mathsf{JT}_{\mathsf{CS}}, \mathsf{J4}_{\mathsf{CS}}, \mathsf{LP}_{\mathsf{CS}}\}$, then there exists a model $\mathcal{M}$ for L_{CS} with $\mathcal{M}, w \nVdash A$ for some world w in $\mathcal{M}$.*
2. *Let CS be an axiomatically appropriate constant specification. If a formula A is not derivable in $\mathsf{L}_{\mathsf{CS}} \in \{\mathsf{JD}_{\mathsf{CS}}, \mathsf{JD4}_{\mathsf{CS}}\}$, then there exists a model $\mathcal{M}$ for L_{CS} with $\mathcal{M}, w \nVdash A$ for some world w in $\mathcal{M}$.*

3 Filtrations

Given the close relationship of models for justification logics to Kripke models, it is not surprising that the two definitions of filtrations look very similar. The major difference is that we have to take the evidence relation into consideration. In modal logic we identify worlds that behave the same way, whereas in justification logic we identify worlds that behave the same way *for the same reason.*

Definition 7 (Filtration). *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ some set of formulae that is closed under subformulae. We define an equivalence relation $=_\Phi$ on W by setting $w =_\Phi v$ if and only if for all $A \in \Phi$*

$$M, w \Vdash A \text{ if and only if } M, v \Vdash A$$

and for all $t : B \in \Phi$

$$\mathcal{E}(t, B, w) \text{ if and only if } \mathcal{E}(t, B, v).$$

We denote the equivalence classes of Φ *by* $[w]_\Phi$. *When* Φ *is clear from the context, we will often only write* $[w]$ *instead of* $[w]_\Phi$.

A model $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ *is called a* filtration of $\mathcal{M}$ through Φ *if it satisfies the following:*

1. $W_\Phi = \{[w]_\Phi \mid w \in W\}$
2. R_Φ *satisfies*
   (R1) *for all* $w, v \in W$ *if* $R(w, v)$, *then* $R_\Phi([w]_\Phi, [v]_\Phi)$
   (R2) *for all* $[w]_\Phi, [v]_\Phi \in W_\Phi$, *if* $R_\Phi([w]_\Phi, [v]_\Phi)$, *then for any* $t : B \in \Phi$ *we have*
   $$\text{if } \mathcal{M}, w \Vdash t : B \text{ then } \mathcal{M}, v \Vdash B$$
3. $\mathcal{E}_\Phi$ *satisfies*
   (E1) *for all* $w \in W$ *and* $t : B \in \Phi$ *we have*
   $$\text{if } \mathcal{M}, w \Vdash t : B \text{ then } (t, B, [w]_\Phi) \in \mathcal{E}_\Phi$$
   (E2) *for all* $w \in W$ *and* $t : B \in \Phi$ *we have*
   $$\text{if } (t, B, [w]_\Phi) \in \mathcal{E}_\Phi \text{ then } (t, B, w) \in \mathcal{E}$$
4. ν_Φ *satisfies for all atomic propositions* $p \in \Phi$
   $$\nu_\Phi(p) = \{[w]_\Phi \mid w \in \nu(p)\}$$

There are two major changes of the definition compared to the case for modal logic. The first change concerns the definition of the equivalence relation to identify worlds. Whereas a modal formula $\Box B$ can only fail due to the existence of an accessible world not satisfying B, a justification formula $t : B$ might fail in two ways: either B is not satisfied in an accessible world or t is not admissible evidence for B at the current world. So we have to refine our equivalence relation to only identify worlds that do not only satisfy the same formulae but also behave the same with respect to the evidence relation. The second change concerns the evidence relation of the filtration: it has to satisfy conditions similar to the Min- and Max-conditions (R1) and (R2) for the accessibility relation.

The crucial property of a filtration of a model through Φ is that the behavior of the model and the filtration is the same with respect to formulae in Φ:

Lemma 8. *Let* $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ *be a model,* Φ *a set of formulae closed under subformulae, and* $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ *a filtration of* $\mathcal{M}$ *through* Φ. *Then for all worlds* $w \in W$ *and formulae* $A \in \Phi$ *we have*

$$\mathcal{M}_\Phi, [w]_\Phi \Vdash A \text{ if and only if } \mathcal{M}, w \Vdash A.$$

Proof. The proof is by induction on the structure of A. The case for propositional variables is immediate by the definition of ν_Φ and the cases for the propositional connectives are immediate by the induction hypothesis. Let us now consider the case $A = t : B$.

First we show the direction from right to left. Assume $\mathcal{M}, w \Vdash t : B$. If $R_\Phi([w]_\Phi, [v]_\Phi)$, then by (R2) we have $\mathcal{M}, v \Vdash B$. By the induction hypothesis we get $\mathcal{M}_\Phi, [v]_\Phi \Vdash B$. Further from (E1) we get $(t, B, [w]_\Phi) \in \mathcal{E}_\Phi$ and thus $\mathcal{M}_\Phi, [w]_\Phi \Vdash t : B$.

For the other direction suppose $\mathcal{M}_\Phi, [w]_\Phi \Vdash t : B$, that is

$$\mathcal{M}_\Phi, [v]_\Phi \Vdash B \text{ for all } [v]_\Phi \text{ with } R_\Phi([w]_\Phi, [v]_\Phi) \tag{1}$$

$$(t, B, [w]_\Phi) \in \mathcal{E}_\Phi \tag{2}$$

If $R(w, v)$, then by (R1) also $R_\Phi([w]_\Phi, [v]_\Phi)$ and by (1) and the induction hypothesis we get $\mathcal{M}, v \Vdash B$. Furthermore, from (2) and (E2) we get $\mathcal{E}(t, B, w)$ and we conclude $\mathcal{M}, w \Vdash t : B$.

A filtration inherits some conditions on the accessibility relations. Furthermore, a filtration through a finite set has finitely many worlds.

Lemma 9. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model, Φ a set of formulae closed under subformulae, and $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ a filtration of $\mathcal{M}$ through Φ.*

1. *If R is serial, so is R_Φ.*
2. *If R is reflexive, so is R_Φ.*
3. *If Φ is finite, then so is W_Φ.*

Proof. The first two claims follow immediately from (R1). The last claim follows from the fact that each element $[w]_\Phi \in W_\Phi$ can be characterized by the set of formulae $A \in \Phi$ that hold in $[w]_\Phi$ as well as the set of formulae $t : B \in \Phi$ with $\mathcal{E}_\Phi(t, B, [w]_\Phi)$ and the fact that $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi)$ has only finitely many elements.

4 Non-transitive Case

As a first example we will define filtrations for logics not containing the j4 axiom.

Definition 10 (Filtration: Non-transitive Case). *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ a set of formulae closed under subformulae. We consider the filtration $\mathcal{M}^{nt}_\Phi = (W^{nt}_\Phi, R^{nt}_\Phi, \mathcal{E}^{nt}_\Phi, \nu^{nt}_\Phi)$ that is given by*

1. *W^{nt}_Φ is the set of equivalence classes induced by $=_\Phi$*
2. *$R^{nt}_\Phi([w], [v])$ if and only if for all $t : B \in \Phi$ we have $\mathcal{M}, w \Vdash t : B$ implies $\mathcal{M}, v \Vdash B$*

3. $\mathcal{E}^{\mathsf{nt}}_{\Phi}$ *is the minimal evidence relation based on* $\mathcal{B}^{\mathsf{nt}}_{\Phi}$*, where*

$$\mathcal{B}^{\mathsf{nt}}_{\Phi}(t, B, [v]) \text{ if and only if } t : B \in \Phi \text{ and } \mathcal{E}(t, B, v).$$

4. ν^{nt}_{Φ} *is given by*

$$\nu^{\mathsf{nt}}_{\Phi}(p) = \begin{cases} \{[w] \mid w \in \nu(p)\} & \text{if } p \in \Phi, \\ \emptyset & \text{otherwise.} \end{cases}$$

Lemma 11. *$\mathcal{M}^{\mathsf{nt}}_{\Phi}$ is a filtration of $\mathcal{M}$ through Φ.*

Proof. We have to check the following conditions.

(R1) Assume $R(w, v)$. If $\mathcal{M}, w \Vdash t : B$, then $\mathcal{M}, v \Vdash B$. Thus we conclude $R^{\mathsf{nt}}_{\Phi}([w], [v])$.

(R2) Let $t : B \in \Phi$ and $R^{\mathsf{nt}}_{\Phi}([w], [v])$. If $\mathcal{M}, w \Vdash t : B$, then we get $\mathcal{M}, v \Vdash B$ immediately from the definition of R^{nt}_{Φ}.

(E1) Assume $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. We have $\mathcal{E}(t, B, w)$ and we immediately get $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ by the definition of $\mathcal{E}^{\mathsf{nt}}_{\Phi}$.

(E2) We show for all $t : B$, not only for those contained in Φ, that for all $w' \in [w]$

$$\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w]) \text{ implies } \mathcal{E}(t, B, w') \ .$$

We proceed by induction on the construction of $\mathcal{E}^{\mathsf{nt}}_{\Phi}$.

- If $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because $\mathcal{B}^{\mathsf{nt}}_{\Phi}(t, B, [w])$, then by definition of $\mathcal{B}^{\mathsf{nt}}_{\Phi}$ we have that $t : B \in \Phi$ and $\mathcal{E}(t, B, w'')$ for some $w'' \in [w]$. By $w' =_{\Phi} w''$ we conclude $\mathcal{E}(t, B, w')$.
- If $t = t_1 + t_2$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because of $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_i, B, [w])$ (for some $i \in \{1, 2\}$), then by induction hypothesis we get that $\mathcal{E}(t_i, B, w')$ and thus also $\mathcal{E}(t_1 + t_2, B, w')$ by the closure conditions on $\mathcal{E}$.
- If $t = t_1 \cdot t_2$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because there is an $A \in \mathrm{Fm}$ such that $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_1, A \to B, [w])$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_2, A, [w])$, then by induction hypothesis $\mathcal{E}(t_1, A \to B, w')$ and $\mathcal{E}(t_2, A, w')$. So, by the closure conditions, we get $\mathcal{E}(t_1 \cdot t_2, B, w')$.
- The case for axiom necessitation is trivial, as we have

$$\mathcal{E}(\underbrace{! \cdots !!}_{n} c, \underbrace{! \cdots !!}_{n-1} c : \cdots :!c : c : A, v)$$

for any world $v \in W$.

5 Transitive Case

The case for logics containing the j4 axiom is a bit more involved, as now the accessibility relation of the filtration has to be transitive, which is not guaranteed by the definition of filtration.

Definition 12 (Filtration: Transitive Case). *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ a set of formulae closed under subformulae. We consider the filtration $\mathcal{M}_\Phi^{tr} = (W_\Phi^{tr}, R_\Phi^{tr}, \mathcal{E}_\Phi^{tr}, \nu_\Phi^{tr})$ that is given by*

1. *W_Φ^{tr} is the set of equivalence classes induced by $=_\Phi$*
2. *$R_\Phi^{tr}([w],[v])$ if and only if for all $t : B \in \Phi$ we have $\mathcal{M}, w \Vdash t : B$ implies $\mathcal{M}, v \Vdash B \wedge t : B$*
3. *$\mathcal{E}_\Phi^{tr}$ is the minimal t-evidence relation based on $\mathcal{B}_\Phi^{tr}$, where*

$$\mathcal{B}_\Phi^{tr}(t, B, [v]) \textit{ if and only if } t : B \in \Phi \textit{ and } \mathcal{M}, v \Vdash t : B.$$

4. *ν_Φ^{tr} is given by*

$$\nu_\Phi^{tr}(p) = \begin{cases} \{[w] \mid w \in \nu(p)\} & \textit{if } p \in \Phi, \\ \emptyset & \textit{otherwise.} \end{cases}$$

As a first step we have to show that $\mathcal{E}_\Phi^{\mathsf{tr}}$ as defined is not only a t-evidence relation but an actual evidence relation.

Lemma 13. *$\mathcal{E}_\Phi^{tr}$ is an admissible evidence relation over $(W_\Phi^{tr}, R_\Phi^{tr})$.*

Proof. We have to show that condition (6) in Definition 1 holds, i.e., we have to show

$$\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [w]) \text{ and } R_\Phi^{\mathsf{tr}}([w],[v]) \text{ imply } \mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [v])$$

So assume $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [w])$ and $R_\Phi^{\mathsf{tr}}([w],[v])$. We now show $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [v])$ by induction on the construction of $\mathcal{E}_\Phi^{\mathsf{tr}}$.

Let $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [w])$ because of $\mathcal{B}_\Phi^{\mathsf{tr}}(t, B, [w])$. We have $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$ by definition of $\mathcal{B}_\Phi^{\mathsf{tr}}$. Since $R_\Phi([w],[v])$, it follows that $\mathcal{M}, v \Vdash B \wedge t : B$ and, in particular, $\mathcal{M}, v \Vdash t : B$. Thus, $\mathcal{B}_\Phi^{\mathsf{tr}}(t, B, [v])$ by definition of $\mathcal{B}_\Phi^{\mathsf{tr}}$, and clearly $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [v])$.

Let us now distinguish the different possible closure conditions from Definition 1:

1. Assume we have $t = t_1 + t_2$ and $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [w])$ because of $\mathcal{E}_\Phi^{\mathsf{tr}}(t_i, B, [w])$ for $i = 1$ or $i = 2$. Then by induction hypothesis $\mathcal{E}_\Phi^{\mathsf{tr}}(t_i, B, [v])$ and thus also $\mathcal{E}_\Phi^{\mathsf{tr}}(t, B, [v])$.
2. The case for $\cdot$ and ! follows immediately from the induction hypothesis in the same manner as the previous case.
3. The case for axiom necessitation (AN) trivially holds.

The accessibility relation for the filtration is transitive.

Lemma 14. *R_Φ^{tr} is transitive.*

Proof. Assume (a) $R_\Phi^{\mathsf{tr}}([w],[v])$ and (b) $R_\Phi^{\mathsf{tr}}([v],[u])$. Suppose $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. By (a) we get $\mathcal{M}, v \Vdash t : B$. Then by (b) we get $\mathcal{M}, u \Vdash B \wedge t : B$. Hence, we conclude $R_\Phi^{\mathsf{tr}}([w],[u])$.

Lemma 15. *$\mathcal{M}^{\mathsf{tr}}_{\Phi}$ is a filtration of $\mathcal{M}$ through Φ.*

Proof. We have to check the following conditions.

(R1) Assume $R(w, v)$. If $\mathcal{M}, w \Vdash t : B$, then $\mathcal{M}, v \Vdash B$ and $\mathcal{M}, w \Vdash !t : t : B$ which implies $\mathcal{M}, v \Vdash t : B$. Thus we conclude $R^{\mathsf{tr}}_{\Phi}([w], [v])$.

(R2) Let $t : B \in \Phi$ and $R_{\Phi}([w], [v])$. If $\mathcal{M}, w \Vdash t : B$, then we get $\mathcal{M}, v \Vdash B$ immediately from the definition of R^{tr}_{Φ}.

(E1) Assume $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. We immediately get $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w])$ by the definition of $\mathcal{E}^{\mathsf{tr}}_{\Phi}$.

(E2) As in the proof of Lemma 11 we can show for all $t : B$ and all $w' \in [w]$

$$\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w]) \text{ implies } \mathcal{E}(t, B, w') \ .$$

6 Decidability

The theorems in this section originate from [13]. We thus only give proof sketches for the sake of brevity.

Definition 16 (Finitary Model). *A model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ is called* finitary *if*

1. *W is finite,*
2. *there exists a finite base $\mathcal{B}$ such that $\mathcal{E}$ is the minimal evidence relation based on $\mathcal{B}$, and*
3. *the set $\{(w, p) \in W \times Prop \mid w \in \nu(p)\}$ is finite.*

Using filtrations we see that if a formula is satisfiable, then it is satisfiable in a finitary model. Thus we have the following

Lemma 17 (Completeness w.r.t. Finitary Models).

1. *Let $\mathsf{L_{CS}} \in \{\mathsf{J_{CS}}, \mathsf{JT_{CS}}, \mathsf{J4_{CS}}, \mathsf{LP_{CS}}\}$ and CS be a constant specification for L. If a formula A is not derivable in $\mathsf{L_{CS}}$, then there exists a finitary model $\mathcal{M}$ for $\mathsf{L_{CS}}$ with $\mathcal{M}, w \nVdash A$ for some world w in $\mathcal{M}$.*
2. *Let $\mathsf{L_{CS}} \in \{\mathsf{JD_{CS}}, \mathsf{JD4_{CS}}\}$ and CS be an axiomatically appropriate constant specification for L. If a formula A is not derivable in $\mathsf{L_{CS}}$, then there exists a finitary model $\mathcal{M}$ for $\mathsf{L_{CS}}$ with $\mathcal{M}, w \nVdash A$ for some world w in $\mathcal{M}$.*

Proof. Let CS be as required above. If A is not derivable in $\mathsf{L_{CS}}$, then by Theorem 6 there exists a model $\mathcal{M}$ for $\mathsf{L_{CS}}$ with $\mathcal{M}, v \nVdash A$ for some world v in $\mathcal{M}$. Now set $\Phi := \mathrm{Sub}(A)$ and let $\mathcal{M}_{\Phi}$ denote either $\mathcal{M}^{\mathsf{nt}}_{\Phi}$ or $\mathcal{M}^{\mathsf{tr}}_{\Phi}$ from Definitions 10 and 12 respectively, depending on whether $\mathsf{L_{CS}}$ contains the j4 axiom. It is easy to see that $\mathcal{M}_{\Phi}$ is a finitary model: by Lemma 9 the set of worlds is finite and, by definition of $\mathcal{M}_{\Phi}$, the evidence relation is finitely based and the valuation function satisfies condition 3 from Definition 16. Finally, since $\mathcal{M}_{\Phi}$ is a filtration of $\mathcal{M}$ through Φ by Lemma 11 or by Lemma 15, by Lemma 8 we have $\mathcal{M}_{\Phi}, [v] \nVdash A$.

Corollary 18. *All statements of Lemma 17 hold if an additional restriction is imposed that the domain of the model $\mathcal{M}$ be a finite subset of $\mathbb{N}$.*

Proof. The claim follows trivially from Lemma 17 by renaming worlds to natural numbers.

The following theorem is a simple instance of Post's theorem [18]: A set is decidable if and only if both the set and its complement are recursively enumerable.

Theorem 19. *A logic is decidable if it is recursively enumerable and is sound and complete with respect to a set $\mathcal{C}$ such that*

1. *$\mathcal{C}$ is a recursively enumerable set of finite models and*
2. *the relation $\mathcal{M}, w \Vdash A$ between models $\mathcal{M} \in \mathcal{C}$, worlds w in $\mathcal{M}$, and formulae A is decidable.*

Proof. We give a proof sketch, for full details cf. [13, Theorem 4.3.3]

Given a formula A, we can simultaneously enumerate theorems $B_0, B_1, \ldots$ of the logic and potential counter-models $\mathcal{M}_0, \mathcal{M}_1, \ldots \in \mathcal{C}$ and at each step check whether (a) $A = B_i$ or (b) $\mathcal{M}_i, w \nVdash A$ for some $w \in \mathcal{M}_i$. Eventually either (a) or (b) will hold for some i, thus indicating whether the logic proves A.

Lemma 20. *Let $\mathsf{L}_{\mathsf{CS}} \in \{\mathsf{J}_{\mathsf{CS}}, \mathsf{JD}_{\mathsf{CS}}, \mathsf{JD4}_{\mathsf{CS}}, \mathsf{JT}_{\mathsf{CS}}, \mathsf{J4}_{\mathsf{CS}}, \mathsf{LP}_{\mathsf{CS}}\}$. The set of finitary models for L_{CS} with the domain being a finite subset of $\mathbb{N}$ is recursively enumerable.*

Proof. We give a proof sketch, for full details cf. [13, Lemma 4.4.6].

It is obvious that the set of such models for J_{CS} can be recursively enumerated. Models of each of the other five logics must additionally satisfy certain conditions on the accessibility relation, some combination of transitivity, reflexivity, and seriality. Since each of these conditions can be effectively verified, the models of J_{CS} that are unsuitable for a given logic can be effectively removed from the enumeration of models for L_{CS}.

Lemma 21. *Let CS be a decidable schematic constant specification and $\mathsf{L}_{\mathsf{CS}} \in \{\mathsf{J}_{\mathsf{CS}}, \mathsf{JD}_{\mathsf{CS}}, \mathsf{JD4}_{\mathsf{CS}}, \mathsf{JT}_{\mathsf{CS}}, \mathsf{J4}_{\mathsf{CS}}, \mathsf{LP}_{\mathsf{CS}}\}$. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a finitary model for L_{CS}. Then the relation $\mathcal{M}, w \Vdash A$ between worlds $w \in W$ and formulae A is decidable.*

Proof. We give a proof sketch, for full details cf. [13, Corollary 4.4.8].

We can show this by induction on the formula A, the cases for propositions and Boolean connectives being trivial.

The crucial step is to show that the relation $\mathcal{E}(t, B, w)$ between terms $t \in \mathrm{Tm}$, formulae $B \in \mathrm{Fm}$ and worlds $w \in W$ is decidable (see [13, Lemma 4.4.7]).

Let $\mathcal{B}$ be the base for the minimal evidence relation $\mathcal{E}$ of $\mathcal{M}$. Given a fixed term t, we will construct a sequence of sets $\mathcal{E}_t^i(w)$ inductively, which can be seen as a partial evidence function that lists all formulae for which t or one of its subterms are admissible evidence at world w.

In order to keep the sets finite and as we are given a schematic constant specification, we will use variables $X, Y, \ldots$ ranging over schemes of formulae and variables $P, Q, \ldots$ ranging over formulae. Also, we assume that our constant specification is given in terms of schemes, i.e.

$$\mathsf{CS} = \{c : X \mid c \text{ is a constant and } X \text{ is a scheme}\}.$$

The sets are defined as follows

$$\mathcal{E}_t^0(w) := \{(s, B) \mid \mathcal{B}(s, B, w) \text{ and } s \in \mathrm{Sub}(t)\} \\ \cup \{(c, X) \mid c : X \in \mathsf{CS} \text{ and } c \in \mathrm{Sub}(t)\}$$

Assume $\mathcal{E}_t^n(w)$ has been constructed, in order to obtain $\mathcal{E}_t^{n+1}(w)$ add the following

- $(s_1 \cdot s_2, Y_1\sigma)$ for any $(s_1, X_1 \to Y_1) \in \mathcal{E}_t^n(w)$ and $(s_2, X_2) \in \mathcal{E}_t^n(w)$ such that the most general unifier σ of X_1 and X_2 exists and $s_1 \cdot s_2 \in \mathrm{Sub}(t)$
- $(s_1 \cdot s_2, Q)$ for any $(s_1, P) \in \mathcal{E}_t^n(w)$ and $(s_2, X_2) \in \mathcal{E}_t^n(w)$ where Q is a fresh variable over formulas and $s_1 \cdot s_2 \in \mathrm{Sub}(t)$
- $(s_1 + s_2, X)$ for any (s_1, X) or $(s_2, X) \in \mathcal{E}_t^n(w)$ with $s_1 + s_2 \in \mathrm{Sub}(t)$
- depending on whether the logic L_{CS} contains the j4 axiom, we distinguish the following two cases: If the logic does not contain the j4 axiom, we add
  - $(\underbrace{!! \cdots !}_{n+1} c, \underbrace{!! \cdots !}_{n} c : \ldots :!c : c : X)$ for any $c : X \in \mathsf{CS}$ with $\underbrace{!! \cdots !}_{n+1} c \in \mathrm{Sub}(t)$

  If the logic contains the j4 axiom, we add
  - $(!s, s : X)$ for any $(s, X) \in \mathcal{E}_t^n(w)$ with $!s \in \mathrm{Sub}(t)$
  - (s, X) for any $(s, X) \in \mathcal{E}_t^n(v)$ with $R(v, w)$ and $s \in \mathrm{Sub}(t)$

All the sets $\mathcal{E}_t^i(w)$ are finite. As W and $\mathrm{Sub}(t)$ are finite, there is an n easily computable from the size of W and the length of t such that $\mathcal{E}_t^n(w) = \mathcal{E}_t^i(w)$ for all $i \geq n$. Furthermore, we have $\mathcal{E}(t, B, w)$ if and only if B unifies with some X such that $(t, X) \in \mathcal{E}_t^n(w)$. Thus, the relation $\mathcal{E}(t, B, w)$ is decidable.

Corollary 22 (Decidability).

1. *Any justification logic in* $\{\mathsf{J}_{\mathsf{CS}}, \mathsf{JT}_{\mathsf{CS}}, \mathsf{J4}_{\mathsf{CS}}, \mathsf{LP}_{\mathsf{CS}}\}$ *with a decidable schematic* CS *is decidable.*
2. *Any justification logic in* $\{\mathsf{JD}_{\mathsf{CS}}, \mathsf{JD4}_{\mathsf{CS}}\}$ *with a decidable, schematic and axiomatically appropriate* CS *is decidable.*

Proof. All logics presented are obviously recursively enumerable. By Corollary 18, Lemma 20 and Lemma 21 all logics presented satisfy the conditions of Theorem 19 and are, therefore, decidable.

7 The Case of Common Knowledge

While the finiteness of the sets of worlds is a key feature of filtrations, the finite bases of our examples are due to the specific setup of the models and are by no means a necessary property of filtrations. On the other hand, if we start with a logic L_{CS} which we already know to be sound and complete with respect to a class of finite models, we can adapt the construction we used to finitely base the evidence function for the filtrations.

Definition 23. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ some set of formulae that is closed under subformulae. The Φ-generated submodel $\mathcal{M} \restriction \Phi$ of $\mathcal{M}$ is defined as $(W, R, \mathcal{E} \restriction \Phi, \nu \restriction \Phi)$ where*

1. *$\mathcal{E} \restriction \Phi$ is the minimal evidence relation based on $\mathcal{B}_\Phi$ where*

$$\mathcal{B}_\Phi(t, B, w) \text{ if and only if } t : B \in \Phi \text{ and } \mathcal{E}(t, B, w)$$

2. *$\nu \restriction \Phi$ is given by*

$$(\nu \restriction \Phi)(p) = \begin{cases} \{w \mid w \in \nu(p)\} & \text{if } p \in \Phi \\ \emptyset & \text{otherwise} \end{cases}$$

Like in the case for filtrations we get the following lemma.

Lemma 24. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model, Φ a set of formulae closed under subformulae, and $\mathcal{M} \restriction \Phi$ the Φ-generated submodel of $\mathcal{M}$. Then for all worlds w in $\mathcal{M}$ and formulae $A \in \Phi$ we have*

$$\mathcal{M} \restriction \Phi, w \Vdash A \text{ if and only if } \mathcal{M}, w \Vdash A.$$

Proof. The proof is by induction on A. The case for atomic propositions is immediate by the definition of $\nu \restriction \Phi$ and the cases for boolean connectives follow immediately by induction hypothesis. Let us consider the case when A is $t : B$.

So assume $\mathcal{M} \restriction \Phi, w \Vdash t : B$. We get $(t, B, w) \in \mathcal{E} \restriction \Phi$ and $\mathcal{M} \restriction \Phi, v \Vdash B$ for all $v \in W$ with $R(w, v)$. The latter gives us $\mathcal{M}, v \Vdash B$ by induction hypothesis whereas from the former we get $(t, B, w) \in \mathcal{E}$ as both $\mathcal{E}$ and $\mathcal{E} \restriction \Phi$ are based on $\mathcal{B}_\Phi$ and $\mathcal{E} \restriction \Phi$ is minimal with that property and hence $\mathcal{E} \restriction \Phi \subseteq \mathcal{E}$. So we have $\mathcal{M}, w \Vdash t : B$.

For the other direction assume $\mathcal{M}, w \Vdash t : B$. We have thus $\mathcal{E}(t, B, w)$ and $\mathcal{M}, v \Vdash B$ for all $v \in W$ with $R(w, v)$. Again, the latter gives us $\mathcal{M} \restriction \Phi, v \Vdash B$ by induction hypothesis and by the definition of $\mathcal{E} \restriction \Phi$ we immediately get $(t, B, w) \in \mathcal{E} \restriction \Phi$ from the former and thus $\mathcal{M} \restriction \Phi, w \Vdash t : B$.

We can use this technique (adapted to the multi-agent case) to establish decidability for the justification logic with common knowledge $\mathsf{LP}^{\mathsf{C}}_h$ that was introduced in [9].

The logic $\mathsf{LP}^{\mathsf{C}}_h$ is a multi-agent version of $\mathsf{LP}_{\mathsf{CS}}$ with additional axioms and operations on terms to deal with mutual and common knowledge. There are separate sets of terms for each agent $i \in \{1, \dots, h\}$ as well as for mutual knowledge E and for common knowledge C. The different kinds of knowledge suggest a slight change of notation. We will write $[t]_{\circledast}B$ to mean "t is a justification term of type $\circledast$ for B", where $\circledast \in \{1, \dots, h, \mathsf{E}, \mathsf{C}\}$. Furthermore $*$ will always denote an element of $\{1, \dots, h, \mathsf{C}\}$.

The logic $\mathsf{LP}_h^{\mathsf{C}}(\mathsf{CS})$ is given by the following axioms as well as the rules for modus ponens and axiom necessitation (for a given constant specification CS):

1. finitely many schemes axiomatizing classical propositional logic
2. $[t]_*(A \to B) \to ([s]_*A \to [t \cdot s]_*B)$ (application)
3. $[t]_*A \vee [s]_*A \to [t + s]_*A$ (sum)
4. $[t]_iA \to A$ (reflexivity)
5. $[t]_iA \to [!t]_i\,[t]_iA$ (inspection)
6. $[t_1]_1A \wedge \dots \wedge [t_h]_hA \to [\langle t_1, \dots, t_h\rangle]_{\mathsf{E}}A$ (tupling)
7. $[t]_{\mathsf{E}}A \to [\pi_i t]_iA$ (projection)
8. $[t]_{\mathsf{C}}A \to [\mathsf{ccl}_1(t)]_{\mathsf{E}}A, \quad [t]_{\mathsf{C}}A \to [\mathsf{ccl}_2(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A$ (co-closure)
9. $A \wedge [t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}}A) \to [\mathsf{ind}(t, s)]_{\mathsf{C}}A$ (induction)

The semantics for $\mathsf{LP}_h^{\mathsf{C}}$ is given by models

$$\mathcal{M} = (W, R_1, \dots, R_h, \mathcal{E}_1, \dots, \mathcal{E}_h, \mathcal{E}_{\mathsf{E}}, \mathcal{E}_{\mathsf{C}}, \nu)$$

where R_i are reflexive, transitive accessibility relations on W and $\mathcal{E}_{\circledast}$ are evidence relations satisfying closure conditions modeled on the axioms of $\mathsf{LP}_h^{\mathsf{C}}$ analogous to the logics presented in Section 2.[1] Furthermore we define $R_{\mathsf{E}} := \bigcup_{i=1}^{h} R_i$ and R_{C} as the transitive closure of R_{E}. A formula being satisfied at a given world is then defined as before with the following crucial case for the formula being of the form $[t]_{\circledast}B$

1) $\mathcal{E}_{\circledast}(t, B, w)$ holds and
2) $\mathcal{M}, v \Vdash B$ for all $v \in W$ with $(w, v) \in R_{\circledast}$.

Using a canonical model construction we can show the soundness and completeness of $\mathsf{LP}_h^{\mathsf{C}}$ with respect to this class of models and as an immediate corollary of this construction (see [9, Theorem 20]) we obtain

Theorem 25. $\mathsf{LP}_h^{\mathsf{C}}(\mathsf{CS})$ *is sound and complete with respect to the class of singleton models for* $\mathsf{LP}_h^{\mathsf{C}}(\mathsf{CS})$*.*

We can easily adapt the Φ-generated submodels from Definition 23 and Lemma 24 to the multi-agent case and turn these singleton models into finitary models. Obviously the class of these finitary, singleton models is recursively enumerable and adapting Lemma 21 to the multi-agent case shows that $\mathsf{LP}_h^{\mathsf{C}}$ satisfies the conditions of Theorem 19. Decidability of $\mathsf{LP}_h^{\mathsf{C}}$ then follows as in the previous section.

Theorem 26. $\mathsf{LP}_h^{\mathsf{C}}(\mathsf{CS})$ *with a decidable schematic* CS *is decidable.*

[1] Note that these closure conditions are very similar and also give rise to a monotone operator as before. This is crucial in adapting the previous proofs to the multi-agent case.

8 Conclusion

We have presented a uniform method of proving decidability for justification logics using a refinement of the finite model property. In order to achieve this property, we have adapted the modal techniques of filtration and generated submodels to justification logics. Apart from reproving the known decidability results for J_{CS}, $\mathsf{JD}_{\mathsf{CS}}$, $\mathsf{JT}_{\mathsf{CS}}$, $\mathsf{J4}_{\mathsf{CS}}$, $\mathsf{JD4}_{\mathsf{CS}}$, and $\mathsf{LP}_{\mathsf{CS}}$, this method has enabled us to establish the decidability of the justification logic with common knowledge introduced in [9].

The main difference from the modal case is the presence of an additional element in models called evidence relation. As evidence relations are in general infinite objects, the filtration has to be performed in such a way that apart from finitizing the set of worlds, also the evidence relation is finitely representable. This finite representation is achieved by using least fixed points of a certain monotone operator that can be read off the axioms of the logic. The existence of the least fixed point is guaranteed when the operator is monotone, which is the case for all the logics considered. Some logics, e.g. justification logics with negative introspection, however, give rise to non-monotone operators. Proving decidability for them requires more involved techniques, see [20].

References

1. Artemov, S.N.: Operational modal logic. Tech. Rep. MSI 95–29, Cornell University (December 1995)
2. Artemov, S.N.: Explicit provability and constructive semantics. Bulletin of Symbolic Logic 7(1), 1–36 (2001)
3. Artemov, S.N.: The logic of justification. The Review of Symbolic Logic 1(4), 477–513 (2008)
4. Artemov, S.N., Fitting, M.: Justification logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Fall 2011 edn. (2011), `http://plato.stanford.edu/archives/fall2011/entries/logic-justification/`
5. Artemov, S.N., Kuznets, R.: Logical omniscience as a computational complexity problem. In: Heifetz, A. (ed.) Proceedings of the Twelfth Conference on Theoretical Aspects of Rationality and Knowledge, TARK 2009, July 6-8, pp. 14–23. ACM, Stanford University (2009)
6. Blackburn, P., Rijke, M., Venema, Y.: Modal logic. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press (2002)
7. Brünnler, K., Goetschi, R., Kuznets, R.: A syntactic realization theorem for justification logics. In: Beklemishev, L.D., Goranko, V., Shehtman, V. (eds.) Advances in Modal Logic, vol. 8, pp. 39–58. College Publications (2010)
8. Bucheli, S., Kuznets, R., Renne, B., Sack, J., Studer, T.: Justified belief change. In: Arrazola, X., Ponte, M. (eds.) LogKCA 2010, Proceedings of the Second ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action, pp. 135–155. University of the Basque Country Press (2010)
9. Bucheli, S., Kuznets, R., Studer, T.: Justifications for common knowledge. Journal of Applied Non-Classical Logics 21(1), 35–60 (2011)

10. Bucheli, S., Kuznets, R., Studer, T.: Partial Realization in Dynamic Justification Logic. In: Beklemishev, L.D., de Queiroz, R. (eds.) WoLLIC 2011. LNCS, vol. 6642, pp. 35–51. Springer, Heidelberg (2011)
11. Fitting, M.: The logic of proofs, semantically. Annals of Pure and Applied Logic 132(1), 1–25 (2005)
12. Kuznets, R.: On the Complexity of Explicit Modal Logics. In: Clote, P.G., Schwichtenberg, H. (eds.) CSL 2000. LNCS, vol. 1862, pp. 371–383. Springer, Heidelberg (2000)
13. Kuznets, R.: Complexity Issues in Justification Logic. Ph.D. thesis, CUNY Graduate Center (May 2008), http://gradworks.umi.com/33/10/3310747.html
14. Kuznets, R.: Self-referential justifications in epistemic logic. Theory of Computing Systems 46(4), 636–661 (2010)
15. Lemmon, E.J., Scott, D.S.: An Introduction to Modal Logic. B. Blackwell, Oxford (1977)
16. Mkrtychev, A.: Models for the Logic of Proofs. In: Adian, S., Nerode, A. (eds.) LFCS 1997. LNCS, vol. 1234, pp. 266–275. Springer, Heidelberg (1997)
17. Pacuit, E.: A note on some explicit modal logics. In: Proceedings of the 5th Panhellenic Logic Symposium, July 25-28, pp. 117–125. University of Athens, Athens (2005)
18. Post, E.L.: Recursively enumerable sets of positive integers and their decision problems. Bull. Am. Math. Soc. 50, 284–316 (1944)
19. Segerberg, K.: An Essay in Classical Modal Logic. Uppsala,Filosofiska Föreningen Och Filosofiska Institutionen Vid Uppsala Universitet (1971)
20. Studer, T.: Decidability for some justification logics with negative introspection. Journal of Symbolic Logic (to appear), http://www.iam.unibe.ch/~tstuder/papers/J5_Decidability.pdf

Interpreted Systems Semantics for Process Algebra with Identity Annotations

Francien Dechesne[1] and Mohammad Reza Mousavi[2]

[1] Philosophy Section, Faculty of Technology, Policy and Management
Delft University of Technology, The Netherlands
[2] Department of Computer Science
Eindhoven University of Technology, The Netherlands

Abstract. Process algebras have been developed as formalisms for specifying the behavioral aspects of protocols. Interpreted systems have been proposed as a semantic model for multi-agent communication. In this paper, we connect these two formalisms by defining an interpreted systems semantics for a generic process algebraic formalism. This allows us to translate and compare the vast body of knowledge and results for each of the two formalisms to the other and perform epistemic reasoning, e.g., using model-checking tools for interpreted systems, on process algebraic specifications. Based on our translation we formulate and prove some results about the interpreted systems generated by process algebraic specifications.

1 Introduction

1.1 Motivation

Process algebras [29,19,3] have evolved in the past three decades into a rich theory of behavioral specification for concurrent and distributed systems. Various process algebras come equipped with rich syntax, rigorous semantics and strong equational and operational reasoning techniques.

Interpreted systems (ISs) [15,30,17,14] have started around the same time as process algebras and have gained popularity as semantic models to include epistemic aspects into multi-agent systems. Since then much research has been devoted into both theory and implementation of interpreted systems.

In this paper, we propose an interpreted systems semantics for a generic process algebra, thereby establishing a link between these two worlds. This link allows one to translate the vast body of knowledge in each of the two realms to the other and benefit from the tools available for both formalisms when dealing with a multi-agent system (e.g., by using model-checkers for interpreted systems for protocols specified in process algebra). Also, algebraic structures of processes and their equational theory can be used to compositionally reason about logical properties, see, e.g., [1].

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 182–205, 2013.

1.2 Related Work

In this paper we aim to show how Interpreted Systems can provide semantics for a generic process algebraic formalism. This paves the way for using model-checking tools based on interpreted systems, e.g., MCMAS [27] and MCK [28], for verifying epistemic properties of process algebraic specifications. This also relates to the line of work in translating operational specification languages to the input languages of the above-mentioned model checkers, see, e.g., [5]. A subsequent goal for the research initiated by this paper is to characterize the class of IS models for different process algebras.

The benefits of combining behavioral (e.g., process algebraic) and epistemic formalisms (e.g., epistemic logic) have been noted by several authors, starting from Halpern and Moses in the seminal [15]. There already exists a rich literature on such combinational frameworks, especially in the application area of security protocols [6,25,7,36,11]. Our work in [9] contributed to this body of knowledge by providing a combinational framework for verifying a rich epistemic temporal language on a process algebraic formalism. Our study of interpreted systems for process algebras is based on the process algebraic framework proposed in [9]. This framework bears relation to the epistemic systems of [34] by the concept of an *appearance map* (renaming function in our framework). Recently, the process-algebraic framework of [9] has been extended in [25] to support probabilistic constructs. In [6], an epistemic temporal logic for the applied pi-calculus is presented (although [6] only allows for single-agent knowledge).

We mention [26], which aims at axiomatizations for interesting classes of ISs (notably *hypercubes*), through a characterization of the epistemic dimension of those ISs as a subclass of S5 Kripke models. This work however disregards the temporal dimension of the ISs.

The intention of our paper relates to the work of Van Benthem et al. on exploring the interface between the Dynamic Epistemic- (DEL) and Epistemic Temporal- (ETL) frameworks [4,21]. Their line of work compares and merges the two, and characterizes the classes of models for ETL generated from DEL models.

There have been several attempts to define a knowledge-based semantics for programming languages [24,20,23,12,37,13]. The closest to our work are [20,23], where a knowledge-based semantics is given to a CSP-like process algebra with local states and assignments. The fundamental difference between the approach of [20,23] and that of the present paper is that there, each agent is supposed to be represented by a sequential process, while in our approach agents may have different observations and perceptions of process algebraic actions and they need not be (although can be) incarnated in a particular process. In other words, in our approach there needs not be a one-to-one mapping between agents and processes.

1.3 Structure of the Paper

In Section 2, we present our generic process algebraic formalism called *CCSi*. The basic definitions of interpreted systems are recalled in Section 3. Then, the semantics of *CCSi* in terms of an interpreted system is presented in 4. Some formal results about the semantic framework are presented in Section 5 and the paper is concluded in Section 6.

2 *CCSi*: A Process Algebraic Formalism

2.1 Syntax

The basic building blocks of a behavioral specification in a process algebra are *atomic actions*. They represent sending, receiving or communicating (synchronizing on) messages as well as internal computations that may or may not be visible to the observers. Atomic actions are composed using various composition operators, leading to myriad process algebras. Here, we confine ourselves to a simple process algebra, inspired by the well-known Calculus of Communicating Systems (CCS) [29]. The formalism studied in this paper (and also in the earlier work) is called *CCSi*, for CCS with identities. We slightly extend our earlier definition of *CCSi* with termination constant ϵ and unbounded choice to allow for infinite branching. We have intentionally chosen for a process algebra with few composition operators to illustrate the ideas. The constructions presented here can be extended in a straight-forward manner to various other process algebras such as those introduced in [19,3].

Let $\mathcal{A}ct$ be a finite set of *action names* $(a, b, a_0, \ldots$ and $!a, ?a, \ldots)$, and let $\mathcal{I}d$ be a finite set of *identities* (of the participating principals or agents) typically denoted by $i, j, \ldots i_1, i_2, \ldots$. Action letters preceded by a question mark or exclamation mark $?a$ and $!a$ represent the receiving and the sending parts of a communication, respectively, which result through synchronization in the communication a.[1] We let Greek letters $\alpha, \beta, \ldots$ range over the complete set of actions $\mathcal{A}ct$ (including the sending- and receiving parts), while letters $a, b, \ldots$ only range over actions without question- and exclamation marks.

Processes in *CCSi* are specified using *decorated actions* $d \in D ::= \{(\mathtt{J})\alpha | \mathtt{J} \subseteq \mathcal{I}d, \alpha \in \mathcal{A}ct\}$, and a global renaming function $\rho : \mathcal{A}ct \to \mathcal{A}ct \cup \{\tau\}$. The intuitive meaning of a decorated action $(\mathtt{J})\alpha$ is that action α is taken visibly to the principals in $\mathtt{J}$ (the intended audience of this α). Principals not in $\mathtt{J}$ will observe the so-called *public appearance* $\rho(\alpha)$ of α. In the signature of ρ, τ denotes the "silent appearance" of an action; it is assumed that $\tau \notin \mathcal{A}ct$ and for any other action α, if $\rho(\alpha) = \tau$, then $(\mathtt{J})\alpha$ becomes unobservable to the principals not in $\mathtt{J}$. We abbreviate $(\mathcal{I}\mathtt{d})\alpha$, i.e., an action visible to everyone, by α. *CCSi*-Processes are then specified as follows, together with a renaming function ρ:

$$Proc \ ::= \ \epsilon \mid D \mid Proc\,; Proc \mid Proc \,||\, Proc \mid \Sigma_{i \in I} Proc_i$$

[1] Here we take a variation on standard CCS, where successful synchronization of a send- and receive action results in a silent action τ. We take the synchronization a in our framework to be the successful communication of a message (which is not a silent action).

where termination process ϵ cannot make any transition, but terminates, $Proc\,;\, Proc$ denotes sequential composition, $Proc \,||\, Proc$ denotes parallel composition, and $\Sigma_{i \in I} Proc_i$ denotes nondeterministic choice among processes $Proc_i$, for each i in the non-empty (and possibly infinite) index set I. We will write $p_1 + \ldots + p_m$ to denote the finite nondeterministic choice $\Sigma_{\{1,\ldots,m\}} p_i$.)

The combination of identity annotations on actions and the action renaming provides different views on the behavior of the system, according to different principals.[2] Modeling passive observation of a system by hiding parts of it to specific principals is already done in the literature (e.g., in [35]), but we will generate the views for all principals simultaneously. This enables talking about properties such as "i knows that j knows that k has communicated message a".

Note that in this formalism there is no direct correspondence between the processes ($Proc$) and agents ($\mathcal{I}d$). We are not so much interested in how the principals behave (what their actions are), but we are interested in what they get to know from what happens. This is in line with the approach taken in the seminal [15], where agents are process*ors*, not process*es* (their framework focuses on knowledge and does not specify the behavior of the system explicitly). But it differs with some earlier work which uses process algebra for the specification of multi-agent systems such as [20,23,12,37,13], where agents *are* modeled as processes. In our approach, an agent may take part in several processes and a process may comprise actions that are visible (communicated by/to) many different principals. This is useful in the behavioral specification of protocols, where an agent may participate in different threads of communications and a single thread may be involved in several synchronisations with different agents.

Example 1. **(Toy Example: Syntax)** Let $\mathcal{A}ct = \{a, a_0, b, c\}$. We elaborate the definitions throughout the paper for the following simple *CCSi*-process p and renaming function ρ.

$$p \doteq ((\mathtt{1})?a \,||\, (\mathtt{2})!a) + ((\mathtt{3})b\,;c)$$

$$\rho(a) = \rho(a_0) \doteq a_0, \rho(b) \doteq \tau, \rho(c) \doteq c$$

Process p features a non-deterministic choice between the following two options:

1. synchronizing on action a; the result of synchronization is directly visible to principals 1 and 2, while the rest precieve this as action a_0, or
2. performing an action b followed by c. Action b is only visible to principal 3, and the rest of the principals do not even notice that an action has taken place. Action c is visible to all principals. Note that as defined earlier, action c abbreviates the decorated action $(\mathcal{I}\mathtt{d})c$ (everyone sees c as it happens).

Example 2. **(Dining Cryptographers: Syntax)** In this example, we give a formal specification of the Dining Cryptographers protocol [8], which has been

[2] We will see that the send- ($!a$) and receive ($?a$) parts of an action will not be explicit in the semantics, but only the result a of their successful communication will be. This means $\rho(?a)$ and $\rho(!a)$ can be defined arbitrarily, or be left undefined.

$$
\begin{array}{lcl}
Crypt(\mathrm{i}) & = & \sum_{b:Bool} \; (\; (\mathrm{i})?pay(\mathrm{i}, b); CryptFlip(\mathrm{i}, b) \;) \\
CryptFlip(\mathrm{i}, b) & = & \sum_{c:Bool} \; (\; (\mathrm{i})flip(\mathrm{i}, c); CryptShare(\mathrm{i}, b, c) \;) \\
CryptShare(\mathrm{i}, b, c) & = & \sum_{d:Bool} \; (\; ((\mathrm{i})!share((\mathrm{i} \mod 2) + 1, c) \,||\, (\mathrm{i})?share(\mathrm{i}, d)) \,; \\
 & & \qquad CryptBcast(\mathrm{i}, b, c, d) \;) \\
CryptBcast(\mathrm{i}, b, c, d) & = & ((\mathrm{i})!bcast(\mathrm{i}, b \oplus c \oplus d) \\
 & & || \sum_{x:Bool}((\mathrm{i})?bcast((\mathrm{i} \mod 2) + 1, x))) \,; \\
 & & \qquad paid(\mathrm{i}, b \oplus c \oplus d \oplus x) \\
Master & = & \quad (\mathtt{M})!pay(1, \top); (\mathtt{M})!pay(2, \bot) \\
 & & + \; (\mathtt{M})!pay(1, \bot); (\mathtt{M})!pay(2, \top) \\
 & & + \; (\mathtt{M})!pay(1, \bot); (\mathtt{M})!pay(2, \bot)
\end{array}
$$

Fig. 1. A *CCSi* model of the Dining Cryptographers protocol for 2 cryptographers

extensively studied in the literature (e.g., in [35,2,22,16,33]). For reasons of presentational simplicity, we give a version with two cryptographers and an external observer.

In general, the scene is about a number of cryptographers (2 in our case) having dinner together. At the end, they learn that the bill has been paid by one of them, or by their master. They do not want to compromise each other's right to anonymity, but they wish to make it known to the public whether the payer was the master or not. (The usual presentation of the setting includes at least three cryptographers, in which case the paying cryptographer -if any- will remain anonymous not just to the public, but to the other cryptographers as well.) To this end, they come up with the following protocol: each neighboring cryptographer generates a shared bit, by flipping a coin; then each cryptographer computes the exclusive or (XOR) of the bits she sees (one in our case) with her own bit, and announces the result — or the flipped result, if she was herself the payer. The XOR of the publicly announced results indicates whether the payer was an insider or the master.

We specify the protocol for an external observer (O), two cryptographers (1 and 2), and the master (M). The observer is assumed to perfectly know the protocol; it tries to comprise the anonymity of the cryptographers and learn about the identity of the payer by looking at the trace of the protocol which has taken place, and comparing it with the possible traces with different payers.

A model of this protocol in our process language is shown in Figure 1.

The model is adopted and adapted from our earlier publication [9] and is close to the CSP description presented in [35], the only significant difference being that the actions are annotated with identities from the set $\mathcal{I}d = \{\mathrm{O}, 1, 2, \mathrm{M}\}$. Note that we use parameters in the basic actions and process definitions only to provide a notational shorthand for the concrete actions and processes resulting from instantiating them. For example, $?pay(\mathrm{i}, b)$ is not defined in our process language but rather it stands for a number of instances such as $?pay(1, \top)$, $?pay(2, \bot)$ each of which are basic actions (obtained by globally replacing i and b with a member of $\{1, 2\}$and $\{\bot, \top\}$ in the process definition each time). In the description of the protocol $\oplus$ denotes exclusive or. Also the process names are syntactic sugar for the processes they define. The behavior of the ith cryptographer is specified by the process *Crypt*(i) and the behavior of the whole DC system as a parallel composition of *Crypt*(i)'s and the *Master* process:

$$DC_2 = Crypt(1) \,||\, Crypt(2) \,||\, Master$$

Note that the observer principal is not mentioned anywhere in the specification and will only be represented in the semantic model of the protocol.

A cryptographer process executes a series of actions corresponding to the three big steps of the protocol: decide whether to pay or not, flip the coins together with the neighbors, and announce the result of XOR-ing the two coins and her own paying bit. The first step is modeled as a statement $pay(\mathrm{i}, b)$, which is in fact a communication step with the *Master*.

The second step is modeled by the processes *CryptFlip*(i) and *CryptShare*(i). Process *Crypt*(i) executes a *flip* action and then shares the result with the neighbor, by executing an action $!share$ which will synchronize with the $?share$ from the neighboring cryptographer. *CryptBcast* models the last phase, announcing the result of one's computation ($!bcast$), receiving the results from all the others ($?bcast$) and concluding for itself that a cryptographer has paid or not ($paid(\mathrm{i}, \top)$, or $paid(\mathrm{i}, \bot)$, respectively).

The renaming function ρ specifies how much of a cryptographers' actions is visible for observing parties. For any $\mathrm{i} \in \{1, 2\}$ and $b \in \{\top, \bot\}$, we define

$$\begin{array}{lll} \rho(pay(\mathrm{i}, b)) \;= pay(\mathrm{i}) & \rho(bcast(\mathrm{i}, b)) = bcast(\mathrm{i}, b) & \rho(share(\mathrm{i}, c)) = share(\mathrm{i}) \\ \rho(flip(\mathrm{i}, b)) = flip(\mathrm{i}) & \rho(paid(\mathrm{i}, b)) \;= paid(\mathrm{i}, b) & \end{array}$$

where $pay(1)$, $bcast(1, \top)$, ... are basic actions.

2.2 Transition Systems Semantics

The operational semantics of *CCSi* (from [9]) is given in Figure 2 in terms of Structural Operational Semantics (SOS) rules [32]. The operational state of *CCSi* is a pair (p, π), where p is a process in the syntax given in Section 2.1 and π is a sequence of decorated actions (a *history*).

We include the history in the operational state of our semantics in order to capture the epistemic aspect. Such a sequence of decorated actions together with the renaming function allow us to construct in each state how each principal perceives what has happened so far. This allows us to evaluate epistemic statements on the semantics. (Note that just process terms as states only would only code the possible future actions, and contain no information about the past - let alone code individual perceptions of the past; the semantics for process $a; p + b; p$ would, for example, after initial branching meet in a single state coded with p.) Using histories and principals' perception, we recover the notions of knowledge and knowledge update in the operational semantics of protocols. If a particular proposition is true in all operational states that are perceived the same for a particular principal, then the principal knows the proposition. The knowledge of principals is updated by appending new perceived actions to the histories.

The operational semantics of a process p is its associated labeled transitions system (with (p, ϵ) as the starting state, where ϵ denotes the empty history of decorated actions) defined by the deduction rules of Figure 2. The transitions in this LTS are of the form $\stackrel{a}{\rightarrow}$, which is, in turn, defined in terms of the auxiliary transition relation $\stackrel{(J)a}{\Rightarrow}$, by stripping off the decorations and blocking non-synchronized sends and receives.

Process ϵ cannot make any transition, but terminates immediately; this is denoted by the termination predicate $\surd$ in the deduction rule (axiom) (ϵ). The semantics of a decorated action is defined by the deduction rule (**a**): the process d can perform the action d (which is concatenated to the history) and then turn into the terminated process ϵ. The operational behavior of nondeterministic choice is defined by the choice in the behavior of its arguments. This is captured by the deduction rule scheme (**ni**) (for each index set I, $i \in I$). Transition semantics of sequential composition is defined by either taking an action from the first component, or termination of the first component followed by an action from the second one, as specified by deduction rules (**s0**) and (**s1**) respectively. The semantics of parallel composition is defined by the interleaving (deduction rules (**p0**)-(**p1**)) and the synchronization (deduction rules (**p2**)-(**p3**)) of the actions of its arguments (we here omit deduction rules (**p1**) and (**p3**), which are, respectively, symmetric copies of (**p0**) and (**p2**)). Termination of nondeterministic choice, sequential composition and parallel composition is specified, respectively, by (**nti**), (**st**), and (**pt**).

Deduction rule (**strip**) strips down the decorated action into plain actions (by removing the intended audience) and ignores send- and receive actions (hence, one could say it enforces synchronization among communicating processes in the trace semantics).

In addition to the operational semantics, we define an epistemic semantics for the process calculus using the indistinguishability relation $\stackrel{i}{\cdots}$, which is defined in terms of the indistinguishability relation $\stackrel{i}{=}$ (see also [9]). Considering the deduction rules for $\stackrel{i}{=}$, reflexivity is captured in deduction rule (**refl**); rule ($= \rho\mathbf{0}$) defines the case for a visible action to principle i; rule ($= \rho\mathbf{1}$) concerns when two invisible actions which have the same public appearance for i; rule ($= \rho\mathbf{2}$) defines the case for an action which is invisible to i but appears to i as another visible action; rule ($= \tau\mathbf{0}$) is about an invisible action which appears as τ to i (i.e., is absolutely unobservable for i). (Again for the sake of brevity, we have omitted symmetric rules for ($= \rho\mathbf{2}$) and ($= \tau\mathbf{0}$).) Deduction rule (**I**) lifts the indistinguishability relation $\stackrel{i}{=}$ from sequences of decorated actions to the

$$(\epsilon)\,\frac{}{(\epsilon,\pi)\surd} \qquad (\mathbf{a})\,\frac{}{(d,\pi)\stackrel{d}{\Rightarrow}(\epsilon,\pi\frown d)}$$

$$(\mathbf{ni})\,\frac{(x_i,\pi)\stackrel{d}{\Rightarrow}(y,\pi')}{(\Sigma_{i\in I}x_i,\pi)\stackrel{d}{\Rightarrow}(y,\pi')}\,i\in I \qquad (\mathbf{nti})\,\frac{(x_i,\pi)\surd}{(\Sigma_{i\in I}x_i,\pi)\surd}\,i\in I$$

$$(\mathbf{s0})\,\frac{(x_0,\pi)\stackrel{d}{\Rightarrow}(y_0,\pi')}{(x_0\,;x_1,\pi)\stackrel{d}{\Rightarrow}(y_0\,;x_1,\pi')} \qquad (\mathbf{s1})\,\frac{(x_0,\pi)\surd \quad (x_1,\pi)\stackrel{d}{\Rightarrow}(y_0,\pi')}{(x_0\,;x_1,\pi)\stackrel{d}{\Rightarrow}(y_0,\pi')}$$

$$(\mathbf{st})\,\frac{(x_0,\pi)\surd \quad (x_1,\pi)\surd}{(x_0\,;x_1,\pi)\surd} \qquad (\mathbf{p0})\,\frac{(x_0,\pi)\stackrel{d}{\Rightarrow}(y_0,\pi')}{(x_0\,||\,x_1,\pi)\stackrel{d}{\Rightarrow}(y_0\,||\,x_1,\pi')}$$

$$(\mathbf{pt})\,\frac{(x_0,\pi)\surd \quad (x_1,\pi)\surd}{(x_0\,||\,x_1,\pi)\surd} \qquad (\mathbf{p2})\,\frac{(x_0,\pi)\stackrel{(\mathtt{J})?a}{\Rightarrow}(y_0,\pi') \quad (x_1,\pi)\stackrel{(\mathtt{J}')!a}{\Rightarrow}(y_1,\pi'')}{(x_0\,||\,x_1,\pi)\stackrel{(\mathtt{J}\cup\mathtt{J}')a}{\Rightarrow}(y_0\,||\,y_1,\pi\frown(\mathtt{J}\cup\mathtt{J}')a)}$$

$$(\mathbf{strip})\,\frac{(x,\pi)\stackrel{(\mathtt{J})a}{\Rightarrow}(y,\pi')}{(x,\pi)\stackrel{a}{\rightarrow}(y,\pi')}$$

$$(=\mathbf{refl})\,\frac{}{\pi\stackrel{i}{\doteq}\pi} \qquad (=\rho\mathbf{0})\,\frac{\pi\stackrel{i}{\doteq}\pi' \quad a=b \quad i\in\mathtt{J}\cap\mathtt{J}'}{\pi\frown(\mathtt{J})a\stackrel{i}{\doteq}\pi'\frown(\mathtt{J}')b}$$

$$(=\rho\mathbf{1})\,\frac{\pi\stackrel{i}{\doteq}\pi' \quad \rho(a)=\rho(b) \quad i\notin\mathtt{J}'\cup\mathtt{J}}{\pi\frown(\mathtt{J})a\stackrel{i}{\doteq}\pi'\frown(\mathtt{J}')b} \qquad (=\rho\mathbf{2})\,\frac{\pi\stackrel{i}{\doteq}\pi' \quad a=\rho(b) \quad i\in\mathtt{J}\setminus\mathtt{J}'}{\pi\frown(\mathtt{J})a\stackrel{i}{\doteq}\pi'\frown(\mathtt{J}')b}$$

$$(=\tau\mathbf{0})\,\frac{\pi\stackrel{i}{\doteq}\pi' \quad i\notin\mathtt{J} \quad \rho(a)=\tau}{\pi\frown(\mathtt{J})a\stackrel{i}{\doteq}\pi'} \qquad (\mathbf{I})\,\frac{\pi_0\stackrel{i}{\doteq}\pi_1}{(x_0,\pi_0)\stackrel{i}{\cdots}(x_1,\pi_1)}$$

Fig. 2. SOS of *CCSi* (cf. [9])

indistinguishability relation $\stackrel{i}{\cdots}$ on the operational state of *CCSi*. As shown in [9], both $\stackrel{i}{\doteq}$ and $\stackrel{i}{\cdots}$ are equivalence relations.

This semantics is introduced here to present the original semantics given in [9] and to compare it with the interpreted systems semantics presented in the subsequent sections. For these and the following definitions, we now provide a running example for clarification.

Example 3. **(Toy Example: Semantics)** Consider process p specified in Example 1. The traces of p can be generated by the SOS-rules as follows:

1. $((1)?a,\langle\rangle)\stackrel{(1)?a}{\Rightarrow}(\epsilon,\langle(1)?a\rangle)$ – rule (**a**)
2. $((2)!a,\langle\rangle)\stackrel{(2)!a}{\Rightarrow}(\epsilon,\langle(2)!a\rangle)$ – rule (**a**)
3. $((1)?a\,||\,(2)!a,\langle\rangle)\stackrel{(1,2)a}{\Rightarrow}(\epsilon\,||\,\epsilon,\langle(1,2)a\rangle)$ – rule (**p2**), using 1,2
4. $(\epsilon\,||\,\epsilon,\langle(1,2)a\rangle)\surd$ – rules (ϵ), (**pt**)
5. $((3)b,\langle\rangle)\stackrel{(3)b}{\Rightarrow}(\epsilon,\langle(3)b\rangle)$ – rule (**a**)
6. $((3)b\,;c,\langle\rangle)\stackrel{(3)b}{\Rightarrow}(\epsilon\,;c,\langle(3)b\rangle)$ – rule (**s0**), using 5
7. $(c,\langle(3)b\rangle)\stackrel{c}{\Rightarrow}(\epsilon,\langle(3)b,c\rangle)$ – rule (**a**)

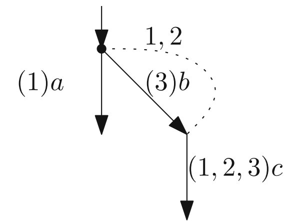


Fig. 3. Operational Semantics of the Toy Example

8. $(\epsilon, \langle(3)b, c\rangle)\surd$ – rule (ϵ)
9. $(\epsilon\,; c, \langle(3)b\rangle) \overset{c}{\Rightarrow} (\epsilon, \langle(3)b, c\rangle)$ – rule (**s1**), using 7,8
10. $((1)?a \,||\, (2)!a, \langle\rangle) \overset{(1)?a}{\Rightarrow} (\epsilon \,||\, (2)!a, \langle(1)?a\rangle) \overset{(2)!a}{\Rightarrow} (\epsilon \,||(\epsilon, \langle(1)?a, (2)!a\rangle)$ – rules (**p0**), (**p1**), using 1,2
11. $((1)?a \,||\, (2)!a, \langle\rangle) \overset{(2)!a}{\Rightarrow} ((1)?a \,||\, \epsilon, \langle(2)!a\rangle) \overset{(1)?a}{\Rightarrow} (\epsilon \,||\, \epsilon, \langle(2)!a, (1)?a\rangle)$ – rule (**p1**), (**p0**), using 2,1
12. $(p, \langle\rangle) \overset{(1,2)a}{\Rightarrow} (\epsilon \,||\, \epsilon, \langle(1, 2)a\rangle)$ – rule (**ni**), using 3
13. $(p, \langle\rangle) \overset{(1)?a}{\Rightarrow} (\epsilon \,||\, (2)!a, \langle(1)?a\rangle) \overset{(2)!a}{\Rightarrow} (\epsilon \,||\, \epsilon, \langle(1)?a, (2)!a\rangle)$ – rule (**ni**) using 10
14. $(p, \langle\rangle) \overset{(2)!a}{\Rightarrow} ((1)?a \,||\, \epsilon, \langle(2)!a\rangle) \overset{(1)?a}{\Rightarrow} (\epsilon \,||\, \epsilon, \langle(2)!a, (1)?a\rangle)$ – rule (**ni**) using 11
15. $(p, \langle\rangle) \overset{(3)b}{\Rightarrow} (\epsilon\,; c, \langle(3)b\rangle) \overset{c}{\Rightarrow} (\epsilon, \langle(3)b, c\rangle)$ – rule (**n1**) using 6, followed by 9
16. $(p, \langle\rangle) \overset{a}{\rightarrow} (\epsilon \,||\, \epsilon, \langle(1, 2)a\rangle)$ – rule (**strip**), using 12
17. $(p, \langle\rangle) \overset{b}{\rightarrow} (\epsilon\,; c, (3)b) \overset{c}{\rightarrow} (\epsilon, \langle(3)b, c\rangle)$ – rule (**strip**) using 15

Note again that rule (**strip**) is restricted to 'closed' actions, i.e., excluding $!a, ?a$, so it does not apply to lines 12 and 13. The set of traces of p is therefore (lines 15 and 16): $\{\langle a\rangle, \langle b, c\rangle\}$.

Now, we also generate the indistinguishability relation through the SOS-rules:

1. For all $i \in \mathcal{I}d$ and for all π: $\pi \overset{i}{=} \pi$ – rule (= **refl**)
2. $\langle(1, 2)a\rangle \overset{3}{=} \langle a_0\rangle$ – rule (= ρ**2**)
3. $\langle(3)b, c\rangle \overset{1}{=} \langle c\rangle$ – rule (= τ**0**)
4. $\langle(3)b, c\rangle \overset{2}{=} \langle c\rangle$ – rule (= τ**0**)

The state space of this example is depicted in Figure 3. In this figure, the initial state is designated with a small incoming arrow. The transitions derived from the operational semantics are drawn as solid arrows and the indistinguishability relation is drawn as a dotted line labeled with the principal identities. (In order not to clutter the figure, we dispensed with the self-loops denoting the reflexivity of the indistinguishability relation.)

Example 4. **(Dining Cryptographers: Semantics)** Consider the specification of dining cryptographers given in Example 2. The complete state space of the

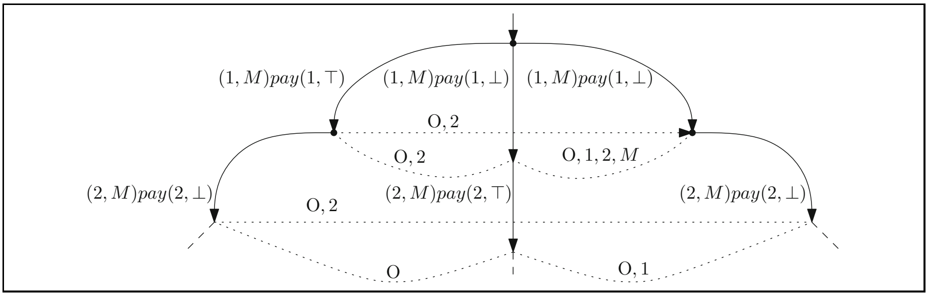


Fig. 4. Operational Semantics of the Dining Cryptographers Protocol

protocol is too large to be studied manually (see [10] for a prototype implementation of a model-checker for a process algebra, using which we performed a mechanized analysis of this protocol). The initial steps of the protocol are depicted in Figure 4. Consider, for example, the leftmost and the middle traces in Figure 4. After the first step of the protocol, principals 1 and 2 observe action $pay(1, \top)$ in the leftmost trace and action $pay(1, \bot)$ in the middle trace and hence, can distinguish the target states of these two actions. Principals O and 2, however, observe $pay(1)$ in both cases and hence the resulting states are indistinguishable to them. After the second step, principal 2 can also distinguish between the two traces, because it can observe $pay(2, \bot)$ as the second action of the leftmost trace, while it can observe $pay(2, \top)$ as the second action of the middle trace. Principal O still cannot distinguish the two traces because the second action appears in both cases as $pay(2)$ to it. Note that modeling this aspect of knowledge about the actions that have taken place and revisions in the knowledge is made possible thanks to the notion of history (of past actions) that is included in the operational state.

Below, we give three completed traces of the protocol, which are continuations of the three initial branches depicted in Figure 4. The protocol starts with the $Master$ synchronising on pay-actions with each cryptographer. These traces do show the essence of the protocol (i.e., the three possible cases for payment) with some choice of coin flips. Our choices for coin flips may seem arbitrary at the first glance, but actually, a particular choice is made to demonstrate how the protocol guarantees anonymity:

1. The first trace, given below, is a continuation of the leftmost trace, in which the first cryptographer has paid; the result of both coin flips in this particular trace is a head ($\top$). For the sake of brevity, in the description of the traces, we only mention the histories and the transitions and omit the process expressions:

$$(DC_2, \langle\rangle) \stackrel{pay(1,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top)\rangle) \stackrel{pay(2,\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot)\rangle) \stackrel{flip(1,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top)\rangle) \stackrel{flip(2,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top)\rangle) \stackrel{share(1,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top)\rangle) \stackrel{share(2,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top)\rangle) \stackrel{bcast(1,\top\oplus\top\oplus\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\top)\rangle) \stackrel{bcast(2,\bot\oplus\top\oplus\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\top), ((\mathtt{1},\mathtt{2}))bcast(2,\bot)\rangle) \stackrel{paid(1,\top\oplus\top\oplus\top\oplus\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\top), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\top)\rangle) \stackrel{paid(2,\bot\oplus\top\oplus\top\oplus\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\top), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\top), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\top), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(2,\top)\rangle) \quad \surd$$

As it can be seen, upon termination, the history indicates that both principals 1 and 2 have announced that a cryptographer has paid and this announcement can be observed by each and every principal.

2. The second trace, given below, is a continuation of the trace in the middle of Figure 4, in which the second cryptographer has paid, the result of the coin flip by crytographer 1 is a head ($\top$) and that of the cryptographer 2 is a tail ($\bot$):

$$(DC_2, \langle\rangle) \quad \overset{pay(1,\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot)\rangle) \quad \overset{pay(2,\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot)\rangle) \quad \overset{flip(1,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top)\rangle) \quad \overset{flip(2,\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot)\rangle) \quad \overset{share(1,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top)\rangle) \quad \overset{share(2,\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top)\rangle) \quad \overset{bcast(1,\bot\oplus\top\oplus\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\top)\rangle) \quad \overset{bcast(2,\top\oplus\top\oplus\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot)\rangle) \quad \overset{paid(1,\bot\oplus\top\oplus\top\oplus\top)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\top)\rangle) \quad \overset{paid(2,\top\oplus\top\oplus\top\oplus\bot)}{\rightarrow}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\bot), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\top), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(2,\top)\rangle) \quad \surd$$

Similar to the previous trace, at the end of the protocol, it has been announced, by both cryptographers, that a cryptographer has paid the bill. The observer, however, cannot distinguish this trace from the first one, and hence, at the end of either of the two traces, cannot establish which cryptographer has paid. Note that the two cryptographers do know (both in the first and the second trace) who has paid the bill: for example, at the end of the above-given trace, cryptographer 2 knows that it has paid the bill, because this trace is distinguishable (by observing $paid(2,\top)$) from any other trace in which it has not paid the bill. Cryptographer 1 also knows that 2 has paid the bill, because it knows after the first step that it has not paid the bill, and after the last step knows that a cryptographer, hence 2, has paid the bill; in other words, this trace is distinguishable from other traces in which 2 has not paid in the observable *pay* and *paid* actions.

3. The last trace is a continuation of the rightmost trace in Figure 4, in which the master has taken care of the bill and no cryptographer has paid. Both coin flips in this trace result in a head ($\top$):

$$(DC_2, \langle\rangle) \quad \xrightarrow{pay(1,\bot)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot)\rangle) \quad \xrightarrow{pay(2,\bot)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot)\rangle) \quad \xrightarrow{flip(1,\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top)\rangle) \quad \xrightarrow{flip(2,\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top)\rangle) \quad \xrightarrow{share(1,\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top)\rangle) \quad \xrightarrow{share(2,\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top)\rangle) \quad \xrightarrow{bcast(1,\bot\oplus\top\oplus\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot)\rangle) \quad \xrightarrow{bcast(2,\bot\oplus\top\oplus\top)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot)\rangle) \quad \xrightarrow{paid(1,\bot\oplus\top\oplus\top\oplus\bot)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\bot)\rangle) \quad \xrightarrow{paid(2,\bot\oplus\top\oplus\top\oplus\bot)}$$

$$(-, \langle(\mathtt{1},\mathtt{M})pay(1,\bot), (\mathtt{2},\mathtt{M})pay(2,\bot), (\mathtt{1})flip(1,\top), (\mathtt{2})flip(2,\top), ((\mathtt{1},\mathtt{2}))share(1,\top), ((\mathtt{1},\mathtt{2}))share(2,\top), ((\mathtt{1},\mathtt{2}))bcast(1,\bot), ((\mathtt{1},\mathtt{2}))bcast(2,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(1,\bot), ((\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{M}))paid(2,\bot)\rangle)$$

After observing this trace, all principals know that the master has taken care of the bill, because they all observe both $paid(2,\bot)$ and $paid(2,\bot)$ and in all traces that are indistinguishable from the present trace (i.e., contain the same observable *paid* actions in the end) no cryptographer has paid. The latter claim can be checked formally, using an exhaustive search of the state space of the procotol; we refer to [10] for the details.

3 Interpreted Systems

In this section, we recall from [17] formal definitions, terminology and notational conventions regarding interpreted systems. Note however that we deviate from the original definition of [17], by restricting to finite runs. We do this in the context of our process language, because our process terms only afford finite behavior. In order to have infinite runs corresponding to our porcesses, we could have every run have an infinite 'stuttering' tail (as is suggested in [15] as well). We will discuss in Section 5 why we have chosen not to do so for the context of this paper.

Definition 1 (Interpreted Systems (Finite Depth)). *Given a set of $n > 0$ agents with identifiers in $\mathcal{I}d = \{1, \ldots, n\}$, and for each agent $i \in \mathcal{I}d$ a set of local states L_i, a global state $\overline{l}$ is an n-tuple $(l_1, \ldots, l_n)$ with $l_i \in L_i$. Let $L = \prod_{i=1}^{n} L_i$ denote the set of global states. A run r is a finite sequence of global states $r(0), r(1), \ldots, r(m)$ for some $m \in \mathbb{N}$. A protocol R is a non-empty set of runs.*

Given a set Φ of atomic logical formulae, a valuation is a function $\nu : L \to \Phi$.

An interpreted system is then a pair (R, ν), where R is a protocol and ν is a valuation.

Two global states are taken to be indistinguishable for an agent if their local states are equal. This defines an equivalence relation for the evaluation of epistemic formulas:

Definition 2 (Indistinguishability Relations in ISs [17]). *Given an interpreted system with set of global states L, for each agent $i \in \mathcal{I}d$ the relation $\overset{i}{\approx} \subseteq L \times L$ is defined by: $\overline{l} \overset{i}{\approx} \overline{l'}$ iff $l_i = l'_i$.*

The valuation part ν of an Interpreted System (which is taken to be defined on the full L) can be specified independently of the protocol part R. As it turns out, linking our framework and Interpreted Systems is essentially about linking our operational semantics to the protocol part of ISs and linking the respective indistinguishability relations. Within this paper we do not yet explore with which logical language, including a meaningful choice for the atoms, it is best to talk about our epistemic-operational models for processes. This will be part of our future work, and for now, we therefore do not specify the ν-part, only the protocol part of our ISs.

4 Interpreted Systems Semantics for *CCSi*

In this section, we define an interpreted systems semantics for the process algebra *CCSi*. We do so by defining the influence of each operational step on the local state of each principal and then aggregating these influences into the definition of a run. The development of this section is only dependent on the definition of an operational semantics, as defined , and hence the same schema can be

$$(\mathbf{is\_}\surd)\,\frac{(x,\pi)\surd}{x_{\surd}}$$

$$(\mathbf{is\_d})\,\frac{(x,\pi)\overset{(\mathtt{J})a}{\Rightarrow}(y,\pi')}{x\overset{[\![(\mathtt{J})a]\!]}{\rightsquigarrow}y}$$

Fig. 5. Influence of a decorated action on the local states

used for any other process algebra (process algebraic formalism) as long as the visibility range (the set of principals to which the action is visible) and the public appearance of each action is provided in the operational semantics.

Before we define the interpreted systems semantics for a *CCSi* process, we first describe what we will call the local and global state of such a semantics: For a process $p \in CCSi$ with renaming function ρ, a local state $l_i \in L_i \subseteq \mathcal{A}ct^*$ is a sequence of actions in p as they appear to agent $i \in \mathcal{I}d$ under ρ. The set of global states L is defined to be $L = \prod_{i=1}^{n} L_i$. The protocol corresponding to p is the set of all runs of p, which are the sequences of global states corresponding to the traces of p. Note that unlike the histories of *CCSi*, which are sequences of *decorated* actions, the local states of *CCSi* are sequences of *actions* (without any decoration).

Definition 3 (Concatenation of Actions-Tuples to Global States). *Consider a global state* $\overline{l} = (l_1, \ldots, l_n)$ *and an n-tuple* $\overline{a} = (a_1, \ldots, a_n)$ *of actions in* $\mathcal{A}ct \cup \{\tau\}$. *Then* $\overline{l}^\frown \overline{a} = (l'_1, \ldots, l'_n)$ *where* $l'_i = l_i$ *if* $a_i = \tau$ *and* $l'_i = l_i \cdot a_i$ *otherwise.*

Definition 4 (Decorated Action Tuple). *For any decorated action* $d = (\mathtt{J})a$ *we define* $[\![d]\!]$ *to be the n-tuple* $(a_1, \ldots, a_n)$ *with* $a_i = a$ *for* $i \in \mathtt{J}$, *and* $a_i = \rho(a)$ *for* $i \notin \mathtt{J}$.

We are now ready to define the interpreted system semantics of processes by defining their associated protocols. Definition 5 defines the protocol associated with the process ϵ to be the singleton set comprising empty local sequences. The protocol associated with a decorated action, is defined by the tuple of local appearances of each action to each agent. Finally, the notion of protocol is lifted in the expected way from decorated actions to processes.

Definition 5 (Interpreted System Protocols for *CCSi*). *The influence of a decorated action (or termination) on the local state of each principal is defined in Figure 5.*

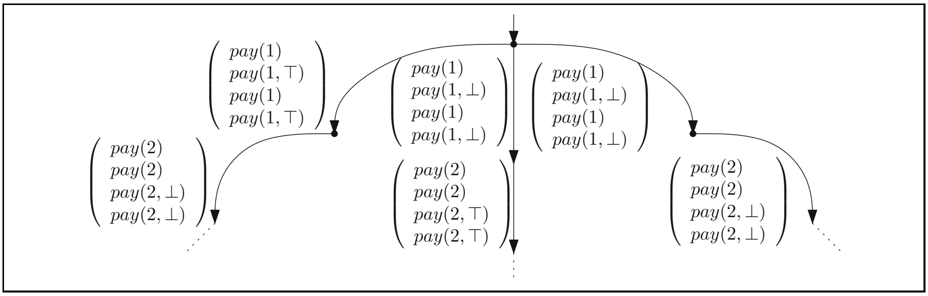


Fig. 6. Interpreted Systems Semantics of the Dining Cryptographers Protocol

We write $p_{\downarrow}$ in the remainder of this definition, to denote that process p can terminate according to the deduction rules of Figure 5, i.e., $p_{\surd}$, or cannot take any transition, i.e., there are no process p' and n-tuple of actions $\overline{a}$ s.t. $p \overset{\overline{a}}{\rightsquigarrow} p'$.

A run of a process p is a sequence $(\overline{l}_0, \overline{l}_1, \ldots, \overline{l}_m)$ of global states, such that there exist processes $p_0, p_1, \ldots, p_m$ with $\overline{l}_0 = (\langle\rangle, \ldots, \langle\rangle)$, $p_0 = p$, $p_{m\downarrow}$, and (if $m > 0$) for each $k < m$, it holds that $\overline{l}_{k+1} = \overline{l}_k \frown \overline{a}$ and $p_k \overset{\overline{a}}{\rightsquigarrow} p_{k+1}$.

The protocol associated with a process p, denoted as $[\![p]\!]$, is the set of all runs of p.

Example 5. **(Toy Example: Interpreted Systems)** The IS semantics for the proces $p = ((1)?a \,||\, (2)!a) + ((3)b\,; c)$ consists of the following runs:

$$(\langle\rangle, \langle\rangle, \langle\rangle), (\langle a\rangle, \langle a\rangle, \langle a_0\rangle)$$

$$\text{and } (\langle\rangle, \langle\rangle, \langle\rangle), (\langle\rangle, \langle\rangle, \langle b\rangle), (\langle c\rangle, \langle c\rangle, \langle b, c\rangle);$$

We see that the process histories of decorated actions are unfolded, through the annotations and the renaming functions, into the local perspectives of the principals in the IS semantics: local states are the sequences of (undecorated) actions from the history as perceived by the corresponding principal.

Example 6. **(Dining Cryptographers: Interpreted Systems)** In Figure 6, the initial steps of the runs of dining cryptographers protocol are depicted. Each tuple depicted in Figure 6, represents the view of principals O, 1, 2, and M, respectively, of the action that has take place.

The runs of the interpreted systems semantics, corresponding to the traces given in Example 4, are given below:

1. The first run corresponds to the leftmost trace in Figures 4 and 6; the global state of the run is a 4-tuple comprising the local states of O, 1, 2, and M, respectively (to save space, we have abbreviated the action names $flip$, $share$ and $bcast$ into fl, sh and bc, respectively):

$$
\begin{pmatrix}
\langle\rangle, \\
\langle\rangle, \\
\langle\rangle, \\
\langle\rangle
\end{pmatrix},
\begin{pmatrix}
\langle \mathit{pay}(1)\rangle, \\
\langle \mathit{pay}(1,\top)\rangle, \\
\langle \mathit{pay}(1)\rangle, \\
\langle \mathit{pay}(1,\top)\rangle
\end{pmatrix},
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top), \mathit{sh}(2,\bot)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top), \mathit{sh}(2,\bot)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top)\rangle, \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top)\rangle, \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top)\rangle
\end{pmatrix},
$$

$$
\begin{pmatrix}
\langle \mathit{pay}(1), \mathit{pay}(2), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top), \mathit{paid}(2,\top)\rangle \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2), \mathit{fl}(1,\top), \mathit{fl}(2), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top), \mathit{paid}(2,\top)\rangle \\
\langle \mathit{pay}(1), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2,\bot), \mathit{sh}(1,\top), \mathit{sh}(2,\bot), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top), \mathit{paid}(2,\top)\rangle \\
\langle \mathit{pay}(1,\top), \mathit{pay}(2,\top), \mathit{fl}(1), \mathit{fl}(2), \mathit{sh}(1), \mathit{sh}(2), \mathit{bc}(1,\top), \mathit{bc}(2,\bot), \mathit{paid}(1,\top), \mathit{paid}(2,\top)\rangle
\end{pmatrix}
$$

There are a couple of interesting observations to be made about the above-given run: firstly, each operational step of the protocol results in appending one action to the local state of each and every principal. This phenomenon, called synchronicity, is because of the particular definition of ρ, which does not map any action to the invisible action τ. Secondly, no principal can observe all actions as they actually happen in the protocol: a cryptographer cannot observe the content of the commuinication between the master and the other cryptographer and the result of its coin flip, the master cannot observe the result of the coin flip for any of the two cryptographer and the communication between the two cryptographers for sharing them, and the observer cannot observe any of the aforementioned information.

2. The second run corresponds to the middle trace in Figures 4 and 6 using the same abbreviations as in the first run:

$$\begin{pmatrix} \langle\rangle, \\ \langle\rangle, \\ \langle\rangle, \\ \langle\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1)\rangle, \\ \langle pay(1,\bot)\rangle, \\ \langle pay(1)\rangle, \\ \langle pay(1,\bot)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2)\rangle, \\ \langle pay(1,\bot), pay(2)\rangle, \\ \langle pay(1), pay(2,\top)\rangle, \\ \langle pay(1,\bot), pay(2,\top)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top)\rangle, \\ \langle pay(1), pay(2,\top), fl(1)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top), sh(2,\bot)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top), sh(2,\bot)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1), sh(2)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top), sh(2,\bot), bc(1,\top)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top), sh(2,\bot), bc(1,\top)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1), sh(2), bc(1,\top)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top)\rangle, \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot), paid(1,\top)\rangle, \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot), paid(1,\top)\rangle, \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top)\rangle \end{pmatrix},$$

$$\begin{pmatrix} \langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle \\ \langle pay(1,\bot), pay(2), fl(1,\top), fl(2), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle \\ \langle pay(1), pay(2,\top), fl(1), fl(2,\bot), sh(1,\top), sh(2,\bot), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle \\ \langle pay(1,\bot), pay(2,\top), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle \end{pmatrix}$$

Consider the first and the second run presented above and consider the local state corresponding to the view of the observer (viz. the first element in the global state); for convenience, we quote the local state of the observer in the final global state of both runs below:

$$\langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle$$
$$\langle pay(1), pay(2), fl(1), fl(2), sh(1), sh(2), bc(1,\top), bc(2,\bot), paid(1,\top), paid(2,\top)\rangle$$

As it can be seen, the local views of the observer in the two runs coincide and following Definition 2, these two runs are indistinguishable for the observer. However, the local states of the other principals differ in one or more actions (namely, *flip* and *pay*). These results have also been established in the operational semantic model of the protocol and they hint at the correspondence between the two semantic models. We formalize and prove this correspondence in the remainder of this paper.

5 Some Formal Results

In this section, we present three types of formal results regarding our interpreted systems semantics for *CCSi*. The first type establishes a correspondence between the operational and the interpreted systems semantics of *CCSi*. The second type of results determine the expressiveness of process algebraic specifications in generating interpreted systems. Finally, we give the third type of results about the characterization of the interpreted systems generated by process algebraic specifications.

Correspondence. The first result, formulated below, relates our interpreted systems semantics with the operational semantics originally defined for *CCSi*. It states that each component of a protocol in the interpreted systems semantics is a local projection on a trace obtained from the operational semantics.

Theorem 6. *For each CCSi process p, the following two statements hold:*

- $r = r(0), \ldots, r(m) \in [\![p]\!] \;\Rightarrow \exists_{p',\pi}(p, \langle\rangle) \rightarrow^* (p', \pi) \wedge p'_{\downarrow} \wedge \forall i \leq n,\, \pi \stackrel{i}{=} \lfloor r(m)_i \rfloor$,
- $\forall_{\pi}(p, \langle\rangle) \rightarrow^* (p', \pi) \wedge p'_{\downarrow} \Rightarrow \exists_{r \in [\![p]\!], m \in \mathbb{N}} r = r(0), ..., r(m) \wedge \forall i \leq n,\, \pi \stackrel{i}{=} \lfloor r(m)_i \rfloor$,

where $\rightarrow^*$ *denotes the reflexive transitive closure of the union of transition relations* $\stackrel{a}{\rightarrow}$ *, and* $\lfloor l_i \rfloor$ *is the sequence of actions in* l_i *lifted to form a history, by reading the actions as decorated implicitly with* $(\mathcal{I}d)$*, i.e. as publicly visible (cf. the convention on p. 184).*

The first item states that for each run in the IS semantics of p, there is a pair (p', π) with process p' terminating, and such that for each i, how i perceives the history π is equal to his local state $r(m)_i$. The proof goes by an induction on the length of the run. The second item states that conversely, for each process p that can terminate after history π, there is a run in the IS semantics of p in which for each i, i's local state at the end of the run captures how i perceives π. The proof of the second item is by an induction on the number of the transitions leading to (p', π).

Expressiveness. Before we study the expressiveness of process algebras in generating interpreted systems, we confine our attention to the set of interpreted systems of which the local states are initialized with the empty history and are updated at each step by at most one action; this is the idea behind the notion of *initialized* and *prefix-closed* interpreted systems defined below.

Definition 7. *Consider an interpreted system* (R, ν) *with sets* L_i *of local states such that* $L_i \subseteq \mathcal{A}ct^*$ *for each* $i \in \mathcal{I}d$*. A run* $r = r(0), \ldots, r(m) \in R$ *is* prefix closed *if for each two consecutive global states* $r(k) = (l_1, \ldots, l_n)$ *and* $r(k+1) = (l'_1, \ldots, l'_n)$ *with* $k < m$*, and each* $i \in \mathcal{I}d$ *it holds that either* $l_i = l'_i$ *or* $l_i \frown \alpha = l'_i$ *for some* $\alpha \in \mathcal{A}ct$*. The run* r *is* initialized *if* $r(0) = (\langle\rangle, \ldots, \langle\rangle)$*. Interpreted system* (R, ν) *is* initialized *and* prefix closed*, if each and every run* $r \in R$ *is initialized, and prefix closed.*

It trivially holds that the interpreted system semantics for our processes are initialized, and prefix-closed. The following theorems show the (lack of) expressiveness of process algebras in generating interpreted systems. The first two theorems show that all initialized and prefix-closed interpreted systems with 1 action or at most 2 agents can be specified by a process algebraic description. The third theorem shows that in the setting with more than 1 action and more than 2 agents, not all initialized and prefix closed interpreted systems can be captured by process algebraic specifications.

Theorem 8. *For an action set* $\mathcal{A}ct$ *with* $\mid \mathcal{A}ct \mid \leq 1$ *(i.e., with cardinality at most 1), for each finite initialized, and prefix-closed interpreted system* (R, ν)*, there exists a process algebraic description* p *and renaming* ρ *such that* $[\![p]\!] = R$*.*

Theorem 9. *Assume that the system comprises at most 2 agents; for each finite initialized and prefix-closed interpreted system* (R, ν)*, there exists a process algebraic description* p *and a renaming* ρ *such that* $[\![p]\!] = R$*.*

Theorem 10. *For an action set of cardinality at least 2 and more than 2 agents, there exist finite initialized and prefix-closed interpreted systems that cannot be generated by any process algebraic specification.*

Proof of Theorem 10. Consider the singleton protocol $\{((\langle\rangle, \langle\rangle, \langle\rangle), (\alpha, \beta, \gamma))\}$, where α, β and γ denote three distinct actions. We claim that this protocol cannot be generated by any process algebraic specification p with the given signature for ρ. Assume towards contradiction, that such a p exists; p cannot have non-ϵ summands or parallel components, since otherwise the protocol cannot be singleton. Hence, p should be an action prefixing followed by ϵ, i.e., is of the form $d; \epsilon$. It follows from Definition 5 that $[\![p]\!] = (\langle\rangle, \langle\rangle, \langle\rangle)^\frown [\![d]\!]$. Without loss of generality assume that $d = (J)\alpha$; then since $\beta \neq \langle\rangle$ and $\gamma \neq \langle\rangle$, it should hold that $\beta = \rho(\alpha)$ and $\gamma = \rho(\alpha)$, which contradicts the assumption that β and γ are distinct. ■

Theorem 10 points out a gap in the expressiveness of our process algebraic specification language. In the proof of the theorem, this shortcoming is traced back to the restrictive nature of our global renaming function: it presumes a dichotomy of actions and their public appearances while interpreted systems allow for several (more than 2) different appearances of actions. This expressiveness gap can be filled in various ways, e.g., by adding the principal identities as a parameter to the signature of the renaming function, thereby allowing for different appearance for different principals. This will be an important next step towards making our framework more general and increasing its expressiveness, especially for application in communication protocols.

Towards Characterization. The properties of the epistemic relations in the IS semantics for the *CCSi* process algebra derive from the signature and the properties of the renaming function. For example, if $\rho(a) = a$ for all a, then the equivalence classes of the $\stackrel{i}{\approx}$ are trivial (all singletons). The one action with a special interpretation, the silent action τ, plays a special role. For this paper, we have excluded τ as member of $\mathcal{A}ct$, but allowed it to be in the range of ρ. If $\rho(a) = \tau$ for some $a \neq \tau$, the renaming function enables modeling that some agents do not notice anything happening, when a actually happens.

If we would have allowed τ to be in $\mathcal{A}ct$ (and thereby in the domain of ρ), for a sensible interpretation of the intuition behind the renaming function, we should probably fix $\rho(\tau) = \tau$, although one could model the "illusion" that something happens when it actually doesn't, by allowing $\rho(\tau) = a$ for some $a \neq \tau$.

The three characterizing properties from [4] can be translated into our formalism as follows:

Synchronicity: if $r, r' \in [\![p]\!]$ and $r(m) \overset{i}{\approx} r'(m')$, then $m = m'$.

This property relates directly to a simple characteristic of the renaming function: we have synchronicity for process p with renaming ρ iff $[\rho(a) \neq \tau$ for all $a \neq \tau$ for which $(\mathtt{J})a \in p$ with $\mathtt{J} \neq \mathcal{I}d]$. If $(\mathtt{J})\tau$ were allowed to occur as action with $\mathtt{J} \neq \mathcal{I}d$, then the extra clause $\rho(\tau) = \tau$ needs to hold.

Perfect Recall: For all $r(m)^\frown[d], r'(m')^\frown[d'] \in [\![p]\!]$: $r(m)^\frown[d] \overset{i}{\approx} r'(m')^\frown[d']$ implies $r(m) \overset{i}{\approx} r'(m')$.

This property was proven to hold for the process algebra with identities in [9]. It holds for synchronous combinations of processes and renaming functions, the proof of which we omit here for the ske of brevity.

Uniform No Miracles: if $r(m) \overset{i}{\approx} r'(m')$ and there are $r''(m''), r'''(m''')$ and d, d' such that $r''(m'')^\frown[d] \overset{i}{\approx} r'''(m''')^\frown[d']$, then $r(m)^\frown[d] \overset{i}{\approx} r'(m')^\frown[d']$.

The investigation of the conditions under which this property holds for the ISs generated from processes and renaming functions is left for future work.

Discussion: Finite vs Infinite Runs. As we indicated before defining Interpreted Systems for our processes in Definition 1, an important difference with the standard account is the fact that we take runs to be finite sequences of global states (corresponding to the finite behavior of our processes) rather than infinite sequences. For future work, we consider generalizing our process language by including recursion, which would incorporate infinitely running processes. If we then generate the corresponding ISs in the way we do in this paper, these would contain both infinite and finite runs. This is still deviating from the standard notion of IS.

In order to turn finite runs (corresponding to finite behavior) into infinite runs could be to add an infinitely stuttering tail: after termination (or the inability to proceed) at time M, generate an infinite tail with $r(m) = r(m')$ for all $m, m' > M$. However, in our current framework, this would make us lose the property of *synchronicity*: for terminating processes then $r(m) \overset{i}{=} r(m')$ for all $m \neq m' > M$. In [15] problems of synchronization in distributed systems are solved by assuming hardware clocks within the processes. For us, implementing such assumption however, would have to be done beyond our process language (remember we take the agents, or principals, in $\mathcal{I}d$ as different entities than the processes).

6 Conclusions and Future Work

In this paper, we defined an interpreted systems semantics for a CCS-based process algebra. The defined semantics can be adopted for any other process-algebraic formalism as long as the visibility range and the public appearance of each atomic action in the process algebra is defined (either by using a richer syntax for atomic actions, or by providing this information as an addendum to the process-algebraic specification). We formally compared the interpreted systems semantics with the original operational semantics of the process algebra and provided a few semantic properties of the generated models by imposing restrictions on the public appearance function and the syntax of the process description.

There are two immediate next steps. The first is to include infinite behavior into the process language, but also to adapt the construction of ISs to generate more standard ISs, i.e. consisting of infinite runs. The second is to develop an epistemic temporal logical language to reason about the processes, by determining which set of propositions will be natural given the process language. Here it is relevant to keep in mind the potential application area of security protocols and the kind of properties relevant there.

We intend to extend this research and study the characteristics of interpreted systems models generated by different process algebras. Furthermore, we would like to mechanize our semantics in a tool in order to be able to verify epistemic properties of process-algebraic descriptions using the tools developed for interpreted systems.

Acknowledgements. We are grateful to the anonymous reviewers for their useful comments, suggestions, and pointers for future work.

References

1. Aceto, L., Birgisson, A., Ingólfsdóttir, A., Mousavi, M.: Decompositional Reasoning about the History of Parallel Processes. In: Arbab, F., Sirjani, M. (eds.) FSEN 2011. LNCS, vol. 7141, pp. 32–47. Springer, Heidelberg (2012)
2. Bhargava, M., Palamidessi, C.: Probabilistic Anonymity. In: Abadi, M., de Alfaro, L. (eds.) CONCUR 2005. LNCS, vol. 3653, pp. 171–185. Springer, Heidelberg (2005)
3. Baeten, J.C.M., Basten, T., Reniers, M.A.: Process Algebra: Equational Theories of Communicating Processes. Cambridge University Press (2009)
4. van Benthem, J., Gerbrandy, J., Hoshi, T., Pacuit, E.: Merging frameworks for interaction. Journal of Philosophical Logic 38(5), 491–526 (2009)
5. Boureanu, I., Cohen, M., Lomuscio, A.: A Compilation Method for the Verification of Temporal-Epistemic Properties of Cryptographic Protocols. Journal of Applied Non-Classical Logics 19(4), 463–487 (2009)
6. Chadha, R., Delaune, S., Kremer, S.: Epistemic Logic for the Applied Pi Calculus. In: Lee, D., Lopes, A., Poetzsch-Heffter, A. (eds.) FMOODS 2009. LNCS, vol. 5522, pp. 182–197. Springer, Heidelberg (2009)

7. Chatzikokolakis, K., Knight, S., Panangaden, P.: Epistemic Strategies and Games on Concurrent Processes. In: Nielsen, M., Kučera, A., Miltersen, P.B., Palamidessi, C., Tůma, P., Valencia, F. (eds.) SOFSEM 2009. LNCS, vol. 5404, pp. 153–166. Springer, Heidelberg (2009)
8. Chaum, D.: The dining cryptographers problem: Unconditional sender and recipient untraceability. Journal of Cryptology 1, 65–75 (1988)
9. Dechesne, F., Mousavi, M.R., Orzan, S.: Operational and Epistemic Approaches to Protocol Analysis: Bridging the Gap. In: Dershowitz, N., Voronkov, A. (eds.) LPAR 2007. LNCS (LNAI), vol. 4790, pp. 226–241. Springer, Heidelberg (2007)
10. Dechesne, F., Mousavi, M.R., Orzan, S.: Operational and epistemic approaches to protocol analysis: bridging the gap. Technical Report CSR-07-15, Department of Computer Science, Eindhoven University of Technology (2007)
11. Dechesne, F., Wang, Y.: To know or not to know: epistemic approaches to security protocol verification. Synthese 177, 51–76 (2010)
12. van Eijk, R., de Boer, F., van der Hoek, W., Meyer, J.-J.C.: Operational Semantics for Agent Communication Languages. In: Dignum, F.P.M., Greaves, M. (eds.) Issues in Agent Communication. LNCS, vol. 1916, pp. 80–95. Springer, Heidelberg (2000)
13. van Eijk, R.M., de Boer, F.S., van der Hoek, W., Meyer, J.-J.Ch.: Process Algebra for Agent Communication: A General Semantic Approach. In: Huget, M.-P. (ed.) Communication in Multiagent Systems. LNCS (LNAI), vol. 2650, pp. 113–128. Springer, Heidelberg (2003)
14. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. MIT Press (1995)
15. Halpern, J.Y., Moses, Y.: Knowledge and Common Knowledge in a distributed environment. In: Proceedings of the 3rd ACM Conference on Principles of Distributed Computing, pp. 50–61. ACM (1984); Reprinted in the Journal of the Association for Computing Machinery 37(3), 549–587 (1990)
16. Halpern, J.Y., O'Neill, K.R.: Anonymity and information hiding in multiagent systems. Journal of Computer Security, 483–514 (2005)
17. Halpern, J., Vardi, M.Y.: Reasoning about knowledge and time in asynchronous systems. In: Proceedings of STOC 1988, pp. 53–65. ACM Press (1988)
18. Hennessy, M., Milner, R.: Algebraic laws for nondeterminism and concurrency. J. ACM 32(1), 137–161 (1985)
19. Hoare, T.: Communicating Sequential Processes. Prentice-Hall (1985)
20. van der Hoek, W., van Hulst, M., Meyer, J.-J.Ch.: Towards an Epistemic Approach to Reasoning about Concurrent Programs. In: de Bakker, J.W., de Roever, W.-P., Rozenberg, G. (eds.) REX 1992. LNCS, vol. 666, pp. 261–287. Springer, Heidelberg (1993)
21. Hoshi, T.: Merging DEL and ETL. J. of Logic, Lang. and Inf. 19, 413–430 (2010)
22. Hughes, D., Shmatikov, V.: Information hiding, anonymity and privacy: A modular approach. Journal of Computer Security 12(1), 3–36 (2004)
23. van Hulst, M., Meyer, J.-J.Ch.: An epistemic proof system for parallel processes: extended abstract. In: Proceeding of TARK 1994, pp. 243–254. ACM Press (1994)
24. Katz, S., Taubenfeld, G.: What Processes Know: Definitions and Proof Methods. In: Proceedings of PODC 1986, pp. 249–262. ACM Press (1986)
25. Kramer, S., Palamidessi, C., Segala, R., Turrini, A., Braun, C.: A quantitative doxastic logic for probabilistic processes and applications to information-hiding. The Journal of Applied Non-Classical Logic 19(4), 489–516 (2009)
26. Lomuscio, A., Ryan, M.: Ideal agents sharing (some!) knowledge. In: Proceedings of ECAI 1998, pp. 557–561. John Wiley and Sons (1998)

27. Lomuscio, A., Qu, H., Raimondi, F.: MCMAS: A Model Checker for the Verification of Multi-Agent Systems. In: Bouajjani, A., Maler, O. (eds.) CAV 2009. LNCS, vol. 5643, pp. 682–688. Springer, Heidelberg (2009)
28. van der Meyden, R., Su, K.: Symbolic model checking the knowledge of the dining cryptographers. In: Proceedings of CSFW 2004, pp. 280–291. IEEE (2004)
29. Milner, R.: A Calculus of Communication Systems. LNCS, vol. 92. Springer, Heidelberg (1980)
30. Parikh, R., Ramanujam, R.: Distributed Processes and the Logic of Knowledge. In: Parikh, R. (ed.) Logic of Programs 1985. LNCS, vol. 193, pp. 256–268. Springer, Heidelberg (1985)
31. Park, D.M.R.: Concurrency and Automata on Infinite Sequences. In: Deussen, P. (ed.) GI-TCS 1981. LNCS, vol. 104, pp. 167–183. Springer, Heidelberg (1981)
32. Plotkin, G.D.: A structural approach to operational semantics. Journal of Logic and Algebraic Programming 60, 17–139 (2004)
33. Raimondi, F., Lomuscio, A.: A tool for specification and verification of epistemic properties in interpreted systems. Electronic Notes in Theoretical Computer Science 85(4) (2004)
34. Richards, S., Sadrzadeh, M.: Aximo: Automated axiomatic reasoning for information update. In: Proceedings of M4M5 2007. ENTCS, vol. 231, pp. 211–225. Elsevier (2009)
35. Schneider, S., Sidiropoulos, A.: CSP and Anonymity. In: Martella, G., Kurth, H., Montolivo, E., Bertino, E. (eds.) ESORICS 1996. LNCS, vol. 1146, pp. 198–218. Springer, Heidelberg (1996)
36. Toninho, B., Caires, L.: A spatial-epistemic logic for reasoning about security protocols. In: Proceedings of SecCo 2010. EPTCS, vol. 51, pp. 1–15 (2010)
37. de Vries, W., de Boer, F.S., van der Hoek, W., Meyer, J.-J.Ch.: A Truly Concurrent Model for Interacting Agents. In: Yuan, S.-T., Yokoo, M. (eds.) PRIMA 2001. LNCS (LNAI), vol. 2132, pp. 16–30. Springer, Heidelberg (2001)

The Duality of State and Observation in Probabilistic Transition Systems

Monica Dinculescu[1], Christopher Hundt[2], Prakash Panangaden[3], Joelle Pineau[3], and Doina Precup[3]

[1] Morgan Stanley, Montreal, Canada
[2] Google Inc., Mountain View, California, USA
[3] McGill University, Montréal, Québec

Abstract. In this paper we consider the problem of representing and reasoning about systems, especially probabilistic systems, with hidden state. We consider transition systems where the state is not completely visible to an outside observer. Instead, there are observables that partly identify the state. We show that one can interchange the notions of state and observation and obtain what we call a dual system. In the case of deterministic systems, the double dual gives a minimal representation of the behaviour of the original system. We extend these ideas to probabilistic transition systems and to partially observable Markov decision processes (POMDPs).

1 Introduction

Learning and planning under uncertainty is a crucial focus of modern AI research. In the AI literature there is much discussion of the role of "state" and there is a point of view that asserts that "a canonical notion of state does not exist." According to this view, state is merely a sufficient statistic for predicting the future. This view has found its most articulate and developed exposition in what is called the "predictive representation of state" or PSRs [18, 24]. The present paper arose from an attempt to understand PSRs from a foundational point of view as well as to understand certain well-known learning algorithms [33] that seem implicitly to use some ideas from duality.

The main point of the present paper is the presence of a duality between state and observation which seems to lie at the heart of the PSR representation. The use of the word "duality" of course evokes connections with the duality between logic and transition systems and of the many dualities known in mathematics: Stone duality and its many variants and extensions, Priestley duality, Gelfand duality and general concrete dualities in category theory; see, for example, the excellent monograph of Johnstone [19] for a general categorical discussion. Duality has been very important in systems theory and expresses a relation between observability and controllability [21].

Much of the work in AI planning and learning under uncertainty is based on the framework of Partially Observable Markov Decision Processes (POMDPs) [20]. In this framework, problems are modeled using discrete states and actions.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 206–230, 2013.

Actions cause stochastic transitions between states. At each time step, a stochastic observation is also generated, based on the current state and the previous action. Much work has been devoted to planning in POMDPs when a model of the system (in terms of the stochastic transitions between states and the probability distributions over observations) is known. Unfortunately, learning POMDPs from data is a very difficult problem. One standard algorithmic solution is expectation maximization (EM) [11], but for POMDPs this approach is plagued by local minima (more so than for other probabilistic models) and works poorly in practice unless a good initial model of the system is used. History-based methods [25] often work better in practice, but are less general. A lot of recent research has been devoted to finding alternative representations for such systems, e.g., diversity-based representation [33], predictive state representations (PSRs) [24] and TD-networks [37]. These approaches aim to combine the generality of POMDPs with the ease of learning of history-based methods. The key idea underlying all of these approaches is that the state of the system is not considered as predefined; instead, it is viewed as a sufficient statistic for making (probabilistic) predictions about future trajectories. However, the models themselves are different and their relationships are only partially understood at the moment. It is our hope that the theory of the present paper will serve as a foundation for these different models and bring into focus the commonalities and differences.

In this paper, we develop a duality theory for POMDPs, which unifies much of the existing work on predictive representations. We show how, for any POMDP, one can develop two alternative representations: a dual machine and a double-dual machine. The key idea in the development is that of making measurements on the system, which we call experiments. Experiments are sequences of actions interspersed with observations. They generalize previous notions of tests from the literature on predictive state representations. Both of the alternative representations that we present allow an accurate prediction of the probability of any experiment. The double-dual representation is of particular interest, because it has a *deterministic* transition structure, and no hidden state. Instead, its states can be thought of as "bundles" of predictions for experiments. As such, this representation holds the promise of much better planning and learning algorithms than those currently available. Our work also generalizes similar representations from automata theory [10] and is closely related to the update graph from [33]. We show how existing predictive representations can be viewed from the perspective of this framework. We also discuss the implications of these alternative representations for learning algorithms, approximate planning algorithms as well as working with continuous observations. A preliminary version of these ideas has appeared in [17].

The main technical result of the present paper is that when one constructs the *double dual* one obtains a *minimal behaviourally equivalent* version of the original system. Of course, as written this cannot be quite right! One should get an isomorphic object when one goes back and forth across a dual situation. Nevertheless, this is what happened with the construction that we present in this paper. This shows that what we have cannot be a pure Stone-type duality. It took a long time to fit these results into a coherent categorical picture. Of course, in the probabilistic setting the systems are infinite so "minimal" does not mean fewest states but is best expressed as a couniversal property; we will come back to this point in the conclusions.

The categorical version of the results of the present paper are being written up in a separate paper by the first author and Nick Bezhanishvili and Clemens Kupke. In that work the authors deal with weighted automata as well. The main idea there is that there is a duality between the transition system and an appropriate algebra *with additional operators*. Thus, for ordinary automata, one has a duality between the automata and boolean algebras equipped with "modal" operators while for the probabilistic case one has C^*-algebras equipped with additional operators. It turns out that the problem of minimizing a transition system can be seen in the dual category as the problem of finding a 0-generated subalgebra. Thus going to the dual, finding the 0-generated subalgebra and returning via the duality automatically minimizes the transition system.

2 Background

In this section we review the definitions of the various kinds of transition systems that we work with in this paper. The type of system most used in applications to planning under uncertainty and learning are partially observable Markov decision processes (POMDPs). There are, however, a number of simpler situations where the duality phenomenon also occurs and we will discuss duality for these systems before going on to POMDPs [20].

The crucial ingredient is the interplay between the dynamics, i.e. the state-to-state transitions, and the observations. In automata theory, the concept of partially observable system is often not made explicit in the definitions. In systems theory [21], and in artificial intelligence, one typically has observations that only partially reveal the state. One, thus, has to reason about behavioural equivalences between systems; bisimulation [26, 29, 30] is the most common of these equivalences[1]. In process algebra one has a similar situation: one does not see the state, only whether actions are "accepted" or "rejected" in a given state. These approaches have equivalent modelling power and concepts like bisimulation can be defined in both settings. In this paper we will always assume that all actions are possible in every state and the behavioural equivalences will take the observations into account. The usual example from process algebra [26, 27] showing the difference between trace equivalence and bisimulation can be mimicked in this setting.

The observations that we use are to be thought of as Boolean valued. Intuitively, one thinks of a black box with a number of buttons (the actions) and a number of lights (the observations). One can press any button and induce some

[1] Though [29] is often cited for bisimulation it is only mentioned there in passing and Park never wrote a paper on bisimulation; the slides [30] are the closest thing to a proper citation for Park's contribution.

kind of internal state transition - which will be invisible - but some of the lights may light up. In the most general case the observations depend on the action and the target state or source state of each transition. In the AI literature it is more usual to consider the observation as depending on the target (posterior) state and we will do that here when we discuss POMDPs. We discuss everything in the context of discrete state spaces so we avoid all measure-theoretic complications and we can work with state-to-state transition probabilities.

Of course, this is a first step. Eventually we will be interested in the case where the observables are real-valued – as in the case of rewards – or indeed vectors of real-values. We will also be interested in the case where the state space is continuous. In that case we will be concerned with developing a good approximation theory so that one can have tractable representations of continuous systems.

2.1 Kripke Automata

We begin with ordinary automata enriched with a notion of observations associated with each state; the observations are deterministic.

Definition 1. *A* ***deterministic Kripke automaton*** *(DKA) is a quintuple*

$$\mathcal{K} = (S, \mathcal{A}, \mathcal{O}, \delta : S \times \mathcal{A} \longrightarrow S, \gamma : S \longrightarrow 2^{\mathcal{O}}).$$

S is a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{O}$ is a set of observations, δ is a transition function and γ is an observation function associated with the states.

The idea here is that there are a number of observations that will be made in a given state; in terms of our previous imagery, for each state there is a set of lights that are turned on in each state. If we call the observations "propositions" instead, this is essentially the definition of Kripke structure [12] used in model checking, except that there are labels on transitions as well, and we have made the transitions deterministic. The automata studied in undergraduate courses [16, 23, 35] are special cases where there is a single observation called "accept."

Note that the observations are associated with the states, like in a Moore machine. Given a state there is a set of observations that are *always* made in that state. One can view γ as a relation; we will switch between γ as defined above and $\hat{\gamma} \subseteq S \times \mathcal{O}$ where $\hat{\gamma}(s, \omega) \Leftrightarrow \omega \in \gamma(s)$. We will usually not even bother to write $\hat{\gamma}$ and just use γ for whatever version is most apt to the situation at hand.

One can take the view that with a state there is always a single observation: the complete description of which lights are on. The picture given above can be encoded in this view by taking the set of observations to be $2^{\mathcal{O}}$. Furthermore, the latter view is a special case of our definition: we are just restricting γ to be a function.

While these two views are the same in the non-probabilistic case, they differ sharply in the probabilistic case. If we were just to give a probability for a given observation in a state, we could not express *correlations* between observations. Thus, when we come to the probabilistic case, we will insist that γ determines a distribution over all possible observations. Of course these observations could be "structured" in some way and we could analyze aspects of this structure. We are planning to explore these ideas in future work.

2.2 Probabilistic Systems

In systems theory, one often considers systems where the transitions and the observations are probabilistic. This gives a probabilistic version of Kripke automata; the interpretation of γ is, as we explained in the paragraph above, generalized to a distribution over the observations for each state.

Definition 2. *A* ***partially observable probabilistic automaton*** *(POPA) is a quintuple*

$$\mathcal{H} = (S, \mathcal{A}, \mathcal{O}, \tau : S \times \mathcal{A} \times S \longrightarrow [0,1], \gamma : S \times \mathcal{O} \longrightarrow [0,1])$$

where $\tau(s, a, \cdot)$ *defines a probability distribution on possible target states and* $\gamma(s, \omega)$ *is the probability of observing* ω *in state* s*. We will often write* $\tau_a(s, \cdot)$ *for* $\tau(s, a, \cdot)$*. We write* $\tau_a(s, X)$*, where* $X \subseteq S$*, for* $\sum_{t \in X} \tau_a(s, t)$*.*

In each state we can observe possibly several (or no) observations. The number $\gamma(s, \omega)$ is the probability that one sees ω, as opposed to not seeing it, given that one is in the state s. The function $\omega \mapsto \gamma(s, \omega)$ is not necessarily a probability distribution on $\mathcal{O}$. It defines a probability distribution on $2^{\mathcal{O}}$. Some of our constructions on partially observable probabilistic automata will yield deterministic variants. The interesting case is where the transitions are deterministic, but the observations are still probabilistic.

Definition 3. *A* ***deterministic automaton with stochastic observations*** *(DASO) is a quintuple*

$$\mathcal{J} = (S, \mathcal{A}, \mathcal{O}, \delta : S \times \mathcal{A} \longrightarrow S, \gamma : S \times \mathcal{O} \longrightarrow [0,1])$$

where $\gamma(s, \omega)$ *is the probability of observing* ω *in state* s*.*

Here the transitions are deterministic, hence, given by a function, but the states can only be partially known through the observations, which are stochastic.

In POMDPs the observations are associated with the transitions rather than with the states.

Definition 4. *A* ***Partially Observable Markov Decision Process*** *(POMDP) is a quintuple*

$$\mathcal{M} = (S, \mathcal{A}, \mathcal{O}, P : S \times \mathcal{A} \times S \longrightarrow [0,1], \gamma : \mathcal{A} \times S \times \mathcal{O} \longrightarrow [0,1])$$

where, as before, S *is a set of states,* $\mathcal{A}$ *is a set of actions,* $\mathcal{O}$ *is a set of observations,* P *is the transition probability function and* γ *gives the observation probabilities. For each* $s \in S$ *and* $a \in \mathcal{A}$*, the function* $\omega \mapsto \gamma(a, s, \omega)$ *is a distribution on* $\mathcal{O}$*.*

Note that here we are requiring that γ defines a distribution. This is done to match the definition used in the extant literature for POMDPs. However, when we construct duals and double duals this will no longer be the case. The number $\gamma(a, s, \omega)$ is the probability that one sees the observation ω, given that the system takes the action a and given that *one ends up in s* after the transition is complete[2]. According to our definition, transitions are triggered by the actions in $\mathcal{A}$; there is no attempt to model probability distributions over the actions. They are intended to be actions chosen by an adversary or scheduler (the verification viewpoint) or actions chosen by a policy external to the system definition (the planning viewpoint). The triple $(S, \mathcal{A}, P)$ forms a labelled Markov process [14] (LMP). If we had a reward as well, it would be a Markov decision process (MDP) [32]. The fact that the state is only partially observable is captured by the fact that $\mathcal{O}$ is different from S and there is no given bijection between S and $\mathcal{O}$.

3 Duality for Kripke Automata

We begin with a simple example which is instructive and gives a foretaste of the rest of the paper. We work with deterministic Kripke automata, i.e. ordinary automata enhanced with a notion of observation, and establish a very pleasing duality between state and observation. The situation is rather special, in that the duality relation is very tight: the stochastic case does not have all the features of the present case. The results of the present section are a slight generalization of a technique known to Brzozowski [10] in 1962. We defer a comparison to Brzozowski's work to the end of this section.

We recall the definitions. Let $\mathcal{K} = (S, \mathcal{A}, \mathcal{O}, \delta, \gamma)$ be a deterministic Kripke automaton. Here S is the set of states, $\mathcal{A}$ an alphabet of input symbols, $\mathcal{O}$ is a set of primitive observations, δ is the transition function $\delta : S \times \mathcal{A} \longrightarrow S$ and $\gamma : S \longrightarrow 2^{\mathcal{O}}$ is a labelling function. One can as well think of the elements of $\mathcal{O}$ as *propositions* capturing basic properties of the states or as *observations* – boolean-valued in this case – that one can make of the states. One can equally well think of γ as having the type $S \times \mathcal{O} \longrightarrow \{T, F\}$. We will emphasize the notion of observation and testing rather than the equivalent notion of proposition and modal formula.

Thinking of the elements of $\mathcal{O}$ as basic observations, we can use them to define a family of tests. We define a *test t* according to the following grammar:

$$t ::== \omega \in \mathcal{O} \mid (a) \cdot t$$

where $a \in \mathcal{A}$.

We say that a state s satisfies or *passes* ω, written $s \models \omega$, if $\omega \in \gamma(s)$ or, equivalently, $\gamma(s, \omega) = T$. We say $s \models (a) \cdot t$ if $\delta(s, a) \models t$. We define $[\![t]\!]_{\mathcal{K}} = \{s \in S | s \models t\}$. Clearly this is exactly the same as defining the tests as modal formulas and defining satisfaction as above.

[2] It would be more natural, perhaps, to make this depend on the source state; we are following the convention used by AI researchers [20].

We define an equivalence relation $\sim_{\mathcal{K}}$ between *tests* as $t_1 \sim_{\mathcal{K}} t_2$ if $[\![t_1]\!]_{\mathcal{K}} = [\![t_2]\!]_{\mathcal{K}}$. Note that this allows us to identify an equivalence class for t with the set of states $[\![t]\!]_{\mathcal{K}}$ that satisfy t. Note that another way of defining this equivalence relation is

$$t \sim_{\mathcal{K}} t' := \forall s \in S.s \models t \iff s \models t'.$$

We also define an equivalence $\equiv$ between *states* in $\mathcal{K}$ as $s_1 \equiv s_2$ if for all tests t on $\mathcal{K}$, $s_1 \models t \iff s_2 \models t$. The equivalence relations $\sim$ and $\equiv$ are clearly closely related: they are the hinge of the duality between states and observations.

We say that $\mathcal{K}$ is *reduced* if it has no $\equiv$-equivalent states. Since there is more than just one observation, in general the relation $\equiv$ is finer than the usual equivalence of automata theory.

Finally, we say that two DKAs $\mathcal{K} = (S, \mathcal{A}, \mathcal{O}, \delta, \gamma)$ and $\mathcal{K}' = (S', \mathcal{A}', \mathcal{O}', \delta', \gamma')$ are *isomorphic* if $\mathcal{A} = \mathcal{A}'$, $\mathcal{O} = \mathcal{O}'$, and there exists a bijection $\phi : S \longrightarrow S'$ such that, for all $s \in S$, $\gamma(s) = \gamma'(\phi(s))$ and for all $a \in \mathcal{A}$ $\phi(\delta(s, a)) = \delta'(\phi(s), a)$.

We define the dual construction as follows.

Definition 5. *Let $\mathcal{K}$ be a Kripke automaton $\mathcal{K} = (S, \mathcal{A}, \mathcal{O}, \delta, \gamma)$. Let T be the set of $\sim_{\mathcal{K}}$-equivalence classes of tests on $\mathcal{K}$. We define $\mathcal{K}' = (S', \mathcal{A}, \mathcal{O}', \delta', \gamma')$ as follows:*

- $S' = T = \{[\![t]\!]_{\mathcal{K}}\}$
- $\mathcal{O}' = S$
- $\delta'([\![t]\!]_{\mathcal{K}}, a) = [\![(a) \cdot t]\!]_{\mathcal{K}}$, $\forall [\![t]\!]_{\mathcal{K}} \in S', a \in \mathcal{A}$
- $\forall [\![t]\!]_{\mathcal{K}} \in S'$ $\gamma'([\![t]\!]_{\mathcal{K}}) = [\![t]\!]_{\mathcal{K}}$ or $\gamma'([\![t]\!]_{\mathcal{K}}, s) = (s \models t)$

The somewhat strange-looking definition of γ' is to be understood as follows. In the machine $\mathcal{K}'$ the observations one can make of the state $[\![t]\!]$ are those states of $\mathcal{K}$ (which are the observations of $\mathcal{K}'$) that satisfy the test t; this set is exactly $[\![t]\!]$. We have interchanged the states and the observations; more precisely we have interchanged equivalence classes of tests - based on the observations - with the states. We have made the states of the old machine the observations of the dual machine. To see the remarkable effect of this interchange we consider the double dual. We will see that the double dual is the minimal automaton with the same behaviour.

Now consider $\mathcal{K}'' = (\mathcal{K}')'$, the dual of the dual. Its states are equivalence classes of $\mathcal{K}'$-tests. Each such class is identified with a set $[\![t']\!]_{\mathcal{K}'}$ of $\mathcal{K}'$-states by which tests in that class are satisfied, and each $\mathcal{K}'$-state is an equivalence class of $\mathcal{K}$-tests. Thus we can look at states in $\mathcal{K}''$ as collections of $\mathcal{K}$-test equivalence classes. It is with this perception in mind that we construct the *Sat* function. To avoid confusion, we will write $\hat{s}$ for a state $s \in S$ when viewed as an observation of $\mathcal{K}'$.

Definition 6. *Let $\mathcal{K}''$ be the double dual of $\mathcal{K}$. For any state $s \in S$ of $\mathcal{K}$ we define $Sat(s) = \{[\![t]\!]_{\mathcal{K}} | s \models t\}$.*

The next lemma shows that these sets are always states of the double dual. The following notation will be useful. A state s of $\mathcal{K}$ is an observation of $\mathcal{K}'$; recall that $\hat{s}$ is s viewed as an observation of $\mathcal{K}'$.

Lemma 7. *Let $s \in S$ be any state in the original automaton $\mathcal{K}$. Then $Sat(s)$ is a state in $\mathcal{K}''$.*

Proof. The observations in $\mathcal{K}'$ are the states in $\mathcal{K}$, i.e. elements of S. Recall that $\hat{s} \in \mathcal{O}'$ is the observation associated with s. Then $[\![\hat{s}]\!]_{\mathcal{K}'}$ is a state in $\mathcal{K}''$, and

$$[\![\hat{s}]\!]_{\mathcal{K}'} = \{[\![t]\!]_{\mathcal{K}} | [\![t]\!]_{\mathcal{K}} \models \hat{s}\} = \{[\![t]\!]_{\mathcal{K}} | s \in [\![t]\!]_{\mathcal{K}}\} = \{[\![t]\!]_{\mathcal{K}} | s \models t\} = Sat(s).$$

∎

In fact *all* the states of the double dual have this form.

Lemma 8. *Let $s'' = [\![t]\!]_{\mathcal{K}'} \in S''$ be any state in $\mathcal{K}''$. Then $s'' = Sat(s_t)$ for some state $s_t \in S$.*

Proof. The proof is by induction on the length of t. The base case is settled by Lemma 7.

Now suppose $t = (a) \cdot t'$ for some t'. Then, by the inductive hypothesis, there is some $s_{t'}$ such that $[\![t']\!]_{\mathcal{K}'} = Sat(s_{t'})$. Let $s_t = \delta(s_{t'}, a)$. Then

$$\begin{aligned} [\![t]\!]_{\mathcal{K}'} &= \{[\![r]\!]_{\mathcal{K}} | [\![r]\!]_{\mathcal{K}} \models t\} = \{[\![r]\!]_{\mathcal{K}} | [\![r]\!]_{\mathcal{K}} \models (a) \cdot t'\} = \{[\![r]\!]_{\mathcal{K}} | \delta'([\![r]\!]_{\mathcal{K}}, a) \models t'\} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \models t'\} = \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \in [\![t']\!]_{\mathcal{K}'}\} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \in Sat(s_{t'})\} = \{[\![r]\!]_{\mathcal{K}} | s'_t \models (a) \cdot r\} \\ &= \{[\![r]\!]_{\mathcal{K}} | \delta(s_{t'}, a) \models r\} = \{[\![r]\!]_{\mathcal{K}} | s_t \models r\} = Sat(s_t). \end{aligned}$$

We have used the induction hypothesis in the first equality of the penultimate line. ∎

Now an immediate consequence of the definitions is:

Observation 9. *$s \equiv s'$ if and only if $Sat(s) = Sat(s')$.*

Now from Lemmas 7 and 8 and the observation we have the following corollary.

Corollary 10. *If $\mathcal{K}$ is reduced then Sat is a bijection from S to S''.*

Proof. Lemma 7 shows that $Sat : S \longrightarrow S''$. The fact that $\mathcal{K}$ is reduced means that $s_1 \neq s_2 \implies Sat(s_1) \neq Sat(s_2)$, which by Observation 9 implies that Sat is injective. Lemma 8 shows that Sat is surjective. Thus Sat is a bijection from S to S''. ∎

The statement of the corollary can be strengthened to show that we actually have an isomorphism of DKAs between $\mathcal{K}$ and $\mathcal{K}''$. We know that the action set $\mathcal{A}$ is the same for both $\mathcal{K}$ and $\mathcal{K}''$, but the observation set $\mathcal{O}$ is not equal to the observation set $\mathcal{O}''$. However, we can transform $\mathcal{K}''$ by just restricting to the original observations in the following way: recall that an observation in $\mathcal{K}''$ is a state in $\mathcal{K}'$, which is an equivalence class of tests in $\mathcal{K}$. We let $\mathcal{O}$ be the observations for $\mathcal{K}''$ and we define a new observation function γ''_T for $\mathcal{K}''$ as

$$\forall s'' \in S'' \; \omega \in \gamma''_T(s'') \iff [\![\omega]\!]_{\mathcal{K}} \in \gamma''(s'').$$

Now we can define $\mathcal{K}''_T$.

Definition 11. *$\mathcal{K}''_T = (S'', \mathcal{A}, \mathcal{O}, \delta'', \gamma''_T)$, with γ''_T defined as above.*

This allows us to establish the required isomorphism.

Theorem 12. *Suppose $\mathcal{K}$ is reduced. Then $\mathcal{K}$ is isomorphic to $\mathcal{K}''_T$.*

Proof. We see that $\mathcal{A}$ and $\mathcal{O}$ are the same in both $\mathcal{K}$ and $\mathcal{K}''_T$. We use Sat as our bijection from S to S''. It remains only to verify the properties of Sat. First, the observations: for all $s \in S$ and all $\omega \in \mathcal{O}$ the following holds:

$$\begin{aligned}\omega \in \gamma(s) &\iff s \models \omega \\ &\iff [\![\omega]\!]_{\mathcal{K}} \in Sat(s) \\ &\iff [\![\omega]\!]_{\mathcal{K}} \in \gamma''(Sat(s)) \\ &\iff \omega \in \gamma''_T(Sat(s)).\end{aligned}$$

We now check the transitions. For any $s \in S$ and any $a \in \mathcal{A}$, let t be such that $Sat(s) = [\![t]\!]_{\mathcal{K}'}$. Then

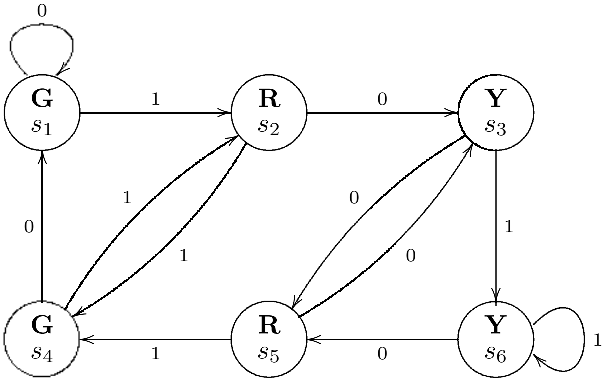
$$\begin{aligned}\delta''(Sat(s), a) &= \delta''([\![t]\!]_{\mathcal{K}'}, a) \\ &= [\![(a) \cdot t]\!]_{\mathcal{K}'} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![r]\!]_{\mathcal{K}} \models (a) \cdot t\} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \models t\} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \in [\![t]\!]_{\mathcal{K}'}\} \\ &= \{[\![r]\!]_{\mathcal{K}} | [\![(a) \cdot r]\!]_{\mathcal{K}} \in Sat(s)\} \\ &= \{[\![r]\!]_{\mathcal{K}} | s \models (a) \cdot r\} \\ &= \{[\![r]\!]_{\mathcal{K}} | \delta(s, a) \models r\} \\ &= Sat(\delta(s, a)).\end{aligned}$$

Thus Sat establishes an isomorphism between $\mathcal{K}$ and $\mathcal{K}''_T$. ∎

Thus the double dual construction produces a machine which is - in a very strong sense - the "minimal version" of the original machine. What we mean by minimal is that no further reduction or collapsing of states is possible. We will expand on this in the conclusions.

In the stochastic case we will not get such a tight correspondence but this gives a preview of what will happen there. In fact analogous results work for the nondeterministic case; in this case the double dual is the minimized version of the equivalent deterministic machine.

An Extended Example. We will explain the concept of duality on a concrete example, using the finite automaton below, where $\mathcal{S} = \{s_1, \ldots, s_6\}$, $\mathcal{A} = \{0, 1\}$ and $\mathcal{O} = \{\mathbf{G}, \mathbf{R}, \mathbf{Y}\}$.



In the above machine $010\mathbf{Y}$, $\mathbf{R}$, and $110\mathbf{R}$ are examples of tests. Suppose that we start with state s_1 in our example. If we follow the sequence of actions 0010, we end up in state s_3, which gives an observation of $\mathbf{Y}$. Thus, we can say that $s_1 \models 0010\mathbf{Y}$. It is easy to see that $s_4 \models 01\mathbf{R}$, $s_1 \models 011\mathbf{G}$ and $s_6 \models 1100\mathbf{Y}$.

Consider the test $0\mathbf{G}$ in the above example. We notice that only s_1 and s_4 satisfy it. In order to find other tests equivalent to it, we should look at tests that only s_1 and s_4, and no other states, satisfy. Other such tests are $\mathbf{G}, 00\mathbf{G}, 000\mathbf{G} \ldots$ etc. Thus, we can say that $[\mathbf{G}] = [0\mathbf{G}] = [00\mathbf{G}] \ldots$. Similarly, we find that only the states s_2 and s_5 satisfy the equivalent tests $[\mathbf{R}] = [11\mathbf{R}] = [101\mathbf{R}] \ldots$, and the states s_3 and s_6 satisfy $[\mathbf{Y}] = [1\mathbf{Y}] = [100\mathbf{Y}] \ldots$.

As we have said before, an equivalence class of tests is identified by the set of states that satisfy these tests. Then, in our example, the equivalence classes of tests are: $t_1 = \{s_1, s_4\}$, $t_2 = \{s_2, s_5\}$ and $t_3 = \{s_3, s_6\}$.

We use the usual labelled transition notion: $s_1 \xrightarrow{a} s_2$ when a transition on action a has source s_1 and target s_2. Notice that not only $s_1 \xrightarrow{0} s_1$, and $s_4 \xrightarrow{0} s_1$, but $s_1 \xrightarrow{1} s_2$, and $s_4 \xrightarrow{1} s_2$ as well. Furthermore, these states have the same observations. This means that s_1 and s_4 have the same behaviour, and thus any test satisfied by one must also be satisfied by the other. Thus, $s_1 \equiv s_4$. Following the same reasoning, we can say that $s_2 \equiv s_5$ and $s_3 \equiv s_6$.

We now construct the **dual machine** of our example described above. Recall that $\mathcal{K} = (S, \mathcal{A}, \mathcal{O}, \delta, \gamma)$. The dual of $\mathcal{K}$ is $\mathcal{K}' = (S', \mathcal{A}, \mathcal{O}', \delta', \gamma')$ where:

$$\begin{aligned} S' &= T = \{[t]\} \\ \mathcal{O}' &= S \\ \delta'([t], a) &= [at] \\ \gamma'([t], s) &= [t] \end{aligned}$$

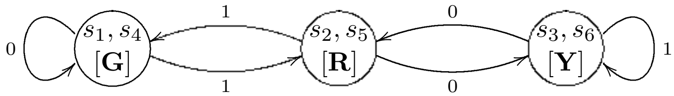
The states of the new machine are the equivalence classes of the original machine, and the new transition function is defined to work with them; the observations are the states of the original machine. What observations do we see? Since the new states of the dual machine are the equivalence classes $[t]$ of the old one, then the observations that we see should be the states that satisfy t; this is just what is given by the new observation function γ'.

We now construct the dual machine. In the original machine, we had determined that we have 3 equivalence classes: $[t_1]$, $[t_2]$ and $[t_3]$. Thus, the new machine has those three states. In addition, the observations are the states of the original machine that satisfy each of the equivalence classes, so transition arrows aside, the dual looks like:

We now add the transition arrows to the dual.

$$\delta([t_1], 0) = [0t_1]$$

We know that $[t_1] = [\mathbf{G}]$, therefore $[0t_1] = [0\mathbf{G}]$. To find out what equivalence class this test belongs to, we have to go through the test backwards. In order to see an observation of **G**, we have to be in either state s_1 or s_4. The action that produced the **Y** observation was 0, so consider what happens if we take a 0 action backwards from s_1: $s_1 \overset{0}{\leftarrow} s_1$ or $s_1 \overset{0}{\leftarrow} s_4$. Note that there is no state s such that $s_4 \overset{0}{\leftarrow} s$. Thus, whatever equivalence class contains $[0\mathbf{G}]$ is identified by the states $\{s_1, s_4\}$. This is just t_1 itself, so we have a 0 transition arrow from $[t_1]$ to itself. Computing the transition arrows in this way we get the dual machine below.



In this special case, the underlying transition graph of the double dual will be isomorphic to the transition graph of the dual but that is just an artefact of this example.

Brzozowski's Algorithm for Minimization. Brzozowski [10] discovered the following intriguing algorithm for minimizing finite state automata viewed as acceptors. Take the transitions and reverse the arrows. In addition interchange the accepting and non-accepting states; the resulting machine is not deterministic, of course. Determinize this machine in the usual way then reverse the result, flip the accepting states and the non accepting states again and determinize again. Remarkably, this gives the minimal deterministic automaton. The reverse operation is exactly our duality construction for the special case of one observation. Brzozowski does not present the duality construction using the logic that we have made explicit, but it is clearly there implicitly. Our presentation is essential for the generalization to the stochastic case. Of course this algorithm can blow up exponentially in the intermediate stage (the construction of the dual).

Despite this it appears to be useful in practice. We should point out that reachability plays a key role in Brzozowski's algorithm so the correspondence is not perfect. A recent paper [8] gives the precise categorical treatment of Brzozowski's algorithm.

4 A Simple Duality for Partially Observable Probabilistic Automata

In this section we work with partially observable probabilistic automata which is a prelude to the treatment of partially observable Markov decision processes (POMDPs). We will develop a duality with deterministic automata with stochastic observations (DASOs) using the probabilistic analogues of the simple formulas of the last two sections. In the AI literature these are called "e-tests" [24], where the e signifies that there is a single observation made at the *end* of a sequence of actions. With these e-tests one does get a pleasant duality theory, but the dual automaton loses much of the information of the original automaton. Nevertheless, this simple duality does capture many aspects of the original behaviour.

We recall the definition of a partially observable probabilistic automaton as $\mathcal{H} = (S, \Sigma, \mathcal{O}, \tau, \gamma)$, where $\tau : S \times \Sigma \times S \longrightarrow [0, 1]$ is the *transition function* and $\gamma : S \times \mathcal{O} \longrightarrow [0, 1]$ is the observation probability function. In our setup the observations are taken from a discrete set. It is not hard to develop the theory with observations taking on, for example, real values, but we will not do that in the present paper because that would involve us in measure-theoretic considerations.

We can define a dual using the same inductive definition for tests as in the deterministic case, but with a probabilistic semantics where a state satisfies a test with a given probability. Thus, the meaning of tests $[\![t]\!]$ should be considered as functions assigning probabilities to states. We will use s to stand for a typical state and ω for a typical observation. The definitions are:

$$[\![\omega]\!]_{\mathcal{H}}(s) = \gamma(s, \omega)$$
$$[\![(a) \cdot t]\!]_{\mathcal{H}}(s) = \sum_{s'} \tau(s, a, s')[\![t]\!]_{\mathcal{H}}(s')$$

We can define an equivalence relation on these tests by $t_1 \sim t_2$ if and only if $[\![t_1]\!] = [\![t_2]\!]$: thus, the equivalence class of t is completely determined by $[\![t]\!]$. We will just use $[\![t]\!]$ rather than the equivalence class $[t]$.

We define the dual $\mathcal{H}' = (S', \Sigma, \mathcal{O}', \tau', \gamma')$ as follows:

- $S' = \{[\![t]\!]_{\mathcal{H}} | t \in \mathcal{F}\}$, where $\mathcal{F}$ is the collection of formulas.
- $\mathcal{O}' = S$
- $\gamma'([\![t]\!]_{\mathcal{H}}, s) = [\![t]\!]_{\mathcal{H}}(s)$
- $\tau([\![t]\!]_{\mathcal{H}}, a, [\![(a) \cdot t]\!]_{\mathcal{H}}) = 1$ (0 otherwise)

Note that the transition function is now completely deterministic: it can be written in the much more perspicuous form $\tau([\![t]\!], a) = [\![(a) \cdot t]\!]$. Thus, the duality construction has made the system deterministic.

We can now consider the double dual, which has $\mathcal{H}'$-equivalence classes of formulas as states. Since the observations of $\mathcal{H}'$ are the states in $\mathcal{H}$, a basic test on $\mathcal{H}'$ would look like $[\![\hat{s}]\!]_{\mathcal{H}'}$ for some $s \in S$, where, as before, we are writing $\hat{s}$ for the state s regarded as an observation of the dual system. We find that

$$\gamma''([\![\hat{s}]\!]_{\mathcal{H}'}, [\![\omega]\!]_{\mathcal{H}}) = [\![\hat{s}]\!]_{\mathcal{H}'}([\![\omega]\!]_{\mathcal{H}}) = \gamma'([\![\omega]\!]_{\mathcal{H}}, \hat{s}) = [\![\omega]\!]_{\mathcal{H}}(s) = \gamma(s, \omega).$$

Tests interpreted on the double dual, applied to states of the double dual that are of the form $[\hat{s}]$ for some state s of the primal machine, define the same functions as they do on when interpreted on the original:

$$\begin{aligned}
[\![a_1 a_2 \cdots a_k \omega]\!]_{\mathcal{H}''}([\hat{s}]_{\mathcal{H}'}) &= [\![\omega]\!]_{\mathcal{H}''}(\delta''([\hat{s}]_{\mathcal{H}'}, a_1 a_2 \cdots a_k)) \\
&= [\![\omega]\!]_{\mathcal{H}''}([a_k \cdots a_2 a_1 \hat{s}]_{\mathcal{H}'}) \\
&= \gamma''([a_k \cdots a_2 a_1 \hat{s}]_{\mathcal{H}'}, [\![\omega]\!]_{\mathcal{H}}) \\
&= [\![a_k \cdots a_2 a_1 \hat{s}]\!]_{\mathcal{H}'}([\![\omega]\!]_{\mathcal{H}}) \\
&= [\![\hat{s}]\!]_{\mathcal{H}'}(\delta'([\omega]_{\mathcal{H}}, a_k \cdots a_2 a_1)) \\
&= \gamma'([\![a_1 a_2 \cdots a_k \omega]\!]_{\mathcal{H}}, \hat{s}) \\
&= [\![a_1 a_2 \cdots a_k \omega]\!]_{\mathcal{H}}(s).
\end{aligned}$$

What we have here is a duality between probabilistic Kripke automata and deterministic automata with probabilistic observations. Once again the duality is mediated by the notion of satisfaction between states and tests and the entire duality theory can be seen as transposing the satisfaction relation.

The formulas that we have considered are very special: observations are made only at the end of a sequence of actions. One can consider tests to be formulas and ask what the effect of adding other logical connectives would be. We will, however, take the view that we are working with "tests" that the system may or may not pass. With this viewpoint it is more natural to consider generalizations that are different from what one would consider by adding more logical connectives to a logic.

As a prelude to the next section, consider what happens with a more general kind of test called an "s-test" [24] in the AI literature. The new feature is that one can make observations after every action. Note, however, that with these more general kinds of tests one does not induce the same functions on the original and double dual. This is because in the double dual the transitions are deterministic so the observations provide no additional information about the state, given the action. A more precise semantics would capture conditional probabilities of a given new state conditioned on the observations made.

5 State Based Duality for POMDPs

In order to obtain a duality theory without a loss of information one needs a more refined notion of experiments, or, equivalently, a richer notion of formulas. The class of s-tests introduced at the end of the last section is not quite the right concept: it does capture all the system dynamics but, it is an ad-hoc class of tests without a nice algebraic structure. We will work with a larger class of tests in which a series of actions can be alternated with an observation. This allows the concatenation of actions with tests, which is essential for duality.

Definition 13. *We define a POMDP as*

$$\mathcal{M} = (S, \mathcal{A}, \mathcal{O}, \delta_{a\in\mathcal{A}} : S \times S \longrightarrow [0,1], \gamma_{a\in\mathcal{A}} : S \times \mathcal{O} \longrightarrow [0,1]).$$

We use the word "tests" almost as before ("e-tests"); we use the word "experiments" in this section for sequences of tests. The formal definitions are as follows.

Definition 14. *A **test** t is a non-empty sequence of actions followed by an observation, i.e. $t = a_1 \cdots a_n\omega$, with $n \geq 1$.*

Definition 15. *An **experiment** is a non-empty sequence of tests $e = t_1 \cdots t_m$ with $m \geq 1$.*

Note that these definitions force one to make an action in order to observe anything. This is a consequence of the way observations are defined; they are associated with actions so we cannot just make an observation in a state. Unlike with POPAs, observations are associated with the *action* and the target state so it makes no sense to regard a simple observation as a test.

In order to proceed with the construction of the dual to a POMDP we extend the definition of the transition function to work on sequences of actions.

Definition 16. *Given a POMDP as in Def. 13 we define a transition function δ_α, where α is a sequence of actions, inductively:*

$$\delta_\epsilon(s,x) = \mathbf{1}_{s=x} \qquad \forall s,x \in S$$
$$\delta_{a\alpha}(s,x) = \sum_{y\in S} \delta_a(s,y)\delta_\alpha(y,x) \qquad \forall s,x \in S.$$

We have written $\mathbf{1}_{s=x}$ for the indicator function viewed as a Kronecker distribution; i.e. $\mathbf{1}_{s=x}$ is 0 unless $s = x$ in which case it is 1.

In order to define the meaning of a state satisfying a test, or an experiment, we need to introduce a ternary symbol, because a test will contain at least one action and thus will cause a transition to a new state. We will define the satisfaction relation between states and tests as a ternary symbol $\langle s|t|q\rangle$ which gives the probability that the system starts in s, is subjected to the test t and ends up in the state q.

Definition 17. *We define $\langle s|t|q\rangle$ by induction on t: $\langle s|a\omega|q\rangle = \delta_a(s,q)\cdot\gamma_a(q,\omega)$ and $\langle s|\alpha a\omega|q\rangle = \sum_r \delta_\alpha(s,r)\langle r|a\omega|q\rangle$.*

We use the same notation for an experiment e: $\langle s|e|q\rangle$ is the probability of the system starting in state s, being subject to the experiment e and ending up in the state q.

It is worth clarifying exactly what it means to say that "a system is subjected to an experiment." If we have an experiment

$$e = a_1^{(1)} \dots a_{n_1}^{(1)} \omega_1 a_1^{(2)} \dots a_{n_2}^{(2)} \omega_2 \dots a_1^{(m)} \dots a_{n_m}^{(1)} \omega_m$$

then the system is subjected to the sequence of actions

$$a_1^{(1)} \dots a_{n_1}^{(1)} a_1^{(2)} \dots a_{n_2}^{(2)} \dots a_1^{(m)} \dots a_{n_m}^{(1)}.$$

The number $\langle s|e|x\rangle$ is the probability that we see the observations $\omega_1\omega_2 \dots \omega_m$ at the appropriate points of the action sequence.

Definition 18. *Given an experiment e the probability $\langle s|e|q\rangle$ is given by the following inductive formula: for a basic experiment $e = t$ the formula is given by Def. 17, for an experiment of the form te we define*

$$\langle s|te|q\rangle = \sum_r \langle s|t|r\rangle\langle r|e|q\rangle.$$

These ternary relations are the fundamental quantities that one can use to carry out the duality constructions. One can define the notion of a state s satisfying test t or experiment e by just summing over the target states. We use the same angle-bracket notation for this.

Definition 19. *We define $\langle s|t\rangle$ to be the probability that a state s satisfies a test t. It is given by the following formula:*

$$\langle s|t\rangle = \sum_q \langle s|t|q\rangle.$$

Similarly $\langle s|e\rangle$ is the probability that s satisfies an experiment e:

$$\langle s|e\rangle = \sum_q \langle s|e|q\rangle.$$

Now we construct the dual and show how to come back. The dual is not a POMDP but a deterministic transition system with stochastic observations.

We proceed as usual by defining an equivalence, this time on experiments; exactly the same definition can be used on tests of course: tests are just simple experiments.

Definition 20. *For experiments e_1, e_2, we say $e_1 \sim_{\mathcal{M}} e_2$ if and only if $\langle s|e_1\rangle = \langle s|e_2\rangle$ for all $s \in S$. Then $[e]_{\mathcal{M}}$ is the $\sim_{\mathcal{M}}$-equivalence class of e.*

The construction of the dual proceeds as before, by making equivalence classes of experiments the states of the dual machine; the states of the primal machine become the observations of the dual machine.

Definition 21. *We define the dual as*

$$\mathcal{M}' = (S', \mathcal{A}, \mathcal{O}', \delta' : S' \times \mathcal{A} \longrightarrow S', \gamma' : S' \times \mathcal{O}' \longrightarrow [0, 1])$$

where

$$\begin{aligned} S' &= \{[e]_{\mathcal{M}}\} \\ \mathcal{O}' &= S \\ \delta'([e]_{\mathcal{M}}, a_0) &= [a_0 e]_{\mathcal{M}} \\ \gamma'([e]_{\mathcal{M}}, s) &= \langle s|e\rangle \end{aligned}$$

As noted before this is a deterministic transition system with stochastic observations.

To get the double-dual we have to use the appropriate construction in the space of the dual machines, i.e. in the space of deterministic transition systems with stochastic observations. This is precisely the simple construction of the last section. Thus, we consider only single-test experiments on the dual (i.e. e-tests), but we allow the action sequences to be empty. Then for a given test $\tau = \alpha s$ of the dual machine[3], where α is a sequence of actions, we have $\langle [t]_{\mathcal{M}} | \tau \rangle = \langle s | \alpha^R t \rangle$, where α^R indicates α with the action order reversed. Equivalence is defined analogously: $\tau_1 \sim_{\mathcal{M}'} \tau_2$ if and only if $\langle x | \alpha_1^R t \rangle = \langle y | \alpha_2^R t \rangle$ for all tests t; where $\tau_1 = \alpha_1 x$ and $\tau_2 = \alpha_2 y$.

We now define the double dual as follows.

Definition 22. *Given a POMDP $\mathcal{M}$ and its dual $\mathcal{M}'$ we construct the double dual $\mathcal{M}'' = (S'', \mathcal{A}', \mathcal{O}'', \delta'', \gamma'')$, which is of the same type as the dual and has the same actions, as follows:*

$$\begin{aligned} S'' &= \{[\tau]_{\mathcal{M}'}\} \\ \mathcal{O}'' &= S' \\ \delta''([\tau]_{\mathcal{M}'}, a_0) &= [a_0 \tau]_{\mathcal{M}} \\ \gamma''([\tau]_{\mathcal{M}'}, [e]_{\mathcal{M}}) &= \langle [e]_{\mathcal{M}} | \tau \rangle = \langle s | \alpha^R e \rangle \qquad (\tau = \alpha s) \end{aligned}$$

We now show that everything is well-defined for the transition functions, which follows more or less immediately from the definitions for observation functions. Note that the ternary symbol is not needed in the actual definition of the dual but it is necessary for the proof that the transition function is well-defined. Essentially the transitions in the dual are given by e goes under an a-action to ae; for this to make sense we need to show that it does not matter which representative of the equivalence class of e is chosen. This is what the next lemma shows.

Lemma 23. *If $e_1 \sim_{\mathcal{M}} e_2$ then $a_0 e_1 \sim_{\mathcal{M}} a_0 e_2$.*

[3] We will use the Greek letter τ for tests of the dual machine.

Proof. $e_1 \sim_{\mathcal{M}} e_2$, so $\langle s|e_1\rangle = \langle s|e_2\rangle$ for all $s \in S$. For $i = 1, 2$ let $e_i = t_1^{(i)} \cdots t_{m_i}^{(i)}$ and $t_1^{(i)} = \alpha^{(i)}\omega^{(i)} = a_1^{(i)} \cdots a_{n_i}^{(i)}\omega^{(i)}$. Then for any state s,

$$\begin{aligned}
\langle s|a_0e_1\rangle &= \sum_{x,y\in S} \langle s|a_0t_1^{(1)}|y\rangle\langle y|t_2^{(1)} \cdots t_{m_1}^{(1)}|x\rangle \\
&= \sum_{x,y\in S} \delta_{a_0\alpha^{(1)}}(s,y)\gamma_{a_{n_1}^{(1)}}(y,\omega^{(1)})\langle y|t_2^{(1)} \cdots t_{m_1}^{(1)}|x\rangle \\
&= \sum_{x,y,z\in S} \delta_{a_0}(s,z)\delta_{\alpha^{(1)}}(z,y)\gamma_{a_{n_1}^{(1)}}(y,\omega^{(1)})\langle y|t_2^{(1)} \cdots t_{m_1}^{(1)}|x\rangle \\
&= \sum_{z\in S} \delta_{a_0}(s,z) \sum_{x,y\in S} \langle z|t_1^{(1)}|y\rangle\langle y|t_2^{(1)} \cdots t_{m_1}^{(1)}|x\rangle \\
&= \sum_{z\in S} \delta_{a_0}(s,z)\langle z|e_1\rangle = \sum_{z\in S} \delta_{a_0}(s,z)\langle z|e_2\rangle = \langle s|a_0e_2\rangle.
\end{aligned}$$

∎

Similarly, for the double dual we have the following lemma.

Lemma 24. *If $\tau_1 \sim_{\mathcal{M}'} \tau_2$ then $a_0\tau_1 \sim_{\mathcal{M}'} a_0\tau_2$ for any action a_0.*

Proof. Let $\tau_1 = \alpha_1 x$ and $\tau_2 = \alpha_2 y$ and assume that $\tau_1 \sim_{\mathcal{M}'} \tau_2$, so $\langle x|\alpha_1^R e\rangle = \langle y|\alpha_2^R e\rangle$ for all experiments e. Then for any experiment e,

$$\langle x|(a_0\alpha_1)^R e\rangle = \langle x|\alpha_1^R a_0 e\rangle = \langle x|\alpha_1^R(a_0e)\rangle = \langle y|\alpha_2^R(a_0e)\rangle.$$

∎

Now that we know that these constructions are well defined the duality is captured by the following theorem.

Theorem 25. *The probability of a state s in the primal satisfying an experiment e, i.e. $\langle s|e\rangle$, is given by $\langle [s]_{\mathcal{M}'}|[e]_{\mathcal{M}}\rangle = \gamma''([s]_{\mathcal{M}'}|[e]_{\mathcal{M}}\rangle$, where $[s]_{\mathcal{M}'}$ indicates the equivalence class of the e-test on the dual which has s as an observation and an empty sequence of actions.*

Proof. Note that $[s]_{\mathcal{M}'}$ is an equivalence class of states of tests of the dual, hence a state of the double dual. Recall that the dual is a DASO so we are using the simple duality here. Note further that $[e]_{\mathcal{M}}$ is an equivalence class of experiments on the primal, which is a state of the dual and hence an observation of the double dual. So, by definition of the angle bracket notation this is just $\gamma''([s]_{\mathcal{M}'}|[e]_{\mathcal{M}}\rangle$. By the definition of the double dual construction we have

$$\gamma''([\tau]_{\mathcal{M}'}|[e]_{\mathcal{M}}\rangle = \langle s|\alpha^R e\rangle,$$

where $\tau = \alpha s$. In our case α is the empty sequence and τ is just s so we get $\gamma''([s]_{\mathcal{M}'}|[e]_{\mathcal{M}}\rangle = \langle s|e\rangle$. ∎

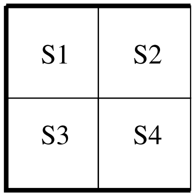


Fig. 1. The navigation domain

Thus, the results of all experiments on the primal can be read off the double dual.

We present two examples illustrating the constructions of this section. The first is taken from [17]. It describes a simple navigation domain as shown in Fig. 1.

The squares represent places where a robot could be. The heavy lines represent walls that cannot be crossed; the walls are painted blue. The robot can take the following actions: `N`,`S`,`E` and `W`. If it is already at the left end, say in square S_1, and attempts to move west (left) it will just stay where it is. It can also make a observation of the colour of its immediate surroundings. If it is in either of the squares S_1 or S_3 it will see blue with probability 1; in the squares S_2 and S_4 it will see red *or* blue each with probability 0.5; the red reading represents a curtain that it could see on the right corresponding to the lightly marked lines on the extreme right edge of the picture. This is an example with deterministic moves but noisy readings. The observations are state based in this case, but we will present it in the POMDP framework.

The system may be represented by the automaton shown in Fig. 2

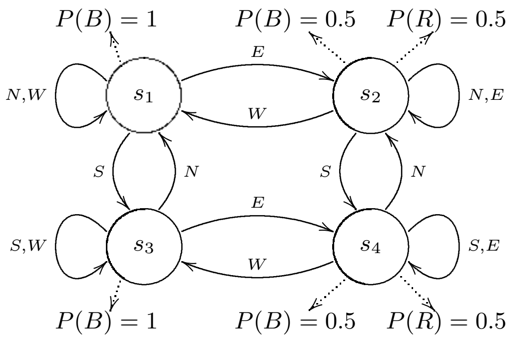


Fig. 2. The navigation domain as a transition system

We can calculate a part of the dual as shown in Fig. 3. The whole dual is of course infinite and one cannot write it down explicitly. Here the states are

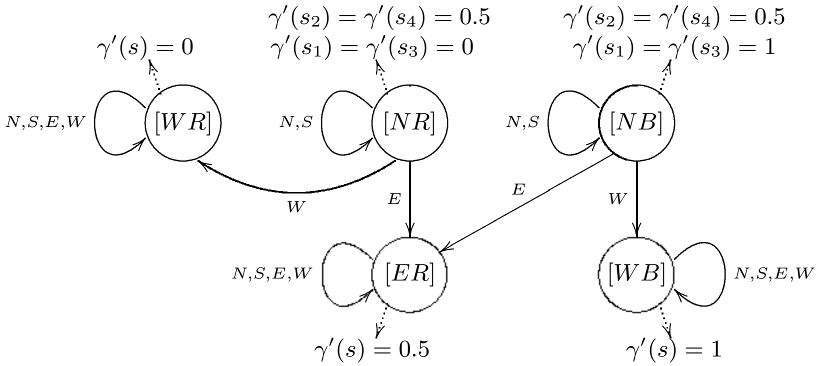


Fig. 3. Part of the dual of the navigation domain

labelled by (equivalence classes of) experiments on the original system, but the transitions are labelled by the same actions as in the primal. The old states are now the observations of the dual. We write $\gamma(s) = 0$ as short for $\gamma(s_i) = 0$ for $i = 1, 2, 3, 4$.

The double dual, shown in Fig. 4, will of course collapse it to a minimal representation. Since the navigation domain is up-down symmetric the collapsed version will just have two states as shown below. This version will completely predict all possible experimental outcomes of the original system.

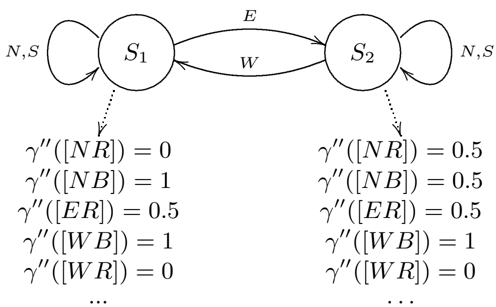


Fig. 4. The double dual of the navigation domain

Our second example shows a situation where the observations depend on the action as well as the target state. This can also be viewed as a four-square navigation domain. There are two observations o_1 and o_2: exactly one of these observations will be made as a transition is taken so we need only specify the probability of one of them. The actions are N, S, E, W which stand for north, south, east and west respectively. This is also a very symmetric domain with top down symmetry.

The primal is shown in Fig. 5.

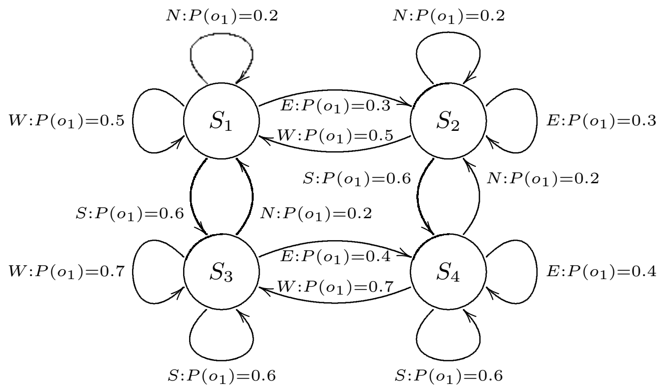


Fig. 5. A POMDP with action dependent observations

The dual is infinite, of course; in Fig. 6 we show a fragment of the dual. We have shown all the states corresponding to tests, i.e. experiments with a single observation and we have suppressed the observations of the dual.

In the table shown in Fig. 7 we display some of the values of γ' for two-observation experiments and all of the values for tests. The first line says: given that that the dual system is in state $[NO_1]$ the observation function for all the s_i is the same and has the value 0.2. Note that, in fact, that for any sequence of N actions followed by O_1 these numbers are the same so all tests of the form N^kO_1 are equivalent.

The double dual, shown in Fig. 8 has two states.

6 Related Work

Rivest and Schapire [33] present an approach to inferring the structure of a finite-state automaton from its input-output behavior, by running "experiments" on the automaton. They rely on an "update graph", which is essentially the dual in our representation, and on e-tests, of the form $a_1 \ldots a_n o$. They also present experiments in which they infer an automaton based on Rubik's cube using this structure. Their work is limited to deterministic automata. Nevertheless, the fact that they could deal with a system with 10^{19} states is very impressive. Their work shows that a very large system can have a much more compact dual.

More recently, predictive state representations [24, 34] (PSR), introduced in the AI community, generalized the work of Rivest and Schapire to the case of stochastic automata. The representation is based on the prediction of s-tests, which are of the form $a_1o_1 \ldots a_no_n$. In the work of Littman et al. [24], each state in a POMDP is viewed as represented by an infinite set of predictions, for

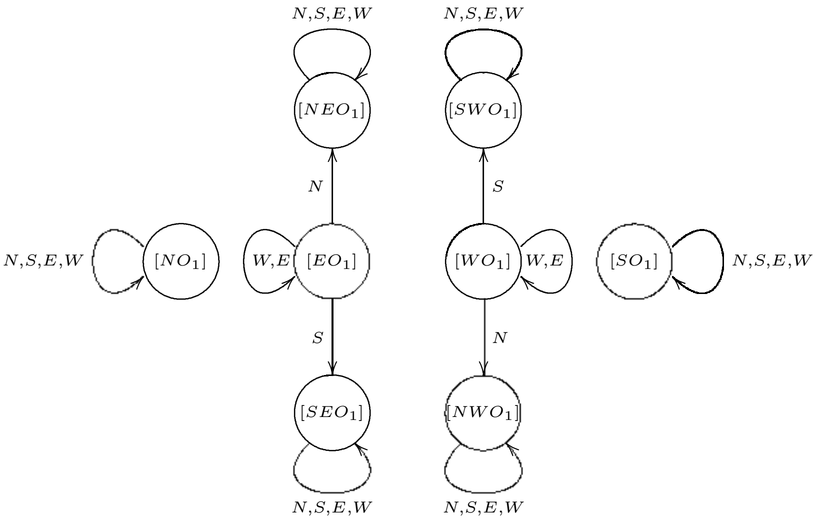
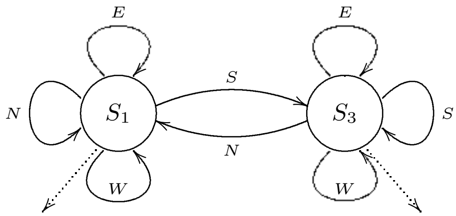


Fig. 6. Part of the dual of the POMDP in Fig. 5

State $[t]$	$\gamma'([t], s_1)$	$\gamma'([t], s_2)$	$\gamma'([t], s_3)$	$\gamma'([t], s_4)$
$[NO_1]$	0.2	0.2	0.2	0.2
$[SO_1]$	0.6	0.6	0.6	0.6
$[EO_1]$	0.3	0.3	0.4	0.4
$[WO_1]$	0.5	0.5	0.7	0.7
$[NEO_1]$	0.3	0.3	0.3	0.3
$[SWO_1]$	0.7	0.7	0.7	0.7
$[SEO_1]$	0.4	0.4	0.4	0.4
$[NWO_1]$	0.5	0.5	0.5	0.5
$[NO_1NO_1]$	0.04	0.04	0.04	0.04
$[SO_1SO_1]$	0.36	0.36	0.36	0.36
$[SO_1NO_1]$	0.12	0.12	0.12	0.12
$[NO_1EO_1]$	0.06	0.06	0.06	0.06
$[NO_1WO_1]$	0.1	0.1	0.1	0.1
$[EO_1EO_1]$	0.09	0.09	0.16	0.16
$[WO_1WO_1]$	0.25	0.25	0.49	0.49
...	...	...	...	...

Fig. 7. Table of observation probabilities in the dual



$$\gamma''([NO_1]) = 0.2 \qquad \gamma''([NO_1]) = 0.2$$
$$\gamma''([SO_1]) = 0.6 \qquad \gamma''([SO_1]) = 0.6$$
$$\gamma''([EO_1]) = 0.3 \qquad \gamma''([EO_1]) = 0.4$$
$$\gamma''([WO_1]) = 0.5 \qquad \gamma''([WO_1]) = 0.7$$
$$\ldots \qquad \ldots$$

Fig. 8. The double dual

each possible test. However, there exists a finite number of linearly independent tests whose predictions are sufficient to compute the prediction for any other tests. Moreover, the number of such "core tests" is at most equal to the number of states in the POMDP. Singh et al. [34] then generalized this approach by considering predictions of s-tests based on histories. They also showed how different frameworks can be represented with a finite number of "core tests." From the point of view presented in this paper, we can view the PSRs as a way of representing the dual or the double dual of a system using the set of linearly independent columns. This view has the advantage of a finite representation, e.g. if the original system is a POMDP. However, this representation does not lend itself easily to approximations. It is our hope that by working directly with the dual or the double-dual, one can develop more easily a theory of approximation for such system.

It seems likely that the best setting to understand the categorical context is that of Chu spaces [5], at least for the deterministic automaton case. Our quotient construction on automata is closely related to the process of forming separated extensional Chu spaces [6]. There is also some similarity to the work of Pratt [31] though there the duality is between states and trajectories (which he calls "schedules").

There are several discussions in the literature on duality in systems theory; see, for example, the excellent paper by Bainbridge [4] and the several very interesting chapters in the book by Kalman, Farb and Arbib [21]. In systems theory one is concerned with *controlling* a system to obtain a desired behaviour. As with our POMDPs, the systems are partially observable; one does not see the state. One only has a *readout map* that maps the states into observables. The "fundamental duality" in this subject is the duality between controllability: the ability to steer a system into a known state, and observability: the ability to determine the state after a series of observations. There has been a rich categorical treatment of this subject: for example, there are several papers by Arbib and Manes on this topic [1–3]. These are largely concerned with the nonprobabilistic situation;

the papers by Arbib and Manes hint at the probabilistic case but we have never seen this spelled out satisfactorily.

There is a whole plethora of dualities in mathematics. The Stone-type dualities establish a duality between logics and transition systems. This has appeared in denotational semantics and is due to Plotkin and Smyth [36]. They establish a relationship between "forward" state-transformer semantics and "backwards" predicate-transformer semantics. Kozen [22] established similar results in the case of probabilistic programs and probabilistic dynamic logic. More recently, there has been very interesting work by Mislove, Ouaknine, Pavlovic and Worrell [28] and by Pavlovic, Mislove and Worrell [13] on duality for labelled Markov processes (LMPs). These are like POMDPs but they do not have the notion of observation. Other research on duality for logics and for transitions systems include the work by Bidoit, Hennicker and Kurz [7] and Bonsangue and Kurz [9].

7 Conclusions

We have shown that, in some informal sense, there is a duality between state and observation, or, more precisely, between state and experiment. In this view a state is an equivalence class of experimental data. It frees us from having to work with arbitrary preconceived notions of state. This view, we hope, will unify many ideas that are currently being investigated for representing systems with hidden state.

It is important to clarify what we mean by minimal. Of course the "minimal" gadget that we construct is larger than the POMDP that one starts with; the latter is finite and the former is infinite. But, among the class of deterministic systems with stochastic observations that represent the same behaviour it is minimal in the following sense. Minimal means that it is the "most highly quotiented version possible." More precisely, suppose that we have a system S of some kind. One can "reduce its size" by defining an appropriate equivalence relation $\sim$ and constructing $S/\sim$. We can say that a system S' is a *minimal realization* of S if it is behaviourally equivalent and if it is the quotient of S by some equivalence relation *and if there is any other system* X which is also a quotient of S then X can be *further quotiented* to yield S'. In the companion categorical paper this is formalized as a couniversal property and comes out naturally. For finite-state systems this is indeed the same as having as few states as possible.

There is much to be investigated from the algorithmic point of view. Perhaps the most pressing issue is a pleasant approximation theory. The dual and the double-dual are both based on exact equivalences. This raises the possibility of working with metric notions [15] and constructing more compact representations based on identifying "nearby" experiments.

Finally, everything has been worked out here for discrete systems. Clearly, for realistic applications one would need to extend the theory to continuous state spaces, and, even more importantly, to continuous observations.

Acknowledgements. Prakash Panangaden has benefitted from discussions with Marcello Bonsangue, Alexander Kurz, Jan Rutten, Alexandra Silva, Ben Worrell and especially Nick Bezhanishvili and Clemens Kupke. He would also like to thank the organizers of the Tbilisi meetings for the invitation to lecture on this at Kutaisi and to enjoy Georgian hospitality.

This research has been supported by a grant from the Office of Naval Research and NSERC. Panangaden's sabbatical visit to the UK, which allowed him to collaborate with Bezhanishvili and Kupke, in 2010-11 was supported by an EPSRC grant to Oxford University. Chris Hundt's summer research was funded by a summer research grant from NSERC.

References

1. Arbib, M.A., Manes, E.G.: Machines in a category: An expository introduction. SIAM Review 16, 163–192 (1974)
2. Arbib, M.A., Manes, E.G.: Foundations of system theory: decomposable systems. Automatica 10, 285–302 (1974)
3. Arbib, M.A., Manes, E.G.: Adjoint machines, state behavior machines and duality. J. Pure Appl. Algebra 6, 313–343 (1975)
4. Bainbridge, E.S.: The fundamental duality of system theory. In: Hartnett, W.E. (ed.) Systems: Approaches, Theories, Applications, pp. 45–61. Reidel (1977)
5. Barr, M.: The chu construction. Theory Appl. Categories 2, 17–35 (1996)
6. Barr, M.: The separated extensional Chu category. Theory Appl. Categories 4(6), 137–147 (1998)
7. Bidoit, M., Hennicker, R., Kurz, A.: Observational logic, constructor-based logic, and their duality. Theor. Comput. Sci. 3(298), 471–510 (2003)
8. Bonchi, F., Bonsangue, M.M., Rutten, J.J.M.M., Silva, A.: Brzozowski's Algorithm (Co)Algebraically. In: Constable, R.L., Silva, A. (eds.) Logic and Program Semantics, Kozen Festschrift. LNCS, vol. 7230, pp. 12–23. Springer, Heidelberg (2012)
9. Bonsangue, M.M., Kurz, A.: Duality for Logics of Transition Systems. In: Sassone, V. (ed.) FOSSACS 2005. LNCS, vol. 3441, pp. 455–469. Springer, Heidelberg (2005)
10. Brzozowski, J.A.: Canonical regular expressions and minimal state graphs for definite events. In: Fox, J. (ed.) Proceedings of the Symposium on Mathematical Theory of Automata. MRI Symposia Series, vol. 12, pp. 529–561. Polytechnic Press of the Polytechnic Institute of Brooklyn (April 1962), book appeared in 1963
11. Chrisman, L.: Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In: Proceedings of the Tenth National Conference on Artificial Intelligence, pp. 183–188 (1992)
12. Clarke, E.M., Grumberg, O., Peled, D.A.: Model Checking. The MIT Press, Cambridge (1999)
13. Pavlovic, D., Mislove, M., Worrell, J.B.: Testing Semantics: Connecting Processes and Process Logics. In: Johnson, M., Vene, V. (eds.) AMAST 2006. LNCS, vol. 4019, pp. 308–322. Springer, Heidelberg (2006)
14. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labeled Markov processes. Information and Computation 179(2), 163–193 (2002)
15. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: A metric for labelled Markov processes. Theoretical Computer Science 318(3), 323–354 (2004)

16. Hopcroft, J.E., Motwani, R., Ullman, J.D.: Introduction to automata theory, languages, and computation, 2nd edn. Addison-Wesley Series in Computer Science. Addison-Wesley-Longman (2001)
17. Hundt, C., Panangaden, P., Pineau, J., Precup, D.: Representing systems with hidden state. In: The Twenty-First National Conference on Artificial Intelligence, AAAI (2006)
18. James, M., Singh, S.: Learning and discovery of predictive state representations in dynamical systems with reset. In: International Conference on Machine Learning, vol. 21, pp. 417–424 (2004)
19. Johnstone, P.: Stone Spaces, Cambridge Studies in Advanced Mathematics, vol. 3. Cambridge University Press (1982)
20. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial Intelligence 101 (1998)
21. Kalman, R.E., Falb, P.L., Arbib, M.A.: Topics in Mathematical Systems Theory. McGraw Hill (1969)
22. Kozen, D.: A probabilistic PDL. Journal of Computer and Systems Sciences 30(2), 162–178 (1985)
23. Kozen, D.: Automata and computability. Undergraduate texts in computer science. Springer (1997)
24. Littman, M., Sutton, R., Singh, S.: Predictive representations of state. In: Advances in Neural Information Processing Systems 14 (NIPS), vol. 14, pp. 1555–1561 (2002)
25. McCallum, A.: Reinforcement Learning with Selective Perception and Hidden State. Ph.D. thesis, University of Rochester (1995)
26. Milner, R.: A Calculus of Communication Systems. LNCS, vol. 92. Springer, Heidelberg (1980)
27. Milner, R.: Communication and Concurrency. Prentice-Hall (1989)
28. Mislove, M., Ouaknine, J., Pavlovic, D., Worrell, J.: Duality for Labelled Markov Processes. In: Walukiewicz, I. (ed.) FOSSACS 2004. LNCS, vol. 2987, pp. 393–407. Springer, Heidelberg (2004)
29. Park, D.: Concurrency and Automata on Infinite Sequences. In: Deussen, P. (ed.) GI-TCS 1981. LNCS, vol. 104, pp. 167–183. Springer, Heidelberg (1981)
30. Park, D.: Title unknown (1981), slides for Bad Honnef Workshop on Semantics of Concurrency
31. Pratt, V.R.: The Duality of Time and Information. In: Cleaveland, W.R. (ed.) CONCUR 1992. LNCS, vol. 630, pp. 237–253. Springer, Heidelberg (1992)
32. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley (1994)
33. Rivest, R.L., Schapire, R.E.: Diversity-based inference of finite automata. Journal of the ACM 41(3), 555–589 (1994)
34. Singh, S., James, M.R., Rudary, M.R.: Predictive state representations: A new theory for modeling dynamical systems. In: Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence (UAI), pp. 512–519 (2004)
35. Sipser, M.: Introduction to the theory of computation. PWS Publishing Company (1997)
36. Smyth, M.: Powerdomains and Predicate Transformers. In: Díaz, J. (ed.) ICALP 1983. LNCS, vol. 154, pp. 662–676. Springer, Heidelberg (1983)
37. Sutton, R.S., Tanner, B.: Temporal-difference networks. In: Advances in Neural Information Processing Systems, vol. 17 (2005)

Model Checking for Modal Intuitionistic Dependence Logic*

Johannes Ebbing[1], Peter Lohmann[1], and Fan Yang[2]

[1] Leibniz University Hannover, Theoretical Computer Science,
Appelstr. 4, 30167 Hannover, Germany
{ebbing,lohmann}@thi.uni-hannover.de
[2] Department of Mathematics and Statistics,
P.O. Box 68 (Gustaf Hällströmin katu 2b),
FIN-00014 University of Helsinki, Finland
fan.yang@helsinki.fi

Abstract. Modal intuitionistic dependence logic ($\mathcal{MIDL}$) incorporates the notion of "dependence" between propositions into the usual modal logic and has connectives which correspond to intuitionistic connectives in a certain sense. It is the modal version of a variant of first-order dependence logic (Väänänen 2007) considered by Abramsky and Väänänen (2009) basing on Hodges' *team semantics* (1997).

In this paper, we study the computational complexity of the model checking problem for $\mathcal{MIDL}$ and its fragments built by restricting the operators allowed in the logics. In particular, we show that the model checking problem for $\mathcal{MIDL}$ in general is PSPACE-complete and that for propositional intuitionistic dependence logic is coNP-complete.

ACM Subject Classifiers: F.2.2 Complexity of proof procedures, F.4.1 Modal logic, D.2.4 Model checking.

Keywords: dependence logic, intuitionistic logic, modal logic, model checking, computational complexity.

1 Introduction

We investigate the computational complexity of model checking problem for modal intuitionistic dependence logic ($\mathcal{MIDL}$). $\mathcal{MIDL}$ incorporates the notion of "dependence" between propositions into the usual modal logic and has connectives which correspond to intuitionistic connectives in a certain sense. Such dependence is a kind of property which is possessed by sets in a Kripke model, model checking problem for $\mathcal{MIDL}$ checks whether a given set in a Kripke model has a modal property which involves implications of dependence. In this paper, we analyze the complexity of model checking problem for fragments of $\mathcal{MIDL}$

* This work was supported by DAAD grant 50740539 and by grant 138163 of the Academy of Finland. It was also partially supported by the EUROCORES LogICCC LINT programme and the NTH Focused Research School for IT Ecosystems.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 231–256, 2013.

defined by restricting the set of operators allowed in the logic. In particular, we show that the model checking problem for $\mathcal{MIDL}$ in general is PSPACE-complete and that for propositional intuitionistic dependence logic is coNP-complete.

The central notion of the logic $\mathcal{MIDL}$ is "dependence", which often occurs in reality together with the notion "independence". Dependence and independence are common phenomena in many fields: from computer science (databases, software engineering, knowledge representation, AI) to social sciences (human history, stock markets). The research of developing an appropriate logical formalism for dependence and independence has been active in recent years in the area of "Logic for Interaction". The first step in this direction dates back to 1960's, when Henkin [Hen61] characterized dependence between variables by extending first-order logic with partially ordered quantifiers. In the second step, Hintikka and Sandu [HS89], [HS96] developed a logic of imperfect information with slashed linear quantifiers, called independence-friendly logic (IF-logic). A compositional semantics for IF-logic, *team semantics*, was given by Hodges [Hod97a], [Hod97b]. The satisfaction relation of team semantics is defined with respect to sets of assignments, called *teams*, instead of single assignments as in the first-order logic case. Based on team semantics, Väänänen [Vää07] introduced first-order dependence logic ($\mathcal{D}$). The logic $\mathcal{D}$ adds into first-order logic a new type of atomic formula $\text{dep}(t_1, \ldots, t_n)$ which describes *functional dependence* of a first-order term t_n on terms $t_1, \ldots, t_{n-1}$, namely, $\mathcal{D}$ formulae are built from first-order and dependence atoms using quantifiers ($\forall x$, $\exists x$) and propositional connectives ($\neg$, $\wedge$, $\vee$). By a method of Enderton [End70] and Walkoe [Wal70], it can be shown that sentences of $\mathcal{D}$ have exactly the same expressive power as sentences of the existential second-order logic. Recent research by Abramsky and Väänänen [AV09] showed that in a more general context of Hodges' construction, the algebraic counterpart of a generalized $\mathcal{D}$ logic (called *BID-logic*) is both a commutative quantale (which carries an interpretation of linear logic) and a complete Heyting algebra (which carries an interpretation of intuitionistic logic). New connectives corresponding to the operations in such an algebraic structure are then introduced into BID-logic, namely, the *intuitionistic implication* $\rightarrow$, the *Boolean disjunction* $\varovee$, as well as the *linear implication* $\multimap$. It was observed in [Yan12] that the constant $\bot$ and connectives $\wedge$, $\varovee$ and $\rightarrow$ satisfy the axioms of Maksimova's logic ([Mak86]), which is a well-known intermediate logic. More surprisingly, the fragment of BID-logic with quantifiers, $\wedge$, $\varovee$, $\rightarrow$, $\bot$ and dependence atoms, called *first-order intuitionistic dependence logic*[1] ($\mathcal{IDL}$), has the same expressive power as the full second-order logic, on the level of sentences ([Yan10]).

The underlying propositional logic of first-order $\mathcal{D}$ and first-order $\mathcal{IDL}$ are called *propositional dependence logic* ($\mathcal{PDL}$) (c.f. [PV05]) and *propositional intuitionistic dependence logic* ($\mathcal{PIDL}$), respectively. The logic $\mathcal{PIDL}$ turns out to be essentially equivalent to *inquisitive logic* defined in [CR11], which is a new logic modeling the exchange of information between intelligent agents (see [Yan12] for

[1] The logic in question is such named only because syntactically it contains the intuitionistic implication, but proof-theoretically, first-order intuitionistic dependence logic is *stronger* than first-order (classical) dependence logic.

further discussions). Both of $\mathcal{PDL}$ and $\mathcal{PIDL}$ have dependence atoms of the form $\text{dep}(p_1,\ldots,p_n,q)$, which have the intuitive meaning that the truth value of the propositional variable q is functionally determined by those of the propositional variables $p_1,\ldots,p_n$. Such dependence does not manifest itself in a single world, play, event or observation, therefore in the semantics, we evaluate the formula $\text{dep}(p_1,\ldots,p_n,q)$ in sets T of assignments (called *teams*). Team T satisfying $\text{dep}(p_1,\ldots,p_n,q)$ means that there is a Boolean function $f:\{0,1\}^n \to \{0,1\}$ such that for every assignment σ in T, $\sigma(q) = f(\sigma(p_1),\ldots,\sigma(p_n))$.

Introducing dependence atoms of the form $\text{dep}(p_1,\ldots,p_n)$ into modal logic ($\mathcal{ML}$), Väänänen [Vää08] defined modal dependence logic ($\mathcal{MDL}$). In contrast to $\mathcal{ML}$, the semantic game of $\mathcal{MDL}$ is a game of imperfect information (see [Vää08]). A compositional semantics of $\mathcal{MDL}$, *team semantics*, can be obtained by generalizing the ideas of [Hod97a], [Hod97b]. The satisfaction relation of $\mathcal{MDL}$ is defined with respect to sets of states of Kripke models, called *teams*, instead of single states as in the $\mathcal{ML}$ case. With respect to expressive power, $\mathcal{MDL}$ is a conservative extension of $\mathcal{ML}$, as the team semantics of the fragment of $\mathcal{MDL}$ without dependence atoms agrees with the usual Kripke semantics of $\mathcal{ML}$. Furthermore, restricted to singleton teams, $\mathcal{MDL}$ can be translated back into $\mathcal{ML}$ ([Sev09]).

For example, the following typical $\mathcal{MDL}$ formula

$$\Box\Diamond\text{dep}(p,q)$$

is said to be satisfied on a team T in a Kripke model if every successor of every state in T has access to some states in such a way that in the set of all these states the value of the propositional variable q is functionally determined by that of p. The following is a practical statement that can be expressed by the above $\mathcal{MIDL}$ formula:

However the environment will be degraded in the next 100 years, it is possible that in 200 years from now, whether the earth will be destroyed depends only on whether there is another planet that crashes into the earth.

As the example suggested, in practice, particularly interesting properties involving dependence are often supposed to be identified from a large amount of data (e.g. a model of the future of the earth). On the other hand, the nature of team semantics gives the logic $\mathcal{MDL}$ more complexity, as the successor search for formulae evaluation has to be done for sets of states. Therefore, the computational aspects of the logic $\mathcal{MDL}$ (and its extensions) deserve studying. Sevenster [Sev09] showed that the satisfiability problem for $\mathcal{MDL}$ is NEXPTIME-complete, and a complete classification of the computational complexity of satisfiability problem for all restrictions of propositional and dependence operators of $\mathcal{MDL}$ is given in [LV10]. In [EL11] the computational complexity of model checking for $\mathcal{MDL}$ and some restrictions of it, e.g. $\mathcal{PDL}$, are proven to be NP-complete.

In this paper, we study the computational complexity of model checking for a natural extension of modal dependence logic, called *intuitionistic modal* *dependence logic*[2] ($\mathcal{MIDL}$), which is obtained by adding intuitionistic implication $\rightarrow$ and Boolean disjunction $\oslash$ into $\mathcal{MDL}$. In the logic $\mathcal{MIDL}$, the connective $\vee$ (called *split disjuction*) can be eliminated ([Yan12]), therefore the underlying propositional logic of $\mathcal{MIDL}$ is essentially $\mathcal{PIDL}$ (which is essentially equivalent to inquisitive logic). Moreover, the team semantics of $\mathcal{MIDL}$ over the usual modal Kripke models K coincides with the Kripke semantics of intuitionistic modal logic defined by Fischer Servi [Ser81] over bi-relation Kripke models with domains consisting of all non-empty subsets of the domains of K (see [Yan12] for detailed discussions).

The model checking problem for $\mathcal{MIDL}$ asks that for a given team T in a given Kripke model K, whether a given $\mathcal{MIDL}$ formula ϕ is satisfied on T. A typical $\mathcal{MIDL}$ formula ϕ describes a modal property which involves implications of dependence statements. By doing model checking, one verifies whether such a property is possessed by a certain set in a Kripke model. For example, the intuitive meaning of the following typical $\mathcal{MIDL}$ formula

$$\Box\Diamond(\mathrm{dep}(p_0, q_0) \rightarrow \mathrm{dep}(p_1, q_1))$$

being satisfied on a team T in a Kripke model is that every successor of every state in T has access to some states in such a way that for the set T' of all these states, every subset of T' which satisfies $\mathrm{dep}(p_0, q_0)$ also satisfies $\mathrm{dep}(p_1, q_1)$. Practical statements expressed by the above $\mathcal{MIDL}$ formulae can be found in many areas. The following is an example of such a statement:

However the environment will be degraded in the next 100 years, it is possible that in 200 years from now, if whether the earth will be destroyed depends only on whether there is another planet that crashes into the earth, then whether the human being will migrate to other planets depends only on whether the crash will occur. (∗)

Following [LV10] and [EL11], we will systematically analyze the complexity of model checking problem for fragments of $\mathcal{MIDL}$ defined by restricting the set of modal operators ($\Diamond$, $\Box$) and propositional operators ($\neg$, $\wedge$, $\vee$, $\oslash$, $\rightarrow$) allowed in the logics. The method of systematically classifying the complexity of logic related problems by restricting the set of operators allowed in formulae was used by Lewis [Lew79] for the satisfiability problem of propositional logic, by Hemaspaandra et al. [Hem05] [HSS10] for the satisfiability problem of modal logic, and by others. The motivation for this approach is twofold: theoretically, this systematic approach may lead to insights into the sources of hardness, i.e., the exact components of the logic that make satisfiability, model checking and other problems hard; practically, by systematically examining all fragments of a logic, one might find useful fragments of the logic in practice with both efficient algorithms and high expressivity.

The detailed complexity results obtained in the paper are listed in Table 1 in the last section.

[2] Here we also use the word "intuitionistic" literally only.

2 Preliminaries

In this section, we define the relevant logics and computational problems.

2.1 Modal Intuitionistic Dependence Logic

The syntax of $\mathcal{MIDL}$ is an extension of that of $\mathcal{ML}$.

Definition 1 (Syntax of $\mathcal{MIDL}$)
Let $p, p_1, \ldots, p_n$ be arbitrary propositional variables. Well-formed formulae (in negation normal form*) of the logic $\mathcal{MIDL}$ are defined by the following grammar*

$$\begin{aligned} \varphi ::= p \;\mid\; \neg p \;\mid\; \mathrm{dep}(p_1, \ldots, p_n) \;\mid \neg\mathrm{dep}(p_1, \ldots, p_n) \;\mid \\ \varphi \wedge \varphi \;\mid\; \varphi \vee \varphi \;\mid\; \varphi \mathbin{\otimes} \varphi \;\mid\; \varphi \rightarrow \varphi \;\mid\; \Box\varphi \;\mid\; \Diamond\varphi \end{aligned}$$

Propositional variables p are also called *propositional atoms*, and formulae of the form $\mathrm{dep}(p_1, \ldots, p_n)$ are called *dependence atoms*. The connectives $\vee$, $\otimes$ and $\rightarrow$ are called *split disjunction*, *Boolean disjunction* and *intuitionistic implication*, respectively. The negation symbol $\neg$ applies only to atoms, namely, all $\mathcal{MIDL}$ formulae are in *negation normal form*. In $\mathcal{MIDL}$, we do not have a classical negation, therefore, unlike in $\mathcal{ML}$, the modalities $\Box$ and $\Diamond$ are not dual to each other, as for instance, the expression $\neg\Diamond\neg\varphi$ is not even a well-formed formula of $\mathcal{MIDL}$.

In this paper, for technical simplicity, we use the expression $\mathrm{dep}(\cdot)$ to stand for a special type of *operator* which acts on propositional variables. Any dependence atom $\mathrm{dep}(p_1, \ldots, p_n)$ ($n \in \mathbb{N}$) is a result of an application of $\mathrm{dep}(\cdot)$. Moreover, all of the atomic negation $\neg$, the connectives $\wedge$, $\vee$, $\otimes$, $\rightarrow$ and the modalities $\Box$, $\Diamond$ are viewed as *operators* as well. Formulae of $\mathcal{MIDL}$ are built from propositional variables and operators. Using these terminologies, we define the syntax of the main sublogics of $\mathcal{MIDL}$ considered in this paper as follows.

Definition 2 ($\mathcal{MDL}$, $\mathcal{PDL}$, $\mathcal{PIDL}$)
Let $\mathcal{L}$ be a sublogic of $\mathcal{MIDL}$ and M a subset of the set of operators occurring in $\mathcal{L}$. Then $\mathcal{L}(M)$ is the sublogic of $\mathcal{L}$ built from propositional variables using operators only from M. We sometimes write $\mathcal{L}(op1, op2, \ldots)$ instead of $\mathcal{L}(\{op1, op2, \ldots\})$.

We define the following important sublogics of $\mathcal{MIDL}$ (see also Figure 1):

- *Modal dependence logic ($\mathcal{MDL}$) $:= \mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \vee, \Box, \Diamond)$,*
- *Propositional dependence logic ($\mathcal{PDL}$) $:= \mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \vee)$,*
- *Propositional intuitionistic dependence logic ($\mathcal{PIDL}$) $:= \mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \otimes, \rightarrow)$.*

As for $\mathcal{ML}$, the semantics of $\mathcal{MIDL}$ is defined with respect to the usual Kripke models.

Definition 3. *A* Kripke model *is a triple $K = (S, R, \pi)$ consisting of a nonempty set S, an accessibility relation $R \subseteq S \times S$, and a labeling (or valuation) function $\pi : S \rightarrow \wp(\textsc{Prop})$, where* Prop *is the set of all propositional variables. The set S is called the* domain *of K. Elements in S are called* states *or* worlds*, while subsets T of S are called* teams*, i.e. a team is a set of states.*

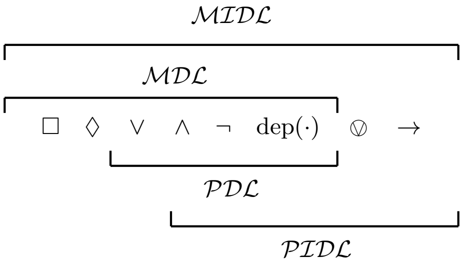


Fig. 1. Sublogics of $\mathcal{MIDL}$

For any team T of a Kripke model K, we define

$$R(T) = \{s \in K \mid \exists s' \in T, \text{ s.t. } s'Rs\}.$$

Elements in $R(\{s\})$ are called *successors* of s. A team T' is called a *successor team* of another team T, in symbols TRT', if every element in T has a successor in T'.

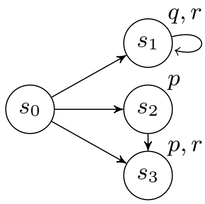
As mentioned in Section 1, dependence between variables does not manifest in single states of Kripke models, thus the satisfaction relation of $\mathcal{MIDL}$ is defined with respect to *sets of states* (or *teams*). A non-compositional game theoretic semantics of $\mathcal{MDL}$ based on *set game* is given in [Vää08], such semantics can be easily generalized to $\mathcal{MIDL}$ ([Yan12]). In this paper, we consider a compositional *team semantics* only. Below, we present the formal definition of team semantics of $\mathcal{MIDL}$; the intuition of the semantics will be discussed afterwards.

Definition 4 (Semantics of $\mathcal{MIDL}$)
We inductively define the notion of an $\mathcal{MIDL}$ formula φ being satisfied *in a Kripke model $K = (S, R, \pi)$ on a team $T \subseteq S$, denoted by $K, T \models \varphi$, as follows:*

$K, T \models p$	*iff*	$p \in \pi(s)$ *for all* $s \in T$
$K, T \models \neg p$	*iff*	$p \notin \pi(s)$ *for all* $s \in T$
$K, T \models \text{dep}(p_1, \ldots, p_n, q)$	*iff*	*for all* $s_1, s_2 \in T$ *it holds that* $\pi(s_1) \cap \{p_1, \ldots, p_n\} = \pi(s_2) \cap \{p_1, \ldots, p_n\}$ *implies* $\pi(s_1) \cap \{q\} = \pi(s_2) \cap \{q\}$
$K, T \models \neg\text{dep}(p_1, \ldots, p_n, q)$	*iff*	$T = \emptyset$
$K, T \models \varphi \wedge \psi$	*iff*	$K, T \models \varphi$ *and* $K, T \models \psi$
$K, T \models \varphi \vee \psi$	*iff*	*there are sets* T_1, T_2 *with* $T = T_1 \cup T_2$ *such that* $K, T_1 \models \varphi$ *and* $K, T_2 \models \psi$
$K, T \models \varphi \otimes \psi$	*iff*	$K, T \models \varphi$ *or* $K, T \models \psi$
$K, T \models \varphi \rightarrow \psi$	*iff*	*for all subsets* $T' \subseteq T$, $K, T' \models \varphi$ *implies that* $K, T' \models \psi$
$K, T \models \Box\varphi$	*iff*	$K, R(T) \models \varphi$
$K, T \models \Diamond\varphi$	*iff*	*there is a set* $T' \subseteq W$ *such that* TRT' *and* $K, T' \models \varphi$

Team semantics is a generalization of the usual Kripke semantics (defined with respect to single states). Atomic facts described by propositions or negated propositions are defined to be true on a team if on each state of the team, they are true in the usual Kripke semantics sense.

The key difference between the usual modal logic $\mathcal{ML}$ and $\mathcal{MIDL}$ is that the latter has a new type of atomic formulae, namely dependence atoms. The intuitive meaning of a dependence atom $\text{dep}(p_1, \ldots, p_n, q)$ being satisfied on a team T is that in the set T, the value of q is *functionally determined* by the values of $p_1, \ldots, p_n$. For example, in the Kripke model K_0 depicted below, for every two states in the team $T_0 = \{s_1, s_2, s_3\}$, if they agree on the valuation of p, then they also agree on the valuation of q, thus the team semantics gives that



$$K_0, T_0 \models \text{dep}(p, q).$$

More precisely, in T_0, the truth value of q is a function of that of p. In this case, such a function $f : \{0, 1\} \to \{0, 1\}$ is defined as: $f(1) = 0$, $f(0) = 1$. This means that in T_0, if p is true, then q is false; if p is false, then q is true. It is also easy to see that

$$K_0, T_0 \not\models \text{dep}(p, r),$$

as p is true on both s_2 and s_3, but r has different truth values on these states.

According to the team semantics, the extreme case "$K, T \models \text{dep}(p)$" means that p has a constant value in the team T. Moreover, on singleton teams or empty team, all dependence atoms are satisfied, i.e. $K, T \models \text{dep}(p_1, \ldots, p_n)$ always holds for $|T| \leq 1$; this fact will be used later in the proofs of the main theorems.

We stipulate that negated dependence atoms $\neg\text{dep}(p_1, \ldots, p_n)$ are only satisfied on the empty team. From the team semantics point of view, by making such a stipulation, we are able to preserve the *empty team property* of $\mathcal{MIDL}$, namely, $K, \emptyset \models \varphi$ holds always for all formulae φ. The readers are referred to [Vää07], [Vää08] for further discussions and a game-theoretical motivation of the team semantics of negated dependence atom. Under such semantics, the atomic negation $\neg$ of $\mathcal{MIDL}$ is clearly different from the *classical negation*, as *law of excluded middle* of dependence atoms fails for both split disjunction $\vee$ and Boolean disjunction $\otimes$, e.g., neither $\text{dep}(p) \vee \neg\text{dep}(p)$ nor $\text{dep}(p) \otimes \neg\text{dep}(p)$ is semantically valid. We sometimes abbreviate the formula $\neg\text{dep}(p_1, \ldots, p_n)$ or $p \wedge \neg p$ as the constant $\bot$ (*falsum*).

The team semantics of modalities is a natural generalization of that of the usual modal logic. The idea is that to evaluate an $\mathcal{MIDL}$ formula with modalities on a team, we consider the successors of all the states of the team altogether.

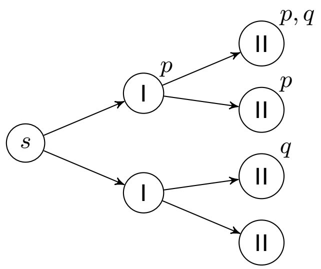
We say that $\Box\phi$ is true on a team T if ϕ is a property that is possessed by the set of all successors of all the points in T, and $\Diamond\phi$ to is said to be true on T if each state in T has access to some states such that these states form a team which satisfies ϕ. The interaction of modalities and dependence atoms gives $\mathcal{MIDL}$ more expressive power than the plain modal logic $\mathcal{ML}$. Below we present a simple and interesting example in this respect.

Example 1. Consider a very simple imperfect information game $\mathcal{G}$ between Player I and Player II with only 2 moves. In the first move of the game, Player I picks a truth value for p. In the second move, player II has to pick a truth value for q without knowing the choice of Player I in the previous move. Then the game ends and we evaluate the formula $\phi := p \vee q$. Player II wins if ϕ is true, otherwise Player I wins. We say that a strategy for a player in $\mathcal{G}$ is a *winning strategy* iff the player wins every play by following the strategy. A winning strategy for Player II is said further to be a *uniform winning strategy* iff in the second move, Player II's choice of the truth value for q is independent of that of Player I for p (see [Hod97a] or [Vää07] for a precise definition of *uniform winning strategies*).

The Kripke model K_1 depicted below represents the game tree of $\mathcal{G}$, where the nodes marked with "I" represent the possible moves for Player I and similarly for the nodes marked with "II". We hope the reader can easily see that judging whether Player II has a winning strategy in the perfect information version of the game $\mathcal{G}$ (i.e., in the second move, Player II is allowed to know Player I's first move) is equivalent to checking whether the $\mathcal{MIDL}$ expression $K_1, \{s\} \models \Box\Diamond(p \vee q)$ holds (or whether the $\mathcal{ML}$ expression $K_1, s \models \Box\Diamond(p \vee q)$ holds). Now, to say that Player II has a uniform winning strategy in $\mathcal{G}$, we further require that

$$K_1, \{s\} \models \Box\Diamond(\text{dep}(q) \wedge (p \vee q))$$

holds, where the dependence atom dep(q) guarantees that the choice of the truth value for q does not depend on that for p, namely, it does not depend on anything else or it has a constant value.



Formulae of $\mathcal{MIDL}$ are *downwards closed*, namely if $K, T \models \varphi$ and $T' \subseteq T$, then $K, T' \models \varphi$. It is possible to define many propositional connectives or operators which would preserve the downwards closure property of the logic, however, the ones considered in this paper (i.e. $\wedge$, $\vee$, $\otimes$ and $\rightarrow$) are of special interests for the following reason. As discussed in [AV09], taking all the downwards closed subsets of $\wp(\mathbf{2}^{\text{PROP}})$, one forms the algebra of the underlying propositional logic of $\mathcal{MIDL}$. Such a structure is both a *commutative quantale* (which carries an interpretation of linear logic) and a *complete Heyting algebra* (which carries an interpretation of intuitionistic logic). The connectives $\wedge$, $\vee$, $\owedge$ and $\to$ of $\mathcal{MIDL}$ correspond to conjunction, multiplicative conjunction (instead of disjunction in team semantics!), intuitionistic disjunction and intuitionistic implication of the algebraic structure, respectively. Readers are referred to [AV09] for further motivation for the connectives[3].

On singleton teams, the team semantics of atomic negation ($\neg$), conjunction ($\wedge$), split/Boolean disjunction ($\vee/\owedge$) and intuitionistic implication ($\to$) agree with the usual Kripke semantics (or in fact Tarskian semantics) of classical negation, classical conjunction, classical disjunction, classical material implication on the single state in plain modal logic $\mathcal{ML}$ (or in fact, classical propositional logic). Especially, on singleton teams, $\mathcal{MDL}$ is, in fact, equivalent to $\mathcal{ML}$ ([Sev09]).

On an arbitrary team, split disjunction $\vee$ *splits* the team into two subteams (not necessarily disjoint) in such a way that each disjunct is satisfied by one of the subteams. We use the standard disjunction symbol "$\vee$" to denote such split disjunction, so $\mathcal{MIDL}$ formulae without any occurrences of dependence atoms, intuitionistic implication $\to$ and Boolean disjunction $\owedge$ have the same syntax as those of plain modal logic $\mathcal{ML}$. In $\mathcal{MIDL}$, $\mathcal{ML}$ formulae φ (viewed as $\mathcal{MIDL}$ formulae) are *flat*, meaning that for all Kripke models K and all teams T of K

$$K, T \models \varphi \quad \text{iff} \quad K, \{s\} \models \varphi \text{ for all } s \in T.$$

Because of this property, on $\mathcal{ML}$ formulae, the split disjunction acts the same way as the usual disjunction of $\mathcal{ML}$ in the following sense. For an $\mathcal{ML}$ formula of the form $\varphi \vee \psi$, by flatness, evaluating $K, T \models \varphi \vee \psi$ is equivalent to evaluating whether for all $s \in T$, $K, \{s\} \models \varphi$ or $K, \{s\} \models \psi$ holds, namely whether there are two subteams of T each satisfying one of the disjuncts, which is exactly the team semantics of split disjunction.

As the name suggested, Boolean disjunction $\owedge$ acts on *full teams* in a Boolean way. Without going into details, we note that the team semantics of the constant $\perp$ and the connectives $\wedge$, $\owedge$, $\to$ over the usual modal Kripke models $K = (S, R, \pi)$ coincides with the Kripke semantics of those in intuitionistic modal logic defined by Fischer Servi [Ser81] over bi-relation Kripke models of the form $K^c = (\wp(S) \setminus \{\emptyset\}, R^c, \supseteq, \pi^c)$ (see [Yan12]). In particular, intuitionistic implication $\to$ acts intuitionistically over the superset accessibility relation in the powerset model K^c corresponding to K.

We say that two $\mathcal{MIDL}$ formulae φ and ψ are *logically equivalent*, in symbols $\varphi \equiv \psi$, if for any Kripke model K and any team T of K

$$K, T \models \varphi \quad \text{iff} \quad K, T \models \psi.$$

[3] In [AV09], another connective, *linear implication* $\multimap$, is considered. However, $\mathcal{MIDL}$ with $\multimap$ does not have the nice empty team property, we leave this case for future research.

Lemma 1. *The following equivalences follow straightforwardly from the semantics:*

a) $\mathrm{dep}(p) \equiv p \otimes \neg p$;
b) $\mathrm{dep}(p_1, \ldots, p_n, q) \equiv \big(\mathrm{dep}(p_1) \wedge \cdots \wedge \mathrm{dep}(p_n)\big) \to \mathrm{dep}(q)$
$$\equiv \bigvee_{i_1,\ldots,i_n \in \{\top,\bot\}} \big(p_1^{i_1} \wedge \cdots \wedge p_n^{i_n} \wedge (q \otimes \neg q)\big),$$
where $p^\top := p$ *and* $p^\bot := \neg p$.

Proof. Easy. □

As shown in the above lemma, dependence atoms $\mathrm{dep}(p_1, \ldots, p_n)$ are definable using the connectives $\vee$, $\otimes$ and $\to$. In particular, we obtain the following equivalent definitions:

$$\mathcal{MIDL} = \mathcal{MIDL}(\neg, \wedge, \vee, \otimes, \to),$$
$$\mathcal{PIDL} = \mathcal{MIDL}(\neg, \wedge, \otimes, \to).$$

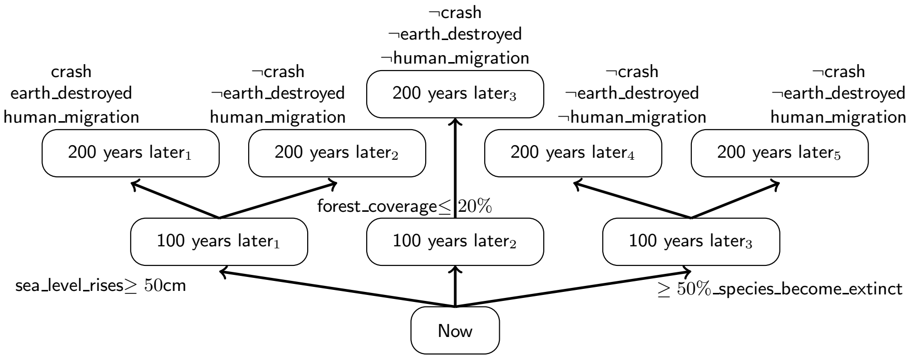
However, in this paper we still include dependence atoms in the language of $\mathcal{MIDL}$, as we will investigate sublogics of $\mathcal{MIDL}$ where not always $\vee$ and $\otimes$ are present.

2.2 Model Checking Problem

Given a Kripke model K, a team T, and an $\mathcal{MIDL}$ formula φ, the *model checking* problem for $\mathcal{MIDL}$ is the problem of deciding whether $K, T \models \varphi$ holds. A typical $\mathcal{MIDL}$ formula expresses a modal property involving functional dependence between propositions and implications of dependence statements. Such properties are commonly found in many fields. Knowing whether these properties hold in certain sets of some system can be very important in many cases. Below we present an example illustrating the applications of $\mathcal{MIDL}$ model checking in practice.

Example 2. Suppose the United Nations wants to build a model (represented as a Kripke model) of the imitation of the future of the earth and human race. Among all the candidate models, the United Nations wants to know which ones are optimistic models from the point of view of environmental degradation. One important criterion of being such an optimistic model is that in the model, the present world has to satisfy the sentence $(*)$ in Section 1. Selecting optimistic models with respect to this criterion is done by implementing an $\mathcal{MIDL}$-MC on the candidate models.

For example, given the below depicted Kripke model K_2, where every symbol is self-explanatory, we achieve the above goal by checking whether



$$K_2, \{\mathsf{now}\} \models \Box\Diamond(\mathrm{dep}(\mathsf{crash}, \mathsf{earth\_destroyed}) \to \mathrm{dep}(\mathsf{crash}, \mathsf{human\_migration}))$$

holds. In this case, the above expression holds, as for the team

$$T_2 = \{\mathsf{200\ years\ later}_1, \mathsf{200\ years\ later}_3, \mathsf{200\ years\ later}_4\},$$

it holds that

$$K_2, T_2 \models \mathrm{dep}(\mathsf{crash}, \mathsf{earth\_destroyed}) \to \mathrm{dep}(\mathsf{crash}, \mathsf{human\_migration}),$$

therefore K_2 is an optimistic model with respect to criterion $(*)$.

Now, we give the formal definition of the model checking problem for $\mathcal{MIDL}$. Following [EL11], we will investigate the source of hardness of the problem by studying also the model checking problem for fragments of $\mathcal{MIDL}$ built by restricting the operators allowed in the logics.

Definition 5 ($\mathcal{MIDL}$-MC, $\mathcal{MDL}$-MC, $\mathcal{PIDL}$-MC). *Let $\mathcal{L}$ be a sublogic of $\mathcal{MIDL}$ and M a subset of the set of operators occurring in $\mathcal{L}$. The* model checking *problem for $\mathcal{L}(M)$ is defined as the decision problem of the set*

$$\mathcal{L}(M)\text{-MC} := \left\{ \langle K, T, \varphi \rangle \;\middle|\; \begin{array}{l} K = (S, R, \pi) \textit{ is a Kripke model, } T \subseteq S,\ \varphi \in \mathcal{L}(M) \textit{ and} \\ K, T \models \varphi \end{array} \right\}.$$

*In case $\mathcal{L}(M) = \mathcal{L}$, we will only write $\mathcal{L}$-*MC *instead of $\mathcal{L}(M)$-*MC*.*

Note that in this paper, we only consider the combined complexity of model checking problem for $\mathcal{MIDL}$, i.e. the input consists of both a model and a formula. One can also consider the data complexity of model checking problem for $\mathcal{MIDL}$, where the formula is fixed and the input consists of a model only. Usually, the data complexity of a model checking problem is lower than the combined complexity. In our case, for a given $\mathcal{MIDL}$ formula with finitely many propositional variables, there are even only finitely many irreducible models (with respect to p-morphisms), therefore the data complexity for $\mathcal{MIDL}$ model checking is not very interesting and is not the topic of this paper.

In the main part of the paper we will sometimes reduce one model checking problem to another in a complexity preserving way. Next, we give the definition of such a reduction.

Definition 6. *Let C be a countable set and $A, B \subseteq C$. Then A is* polynomial-time many-one reducible *to B, in symbols $A \leq^{\mathsf{P}}_{\mathrm{m}} B$, iff there is a* reduction function *$f : C \to C$ such that f is computable in polynomial time and for all $x \in C$ it holds that $x \in A$ iff $f(x) \in B$. We write $A \equiv^{\mathsf{P}}_{\mathrm{m}} B$ if both $A \leq^{\mathsf{P}}_{\mathrm{m}} B$ and $B \leq^{\mathsf{P}}_{\mathrm{m}} A$ hold.*

Most complexity classes $\mathcal{C}$ with $\mathsf{P} \subseteq \mathcal{C}$ (e.g. PSPACE, coNP) are closed under $\leq^{\mathsf{P}}_{\mathrm{m}}$, i.e. if $A \leq^{\mathsf{P}}_{\mathrm{m}} B$ and $B \in \mathcal{C}$, then also $A \in \mathcal{C}$.

We end this section by pointing out that for any set M of $\mathcal{MIDL}$ operators, the complexity of $\mathcal{MIDL}(M)$-MC is independent of the presence of atomic negation $\neg$ in $\mathcal{MIDL}(M)$, namely:

Fact 1. $\mathcal{MIDL}(M)$-MC $\equiv^{\mathsf{P}}_{\mathrm{m}}$ $\mathcal{MIDL}(M \setminus \{\neg\})$-MC.

This is basically because given an $\mathcal{MIDL}(M)$-MC instance $\langle K, T, \varphi\rangle$, in the $\mathcal{MIDL}(M)$ formula φ, if one replaces all negated dependence atom $\neg\mathrm{dep}(p_1, \ldots, p_n)$ by a fixed fresh propositional variable r, and every occurrence of every negated atomic subformula $\neg p$ by a fresh propositional variable p', and modifies the valuation of K in such a way that r is made to be true nowhere and p' is made to be true only on the states where p is false, then the resulting formula φ' and Kripke model K' would satisfy

$$K, T \models \varphi \iff K', T \models \varphi'.$$

3 Complexity of Model Checking for fragments of $\mathcal{MIDL}$

In this section we study the complexity of model checking for fragments of $\mathcal{MIDL}$ and obtain the results listed in Table 1. The results for the fragments where the intuitionistic implication $\to$ is not present have been obtained already in [EL11], so we will only consider the cases where $\to$ is involved. We start with giving a PSPACE algorithm for $\mathcal{MIDL}$-MC.

Theorem 2. $\mathcal{MIDL}$-MC *is in* PSPACE.

Proof. To prove the theorem, it suffices to give an algorithm solving the problem that can be implemented on an alternating Turing machine running in polynomial time (AP Turing machines) ([CKS81]). An AP Turing machine uses an extension of ordinary nondeterministic guessing. Here the algorithm can switch between two guessing modes, namely universal and existential guessing. Existential guessing is the nondeterministic guessing mode of NP machines, whereas universal guessing is the guessing mode of coNP machines. When the number of alternations is unbounded, AP Turing machines decide the problems of the complexity class PSPACE.

Now, to prove the theorem, we consider a top down algorithm which has as input a Kripke model K, an $\mathcal{MIDL}$ formula φ, and a team T of K. The output of the algorithm is "true" if and only if $K, T \models \varphi$. In the cases

$$\varphi \in \{p, \neg p, \mathrm{dep}(p_1, \ldots, p_n), \neg\mathrm{dep}(p_1, \ldots, p_n), \psi \wedge \chi, \psi \otimes \chi\},$$

the algorithm checks whether $K, T \models \varphi$ according to the team semantics in an obvious way. These cases are deterministic and can be done in PSPACE.

If $\varphi = \psi \vee \chi$, or $\varphi = \Diamond\psi$ the algorithm guesses existentially the right fragmentation of the team and the right succeeding team, respectively. In the case of $\varphi = \psi \to \chi$ the algorithm checks universally if for every team $T' \subseteq T$ (i.e. every computation path) with $K, T' \models \psi$, it also holds that $K, T' \models \chi$. Altogether the algorithm can be implemented on an alternating Turing machine running in polynomial time or – equivalently – on a deterministic machine using polynomial space.

For the full algorithm (Listing 1.1) see Lemma 2 in the Appendix. □

If we forbid split disjunction $\vee$ and diamond $\Diamond$ in the sublogic of $\mathcal{MIDL}$ in question, the complexity of the above algorithm drops to coNP.

Corollary 1. $\mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \otimes, \to, \Box)$-MC *is in* coNP. *In particular,* $\mathcal{PIDL}$-MC *is in* coNP.

Proof. In Listing 1.1, existential guessing only applies to the cases $\varphi = \psi \vee \chi$ and $\varphi = \Diamond\psi$. □

If neither dependence atoms nor Boolean disjunction $\otimes$ are allowed in the logic, the model checking problem can even be decided in deterministic polynomial time.

Theorem 3. $\mathcal{MIDL}(\neg, \wedge, \vee, \to, \Box, \Diamond)$-MC *is in* P.

Proof. First, as pointed out in Section 2.1, the logic $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$ and the usual modal logic $\mathcal{ML}$ (taking all formulae in negation normal form) have the same syntax, thus as $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$ formulae are all flat, it can be easily shown that

$$\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)\text{-MC} \equiv^{\mathsf{P}}_{\mathrm{m}} \mathcal{ML}\text{-MC}.$$

We know by [CES86] that $\mathcal{ML}$-MC is in P, so it suffices to show that

$$\mathcal{MIDL}(\neg, \wedge, \vee, \to, \Box, \Diamond)\text{-MC} \equiv^{\mathsf{P}}_{\mathrm{m}} \mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)\text{-MC}. \quad (1)$$

Let φ and ψ be $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$ formulae. Viewing φ as an $\mathcal{ML}$ formula, we have that in $\mathcal{ML}$, the formula $\neg\varphi$ has an equivalent formula $\varphi^{\neg}$ in negation normal form; such a formula $\varphi^{\neg}$ can also be viewed as an $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$ formula. In $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$, it is not hard to prove by induction that

$$\varphi \to \psi \;\equiv\; \varphi^{\neg} \vee \psi. \quad (2)$$

Now, for each $\mathcal{MIDL}(\neg, \wedge, \vee, \to, \Box, \Diamond)$ formula φ, starting from the innermost intuitionistic implication $\to$, by applying Equation (2), we may eliminate all the occurrences of the connective $\to$ in φ and obtain an equivalent formula φ^* in the language of $\mathcal{MIDL}(\neg, \wedge, \vee, \Box, \Diamond)$. Such a translation can clearly be done in polynomial time, thus Equation (1) is obtained. □

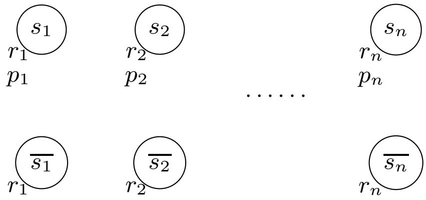


Fig. 2. Kripke model K in the proof of Theorem 4

In the remaining part of this section we provide hardness proofs for model checking problems for various sublogics of $\mathcal{MIDL}$. We first consider the sublogics without diamond $\Diamond$ and split disjunction $\vee$.

Theorem 4. *$\mathcal{PIDL}(\wedge, \varovee, \rightarrow)$-MC is* coNP*-hard.*

Proof. Noting that by Lemma 1 and Fact 1,

$$\begin{aligned}\mathcal{PIDL}(\wedge, \varovee, \rightarrow)\text{-MC} &\equiv^{\mathsf{P}}_{\mathrm{m}} \mathcal{PIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \varovee, \rightarrow)\text{-MC} \\ &\equiv^{\mathsf{P}}_{\mathrm{m}} \mathcal{PIDL}\text{-MC},\end{aligned}$$

it suffices to give a polynomial-time reduction from the known coNP-complete problem

$$\textsc{Taut} := \{\varphi \text{ a classical propositional logic } (\mathcal{CPL}) \text{ formula} \mid \varphi \text{ is a tautology}\}$$

to $\mathcal{PIDL}$-MC. For this purpose let φ be an arbitrary formula in the language of classical propositional logic in negation normal form and let $p_1, \ldots, p_n$ be the propositional variables occurring in φ. Let $r_1, \ldots, r_n$ be new propositional variables and $K = (S, R, \pi)$ the Kripke model shown in Figure 2 and formally defined by

$$\begin{aligned}S &:= \{s_1, \ldots, s_n, \overline{s_1}, \ldots, \overline{s_n}\}, \\ R &:= \emptyset, \\ \pi(s_i) &:= \{r_i, p_i\}, \\ \pi(\overline{s_i}) &:= \{r_i\}.\end{aligned}$$

Next, define a $\mathcal{PIDL}$ formula ψ as follows:

$$\psi \;:=\; \alpha_n \rightarrow \varphi^{\rightarrow},$$

where

$$\alpha_n := \bigwedge_{i=1}^{n} (r_i \rightarrow \mathrm{dep}(p_i)),$$

and $\varphi^{\rightarrow}$ is defined inductively as follows:

$$\begin{aligned}p_i^{\rightarrow} &:= r_i \rightarrow p_i, \\ (\neg p_i)^{\rightarrow} &:= r_i \rightarrow \neg p_i, \\ (\varphi \wedge \psi)^{\rightarrow} &:= \varphi^{\rightarrow} \wedge \psi^{\rightarrow}, \\ (\varphi \vee \psi)^{\rightarrow} &:= \varphi^{\rightarrow} \varovee \psi^{\rightarrow}.\end{aligned}$$

Now we will show that $\varphi \in$ TAUT iff $K, S \models \psi$.

The general idea of the proof is as follows. By the construction, each team T of K satisfying the formula α_n contains at most one of the states s_i and $\overline{s_i}$, for each i. In the Kripke model K, the state s_i simulates positive truth assignments for p_i (i.e. assignments σ such that $\sigma(p_i) = \top$), while $\overline{s_i}$ simulates negative truth assignments for p_i (i.e. assignments σ such that $\sigma(p_i) = \bot$). Thus, any maximal such team T simulates a truth assignment σ_T for $p_1, \ldots, p_n$. Moreover, under the assignment σ_T , the $\mathcal{CPL}$ formula φ will behave exactly as the $\mathcal{PIDL}$ formula $\varphi^{\rightarrow}$ does under the team T, in the sense that $\sigma_T(\varphi) = \top$ iff $K, T \models \varphi^{\rightarrow}$. The required equivalence will then follow.

Formally, first suppose $\varphi \in$ TAUT. Let $T_0 \subseteq S$ be an arbitrary team such that $K, T_0 \models \alpha_n$. If T_0' is a nonempty subteam of T_0 such that $K, T_0' \models r_i$, then by the construction of K, $T_0' \subseteq \{s_i, \overline{s_i}\}$. But as $K, T_0' \models r_i \rightarrow \mathrm{dep}(p_i)$, we must have that $K, T_0' \models \mathrm{dep}(p_i)$, so p_i has a constant value in T_0', which means that T_0' contains only one state of s_i and $\overline{s_i}$. Therefore, team T_0 contains at most one state of s_i and $\overline{s_i}$ for each i.

Now, there is a maximal team $T \supseteq T_0$ with $K, T \models \alpha_n$. Team T contains exactly one of the states s_i and $\overline{s_i}$ for each $1 \leq i \leq n$, therefore T induces a truth assignment σ_T for $p_1, \ldots, p_n$ defined as follows:

$$\sigma_T(p_i) := \begin{cases} \top \text{ if } s_i \in T, \\ \bot \text{ if } \overline{s_i} \in T. \end{cases}$$

Such team T and its *induced truth assignment* σ_T are in one-one correspondence; moreover, the assignment σ_T makes the $\mathcal{CPL}$ formula φ true if and only if the team T satisfies the $\mathcal{PIDL}$ formula $\varphi^{\rightarrow}$. We show this by showing a more general claim as follows:

Claim. For all subformulae χ of φ, it holds that $\sigma_T(\chi) = \top$ iff $K, T \models \chi^{\rightarrow}$.

Proof (Proof of Figure 2). An easy inductive proof. We only show the case that $\chi = \neg p_i$. First suppose $\sigma_T(\neg p_i) = \top$. Then $\overline{s_i} \in T$ and $s_i \notin T$, thus, for all non-empty team $T' \subseteq T$ such that $K, T' \models r_i$, we must have that $T' = \{\overline{s_i}\}$, hence $K, T' \models \neg p$, which implies that $K, T \models r_i \rightarrow \neg p_i$.

Conversely, suppose $K, T \models r_i \rightarrow \neg p_i$. Then we must have that $s_i \notin T$, thus by the maximality of T, $\overline{s_i} \in T$ and $\sigma_T(p_i) = \bot$, i.e. $\sigma_T(\neg p_i) = \top$. □

Now, as $\varphi \in$ TAUT, it holds that $\sigma_T(\varphi) = \top$, which by Figure 2 implies that $K, T \models \varphi^{\rightarrow}$, hence, as $T_0 \subseteq T$, by downward closure property, $K, T_0 \models \varphi^{\rightarrow}$. As T_0 was chosen arbitrarily with $T_0 \subseteq S$ and $K, T_0 \models \alpha_n$, this implies that $K, S \models \psi$.

Conversely suppose that $K, S \models \psi$ and σ is an arbitrary truth assignment for $p_1, \ldots, p_n$. The truth assignment σ induces a team T_σ defined by

$$T_\sigma := \{s_i \mid \sigma(p_i) = \top\} \cup \{\overline{s_i} \mid \sigma(p_i) = \bot\}.$$

Clearly, $K, T_\sigma \models \alpha_n$, thus $K, T_\sigma \models \varphi^{\rightarrow}$, which by Figure 2 implies that $\sigma(\varphi) = \top$. Since σ was chosen arbitrarily, this means that $\varphi \in$ TAUT. □

Theorem 5. *For all* $\{\wedge, \ovee, \to\} \subseteq M \subseteq \{\neg, \mathrm{dep}(\cdot), \wedge, \ovee, \to, \Box\}$, $\mathcal{MIDL}(M)$-MC *is* coNP*-complete. In particular,* $\mathcal{PIDL}$-MC *is* coNP*-complete.*

Proof. Follows from Corollary 1 and Theorem 4. □

Next, we analyze the complexity of model checking for fragments of $\mathcal{MIDL}$ containing split disjunction $\vee$ and intuitionistic implication $\to$.

Theorem 6. *Let* $M \supseteq \{\mathrm{dep}(\cdot), \wedge, \vee, \to\}$*. Then* $\mathcal{MIDL}(M)$-MC *is* PSPACE*-complete.*

Proof. The upper bound follows from Theorem 2. For the lower bound we give a reduction from the well-known PSPACE-complete problem, true quantified Boolean formulae QBF. Let $\psi = \forall x_1 \exists x_2 \ldots \forall x_{n-1} \exists x_n \, \varphi$ be a QBF instance, where φ is quantifier-free. We assume without loss of generality that n is even and that $\varphi = C_1 \wedge \cdots \wedge C_m$ is in 3CNF with

$$C_j = \alpha_{j0} \vee \alpha_{j1} \vee \alpha_{j2} \quad (1 \leq j \leq m)$$

for distinct propositional literals $\alpha_{j0}, \alpha_{j1}, \alpha_{j2}$. Let

$$r_1, \ldots, r_n, p_1, \ldots, p_n, c_1, \ldots, c_m, c_{10}, \ldots, c_{m0}, c_{11}, \ldots, c_{m1}, c_{12}, \ldots, c_{m2}$$

be distinct propositional variables. The corresponding $\mathcal{MIDL}(\mathrm{dep}(\cdot), \wedge, \vee, \to)$-MC instance is defined as $(K = (S, R, \pi), S, \theta)$, where

- $S := \{s_1, \ldots, s_n, \overline{s_1}, \ldots, \overline{s_n}\}$,
- $R := \emptyset$,
- $\pi(s_i) = \{r_i, p_i\} \cup \{c_j, c_{j0} \mid \alpha_{j0} = x_i,\ 1 \leq j \leq m\} \cup \{c_j, c_{j1} \mid \alpha_{j1} = x_i,\ 1 \leq j \leq m\} \cup \{c_j, c_{j2} \mid \alpha_{j2} = x_i,\ 1 \leq j \leq m\}$,
- $\pi(\overline{s_i}) = \{r_i\} \cup \{c_j, c_{j0} \mid \alpha_{j0} = \neg x_i,\ 1 \leq j \leq m\} \cup \{c_j, c_{j1} \mid \alpha_{j1} = \neg x_i,\ 1 \leq j \leq m\} \cup \{c_j, c_{j2} \mid \alpha_{j2} = \neg x_i,\ 1 \leq j \leq m\}$

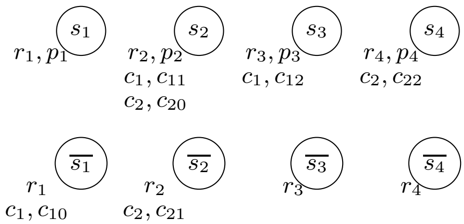
  (see Figure 3 for an example of the construction of K),
- $\theta := \delta_1$, where

$$\begin{aligned} \delta_{2k-1} &:= (r_{2k-1} \to \mathrm{dep}(p_{2k-1})) \to \delta_{2k} \quad (1 \leq k \leq n/2), \\ \delta_{2k} &:= (r_{2k} \wedge \mathrm{dep}(p_{2k})) \vee \delta_{2k+1} \quad (1 \leq k < n/2), \\ \delta_n &:= (r_n \wedge \mathrm{dep}(p_n)) \vee \varphi', \end{aligned}$$

  and

$$\varphi' := \bigwedge_{j=1}^{m} \Big(c_j \to \Big(\ (\mathrm{dep}(c_{j0}) \wedge \mathrm{dep}(c_{j1}) \wedge \mathrm{dep}(c_{j2})) \vee (\mathrm{dep}(c_{j0}) \wedge \mathrm{dep}(c_{j1}) \wedge \mathrm{dep}(c_{j2})) \ \Big)\Big),$$

  i. e. $\theta = \ (r_1 \to \mathrm{dep}(p_1)) \to \Big((r_2 \wedge \mathrm{dep}(p_2)) \vee \cdots \ \cdots \ \cdots \vee \Big((r_{n-1} \to \mathrm{dep}(p_{n-1})) \to \Big((r_n \wedge \mathrm{dep}(p_n)) \vee \varphi' \Big) \Big) \cdots \Big).$



The model K for $\varphi = (\neg x_1 \vee x_2 \vee x_3) \wedge (x_2 \vee \neg x_2 \vee x_4)$

Fig. 3. Example of the Kripke model construction in the proof of Theorem 6

We will now show that $\psi \in$ QBF iff $K, S \models \theta$.

The general idea of the proof is that the alternating operators $\rightarrow$ and $\vee$ in the $\mathcal{MIDL}$ formula θ simulate the alternating quantifiers $\forall$ and $\exists$ in the QBF formula ψ, respectively. In the QBF formula ψ, for each $1 \leq k \leq n/2$, the universal quantifier $\forall x_{2k-1}$ corresponds to the $\mathcal{MIDL}$ formula δ_{2k-1} in θ , and we have that δ_{2k-1} is satisfied in a team T iff δ_{2k} is satisfied in all subteams $T_{2k-1} \subseteq T$ which satisfy $r_{2k-1} \rightarrow \text{dep}(p_{2k-1})$, i. e. all subteams containing at most one of the states s_{2k-1} and $\overline{s_{2k-1}}$. Every existential quantifier $\exists x_{2k}$ in ψ corresponds to the subformula δ_{2k} of θ, and δ_{2k} is satisfied in a team T iff T can be split into T_{2k} and T'_{2k} such that $K, T_{2k} \models \delta_{2k+1}$ and $K, T'_{2k} \models r_{2k} \wedge \text{dep}(p_{2k})$, i. e. δ_{2k+1} has to be satisfied in a team with exactly one of the states s_{2k} and $\overline{s_{2k}}$.

In the Kripke model K, we start with the team S of all states and then for every nesting level i of $\rightarrow$ or $\vee$ we drop exactly one of the states s_i and $\overline{s_i}$ according to the truth assignment of the Boolean variable x_i quantified by $\forall$ or $\exists$. We do this until we arrive at a team T_n that contains exactly one of the states s_i and $\overline{s_i}$ for all $i \in \{1, \ldots, n\}$. This team T_n is in fact the team induced by the complement of a truth assignment σ for $x_1, \ldots, x_n$ (in a similar sense with that in the proof of Theorem 4) and $K, T_n \models \varphi'$ iff $\sigma(\varphi) = \top$. Then the required equivalence will follow.

Formally, first suppose that $\psi \in$ QBF. During the proof, we will construct a truth assignment σ for $x_1, \ldots, x_n$ such that $\sigma(\varphi) = \top$ by choosing values for

$$\sigma(x_1), \sigma(x_3), \ldots, \sigma(x_{n-1}).$$

The assumption guarantees that an appropriate value for each $\sigma(x_{2k})$ $(k \in \{1, \ldots, n/2\})$ exists and they are determined by the values of $\sigma(x_1), \sigma(x_3), \ldots, \sigma(x_{2k-1})$.

We have to show that

$$K, S \models (r_1 \rightarrow \text{dep}(p_1)) \rightarrow \delta_2.$$

By the downward closure property , it suffices to show that for the maximal teams $T_1 \subseteq S$ such that $K, T_1 \models r_1 \rightarrow \text{dep}(p_1)$, namely the teams $S \setminus \{s_1\}$ and

$S \setminus \{\overline{s_1}\}$, it holds that

$$K, T_1 \models \delta_2, \text{ i.e. } K, T_1 \models (r_2 \wedge \text{dep}(p_2)) \vee \delta_3.$$

Choose the value of $\sigma(x_1)$ according to T_1 by letting

$$\sigma(x_1) := \begin{cases} \top \text{ if } T_1 = S \setminus \{s_1\}, \\ \bot \text{ if } T_1 = S \setminus \{\overline{s_1}\}. \end{cases}$$

Note that $\sigma \restriction \{x_1\}$ is defined as the *complement* of the truth assignment induced by T_1 – which was used in the proof of Theorem 4. We are going to continue to define the truth assignment as the complement of the induced one. The reason for this will become clear when we show the connection between φ and φ' in the end.

Since $\psi \in$ QBF, by our discussion above, an appropriate value of $\sigma(x_2)$ exists and is determined by $\sigma(x_1)$. Now we split the team T_1 into T_2 and T_2' according to the value of $\sigma(x_2)$ by letting

$$T_2' := \begin{cases} \{s_2\} \text{ if } \sigma(x_2) = \top, \\ \{\overline{s_2}\} \text{ if } \sigma(x_2) = \bot, \end{cases}$$

and $T_2 = T_1 \setminus T_2'$. Clearly, $K, T_2' \models r_2 \wedge \text{dep}(p_2)$ and it suffices to check that $K, T_2 \models \delta_3$.

As we have shown, to prove $K, S \models \delta_1$, it is enough to show that for every T_1 chosen as above, the above constructed T_2 satisfies $K, T_2 \models \delta_3$. By repeating the same arguments and constructions $n/2$ times, it remains to show that $K, T_n \models \varphi'$. Note that T_n and σ satisfy

$$\begin{aligned} s_i \in T_n &\iff \sigma(x_i) \quad = \bot, \\ \overline{s_i} \in T_n &\iff \sigma(\neg x_i) = \bot, \end{aligned} \tag{3}$$

moreover, $\sigma(\varphi) = \top$, i.e. for all $j \in \{1, \ldots, m\}$ it holds that $\sigma(\alpha_{j0}) = \top$, $\sigma(\alpha_{j1}) = \top$ or $\sigma(\alpha_{j2}) = \top$.

Now let $j \in \{1, \ldots, m\}$ arbitrarily chosen, and suppose, for example, $\alpha_{j0} = \neg x_{i_0}$, $\alpha_{j1} = \neg x_{i_1}$, $\alpha_{j2} = x_{i_2}$ for some $i_0, i_1, i_2 \in \{1, \ldots, n\}$, and $\sigma(\neg x_{i_1}) = \top$. Let $T \subseteq T_n$ be arbitrarily chosen such that $K, T \models c_j$. Then, by the construction of K, $T \subseteq \{\overline{s_{i_0}}, \overline{s_{i_1}}, s_{i_2}\}$, and furthermore, by (3) we obtain $\overline{s_{i_1}} \notin T$. Hence, $T \subseteq \{\overline{s_{i_0}}, s_{i_2}\}$, therefore

$$K, T \models \big(\text{dep}(c_{j0}) \wedge \text{dep}(c_{j1}) \wedge \text{dep}(c_{j2})\big) \vee \big(\text{dep}(c_{j0}) \wedge \text{dep}(c_{j1}) \wedge \text{dep}(c_{j2})\big),$$

as dependence atoms are always satisfied on singleton teams. Since j was chosen arbitrarily, it follows that $K, T_n \models \varphi'$.

Conversely, suppose $K, S \models \theta$. Choose arbitrarily the values for

$$\sigma(x_1), \sigma(x_3), \ldots, \sigma(x_{2n-1})$$

and define the values for

$$\sigma(x_2), \sigma(x_4), \ldots, \sigma(x_{2n})$$

by reversing the above arguments and constructions, and repeating them $n/2$ times, we then arrive at (3) again. The crucial observation is that when evaluating $(r_{2k-1} \to \mathrm{dep}(p_{2k-1})) \to \delta_{2k}$ we only need to consider the maximal teams satisfying $r_{2k-1} \to \mathrm{dep}(p_{2k-1})$ and there are exactly two of those, one without s_{2k-1} and the other one without $\overline{s_{2k-1}}$. And when evaluating $(r_{2k} \wedge \mathrm{dep}(p_{2k})) \vee \delta_{2k+1}$ we have to consider only the complements of the maximal teams satisfying $r_{2k} \wedge \mathrm{dep}(p_{2k})$ and again there are exactly two, one without s_{2k} and the other one without $\overline{s_{2k}}$.

It remains to show that $\sigma(\varphi) = \top$. That is to show that $\sigma(\alpha_{j0} \vee \alpha_{j1} \vee \alpha_{j2}) = \top$ for an arbitrarily chosen $j \in \{1, \dots, m\}$. Suppose, for example,

$$\alpha_{j0} = x_{i_0},\ \alpha_{j1} = x_{i_1} \text{ and } \alpha_{j2} = \neg x_{i_2}.$$

Now let $T \subseteq T_n$ be the maximal team such that $K, T \models c_j$. Then, by construction of K, we have that $T \subseteq \{s_{i_0}, s_{i_1}, \overline{s_{i_2}}\}$. Since $K, T_n \models \varphi'$, it holds that

$$K, T \models \big(\mathrm{dep}(c_{j0}) \wedge \mathrm{dep}(c_{j1}) \wedge \mathrm{dep}(c_{j2})\big) \vee \big(\mathrm{dep}(c_{j0}) \wedge \mathrm{dep}(c_{j1}) \wedge \mathrm{dep}(c_{j2})\big)$$

and thus, by construction of K, it follows that $|T| \leq 2$. Say $s_{i_1} \notin T$, then, by maximality of T, we obtain that $s_{i_1} \notin T_n$ which, by (3), means that $\sigma(x_{i_1}) = \top$, i.e. $\sigma(\alpha_{j1}) = \top$. Hence, as j was chosen arbitrarily, it follows that $\sigma(\varphi) = \top$. □

Finally, we study the model checking problem for sublogics of $\mathcal{MIDL}$ including diamond $\Diamond$, Boolean disjunction $\varovee$ and intuitionistic implication $\to$.

Theorem 7. *Let* $M \supseteq \{\wedge, \varovee, \to, \Diamond\}$. *Then* $\mathcal{MIDL}(M)$-MC *is* PSPACE-*complete.*

Proof. The upper bound again follows from Theorem 2. For the lower bound we give a polynomial-time reduction from QBF to $\mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \varovee, \to, \Diamond)$-MC, which implies the desired result since

$$\mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \varovee, \to, \Diamond)\text{-MC} \equiv^{\mathsf{P}}_{\mathrm{m}} \mathcal{MIDL}(\wedge, \varovee, \to, \Diamond)\text{-MC},$$

by Lemma 1 and Fact 1.

Let $\psi = \forall x_1 \exists x_2 \dots \forall x_{n-1} \exists x_n\, \varphi$ with n even (without loss of generality we may assume) and φ quantifier-free. The corresponding $\mathcal{MIDL}(\neg, \mathrm{dep}(\cdot), \wedge, \varovee, \to, \Diamond)$-MC instance is defined as $(K = (S, R, \pi), T, \theta)$ where

– $S := \bigcup_{1 \leq i \leq n} S_i, \quad R := \bigcup_{1 \leq i \leq n} R_i$ and for $1 \leq i \leq n/2$

$$\begin{aligned}
S_{2i-1} &:= \{s_{2i-1}, \overline{s_{2i-1}}\} \\
S_{2i} &:= \{s_{2i}, \overline{s_{2i}}\} \cup \{t_i\} \cup \{t_{i1}, \cdots, t_{i(i-1)}\} \\
R_{2i-1} &:= \{(s_{2i-1}, s_{2i-1}), (\overline{s_{2i-1}}, \overline{s_{2i-1}})\} \\
R_{2i} &:= \{(t_i, t_{i1}), (t_{i1}, t_{i2}), \cdots, (t_{i(i-2)}, t_{i(i-1)})\} \\
&\quad \cup \{(t_{i(i-1)}, s_{2i}), (t_{i(i-1)}, \overline{s_{2i}})\} \\
&\quad \cup \{(s_{2i}, s_{2i}), (\overline{s_{2i}}, \overline{s_{2i}})\}
\end{aligned}$$

- $\pi(s_j) := \{r_j, p_j\}$,
- $\pi(\overline{s_j}) := \{r_j\}$,
- $\pi(t) := \emptyset$, for $t \notin \{s_j, \overline{s_j} \mid 1 \leq j \leq n\}$

  (Figure 4 depicts the construction of K),
- $T := \{s_i, \overline{s_i} \mid 1 \leq i \leq n,\ i \text{ odd}\} \cup \{t_i \mid 1 \leq i \leq n/2\}$;
- $\theta = \delta_1$, where

$$\begin{aligned} \delta_{2k-1} &:= (r_{2k-1} \rightarrow \mathrm{dep}(p_{2k-1})) \rightarrow \delta_{2k} \quad (1 \leq k \leq n/2), \\ \delta_{2k} &:= \Diamond \delta_{2k+1} \quad (1 \leq k < n/2), \\ \delta_n &:= \Diamond \varphi^{\rightarrow}, \end{aligned}$$

  and $\varphi^{\rightarrow}$ is generated from φ by the same inductive translation as in the proof of Theorem 4,

$$\text{i.e. } \theta = \big(r_1 \rightarrow \mathrm{dep}(p_1)\big) \rightarrow \Diamond\Big(\cdots \quad \cdots \rightarrow \Diamond \Big(\big(r_{n-1} \rightarrow \mathrm{dep}(p_{n-1})\big) \rightarrow \Diamond\varphi^{\rightarrow}\Big) \ldots \Big).$$

We will show that $\psi \in \text{QBF}$ iff $K, T \models \theta$. The idea, analogous to the proof of Theorem 6, is that the alternating operators $\rightarrow$ and $\Diamond$ in the $\mathcal{MIDL}$ formula θ simulate the quantifiers $\forall$ and $\exists$ in the QBF formula ψ, respectively. Note that

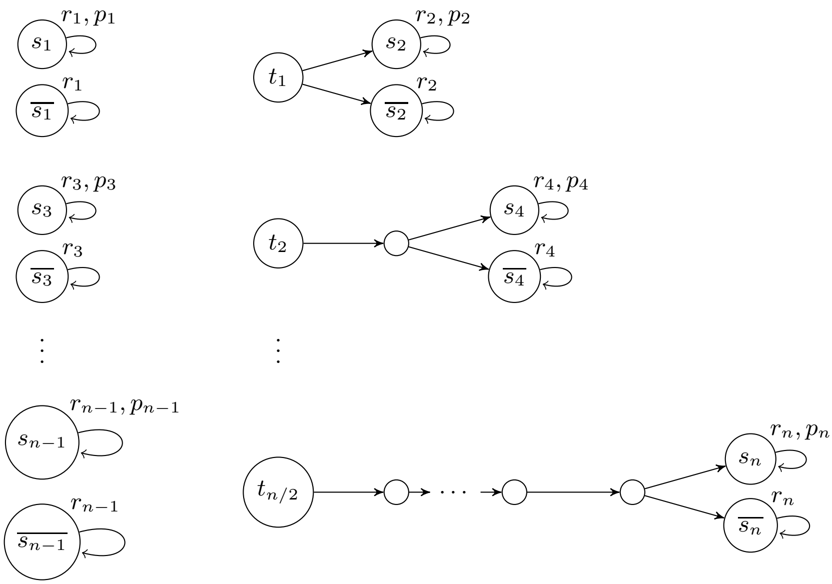


Fig. 4. Kripke model K in the proof of Theorem 7

the model K can be viewed as the disjoint union of n small models K_i (as shown in Figure 4), where each K_i contains for p_i a positive state s_i with a loop and a negative state $\overline{s_i}$ with a loop, and for even $1 \leq i \leq n$, K_i also contains a chain of length $(i/2-1)$ with a split leading to s_i and $\overline{s_i}$ at the end. The states in every K_i with no proper predecessors (all s_{2k-1}, $\overline{s_{2k-1}}$, t_k's for $1 \leq k \leq n/2$, these form the team T) can be viewed as starting states and those with no proper successors (all s_i, $\overline{s_i}$'s for $1 \leq i \leq n$) can be viewed as final states. In the proof, we start with the team T of all starting states, and then for every nesting level i of $\rightarrow$ we drop one of the states s_{2i+1} and $\overline{s_{2i+1}}$, while for every nesting level i of $\Diamond$ we simultaneously move forward on the chains and thereby choose one of the states s_{2i} and $\overline{s_{2i}}$. We do this until we arrive at a team T_n that contains exactly one of the final states s_i and $\overline{s_i}$ for each $i \in \{1, \ldots, n\}$. This team T_n induces a truth assignment σ for $x_1, \ldots, x_n$ as in the proof of Theorem 4, and $K, T_n \models \varphi^{\rightarrow}$ iff $\sigma(\varphi) = \top$.

Now, suppose that $\psi \in$ QBF. We have to show that

$$K, T \models (r_1 \rightarrow \mathrm{dep}(p_1)) \rightarrow \delta_2.$$

It then suffices to show that for the maximal teams $T_1 \subseteq T$ such that $K, T_1 \models r_1 \rightarrow \mathrm{dep}(p_1)$, it holds that

$$K, T_1 \models \delta_2, \text{ i. e. } K, T_1 \models \Diamond\delta_3.$$

Analogous to the proof of Theorem 6, choose the value of $\sigma(x_1)$ according to T_1 by letting

$$\sigma(x_1) := \begin{cases} \bot \text{ if } T_1 = T \setminus \{s_1\}, \\ \top \text{ if } T_1 = T \setminus \{\overline{s_1}\} \end{cases}$$

(but here $\sigma \upharpoonright \{x_1\}$ is defined as the truth assignment induced by T_1 instead of the complementary one, as in the proof of Theorem 6). By a similar argument with that in the proof of Theorem 6, an appropriate value of $\sigma(x_2)$ exists and is determined by $\sigma(x_1)$. We now choose T_2 such that $T_1 R T_2$ as follows:

$$T_2 := \begin{cases} R(T_1) \setminus \{\overline{s_2}\} \text{ if } \sigma(x_2) = \top, \\ R(T_1) \setminus \{s_2\} \text{ if } \sigma(x_2) = \bot, \end{cases}$$

It suffices to check that $K, T_2 \models \delta_3$. Again, analogous to the proof of Theorem 6, by repeating the universal and the existential arguments $n/2$ times, it remains to show that $K, T_n \models \varphi^{\rightarrow}$. And, analogous to (3), T_n and σ satisfy

$$\begin{aligned} s_i \in T_n &\iff \sigma(x_i) = \top, \\ \overline{s_i} \in T_n &\iff \sigma(\neg x_i) = \top, \end{aligned} \tag{4}$$

and moreover $\sigma(\varphi) = \top$.

Table 1. Classification of complexity for fragments of $\mathcal{MIDL}$-MC
All results are completeness results except for the P cases which are upper bounds.

$\Box$	$\Diamond$	$\wedge$	$\vee$	$\varovee$	$\neg$	$\rightarrow$	dep(·)	Complexity	Reference
$*$	$*$	$+$	$+$	$*$	$*$	$+$	$+$	PSPACE	Theorem 6
$*$	$*$	$+$	$+$	$+$	$*$	$+$	$*$	PSPACE	Theorem 6, Lemma 1
$*$	$+$	$+$	$*$	$+$	$*$	$+$	$*$	PSPACE	Theorem 7
$*$	$-$	$+$	$-$	$+$	$*$	$+$	$*$	coNP	Theorem 5
$*$	$*$	$*$	$*$	$-$	$*$	$*$	$-$	P	Theorem 3
$*$	$*$	$*$	$*$	$*$	$*$	$-$	$*$	P / NP	[EL11]

$+$: operator present $-$: operator absent $*$: complexity independent of operator

Noting that σ is the truth assignment induced by T_n, by Figure 2 in the proof of Theorem 4, we obtain that $K', T_n \models \varphi^{\rightarrow}$, where $K' = (S', R', \pi')$ with

- $S' = \{s_i, \overline{s_i} \mid 1 \leq i \leq n\}$,
- $R' = \emptyset$,
- $\pi' = \pi \restriction S'$

(i.e., K' is the model constructed in the proof of Theorem 4, which can also be viewed as a submodel of K). Next, since $\varphi^{\rightarrow}$ is modality-free, it follows that $K'', T_n \models \varphi^{\rightarrow}$, where K'' is the submodel of K generated by S' (namely K' with all the loops). Finally, as it is easy to check by induction that truth of $\mathcal{MIDL}$ formulae with respect to teams is invariant under taking generated submodels, we conclude that $K, T_n \models \varphi^{\rightarrow}$.

Conversely, suppose that $K, T \models \theta$. As in the proof of Theorem 6 we can reverse the above constructions and arrive at (4) again. The crucial point is that when evaluating $\Diamond\delta_{2k+1}$ we only need to consider minimal successor teams.

Now, by the construction of T_n, we have that $K, T_n \models \varphi^{\rightarrow}$. Reversing the above argument, by Figure 2 in the proof of Theorem 4, we obtain that $\sigma(\varphi) = \top$. $\Box$

4 Concluding Remarks

Table 1 contains all results we have obtained in this paper. We have shown that model checking for $\mathcal{MIDL}$ in general is PSPACE-complete and that this still holds if we forbid $\Box$, dep(·) and either $\Diamond$ or $\vee$. If we forbid $\Diamond$ and $\vee$, on the other hand, the complexity drops to coNP. In particular, $\mathcal{PIDL}$-MC is coNP-complete.

Note that some cases are missing in Table 1, e. g. the one where only conjunction is forbidden, the one where only both disjunctions are forbidden and the one from Theorem 7 but with dependence atoms allowed instead of Boolean disjunction. Also, we point out that the computational complexity of the satisfiability problem for $\mathcal{MIDL}$ is still open.

References

[AV09] Abramsky, S., Väänänen, J.: From IF to BI. Synthese 167(2), 207–230 (2009)

[CES86] Clarke, E.M., Emerson, E.A., Sistla, A.P.: Automatic verification of finite-state concurrent systems using temporal logic specifications. ACM Trans. Program. Lang. Syst. 8(2), 244–263 (1986)

[CKS81] Chandra, A.K., Kozen, D.C., Stockmeyer, L.J.: Alternation. J. ACM 28(1), 114–133 (1981)

[CR11] Ciardelli, I., Roelofsen, F.: Inquisitive logic. Journal of Philosophical Logic 40(1), 55–94 (2011)

[EL11] Ebbing, J., Lohmann, P.: Complexity of model checking for modal dependence logic, CoRR abs/1104.1034v1 (2011)

[End70] Enderton, H.: Finite partially-ordered quantifiers. Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik (16), 393–397 (1970)

[Hem05] Hemaspaandra, E.: The complexity of poor man's logic, CoRR cs.LO/9911014v2 (2005)

[Hen61] Henkin, L.: Some remarks on infinitely long formulas. In: Infinitistic Methods, Proceedings Symposium Foundations of Mathematics, pp. 167–183. Pergamon, Warsaw (1961)

[Hod97a] Hodges, W.: Compositional semantics for a langauge of imperfect information. Logic Journal of the IGPL 5, 539–563 (1997)

[Hod97b] Hodges, W.: Some Strange Quantifiers. In: Mycielski, J., Rozenberg, G., Salomaa, A. (eds.) Structures in Logic and Computer Science. LNCS, vol. 1261, pp. 51–65. Springer, Heidelberg (1997)

[HS89] Hintikka, J., Sandu, G.: Informational independence as a semantical phenomenon. In: Fenstad, J.E., Frolov, I.T., Hilpinen, R. (eds.) Logic, Methodology and Philosophy of Science, vol. 8, pp. 571–589. Elsevier, Amsterdam (1989)

[HS96] Hintikka, J., Sandu, G.: Game-theoretical semantics. In: van Benthem, J., ter Meulen, A. (eds.) Handbook of Logic and Language. Elsevier (1996)

[HSS10] Hemaspaandra, E., Schnoor, H., Schnoor, I.: Generalized modal satisfiability. J. Comput. Syst. Sci. 76(7), 561–578 (2010)

[Lew79] Lewis, H.: Satisfiability problems for propositional calculi. Mathematical Systems Theory 13, 45–53 (1979)

[LV10] Lohmann, P., Vollmer, H.: Complexity Results for Modal Dependence Logic. In: Dawar, A., Veith, H. (eds.) CSL 2010. LNCS, vol. 6247, pp. 411–425. Springer, Heidelberg (2010)

[Mak86] Maksimova, L.: On maximal intermediate logics with the disjunction property. Studia Logica 45(1), 69–75 (1986)

[PV05] Parikh, R., Väänänen, J.: Finite information logic. Annals of Pure and Applied Logic 134(1), 83–93 (2005), Papers Presented at the 9th Workshop on Logic, Language, Information and Computation (WoLLIC 2002)

[Ser81] Servi, G.F.: Semantics for a class of intuitionistic modal calculi. In: Dalla Chiara, M.L. (ed.) Italian Studies in the Philosophy of Science, pp. 59–72. D. Reidel Publishing Company (1981)

[Sev09] Sevenster, M.: Model-theoretic and computational properties of modal dependence logic. Journal of Logic and Computation 19(6), 1157–1173 (2009)

[Vää07] Väänänen, J.: Dependence logic: A new approach to independence friendly logic. London Mathematical Society Student Texts, vol. 70. Cambridge University Press (2007)

[Vää08] Väänänen, J.: Modal dependence logic. In: Apt, K.R., van Rooij, R. (eds.) New Perspectives on Games and Interaction. Texts in Logic and Games, vol. 4, pp. 237–254. Amsterdam University Press (2008)

[Wal70] Walkoe, W.: Finite partially-ordered quantification. Journal of Symbolic Logic (35), 535–555 (1970)

[Yan10] Yang, F.: Expressing second-order sentences in intuitionistic dependence logic. In: Dependence and Independence in Logic Proceedings, pp. 118–132 (2010)

[Yan12] Yang, F.: Modal intuitionistic dependence logic (2012) (manuscript)

Appendix

Lemma 2. *There is a* PSPACE *algorithm for $\mathcal{MIDL}$-MC.*

Proof. As it is outlined in the proof of Theorem 2, we give an algorithm that runs in polynomial time on an alternating Turing machine.

Algorithm 1.1. check($K = (S, R, \pi)$,φ,T)

```
case φ
when φ = p
  foreach s ∈ T
    if not p ∈ π(s) then
      return false
  return true

when φ = ¬p
  foreach s ∈ T
    if p ∈ π(s) then
      return false
  return true

when φ = dep(p_1, ..., p_n)
  foreach (s, s') ∈ T × T
   if π(s) ∩ {p_1, ..., p_{n-1}} = π(s') ∩ {p_1, ..., p_{n-1}} then
          // i.e., s and s' agree on the valuations of p_1, ..., p_{n-1}
    if (q ∈ π(s) and not q ∈ π(s')) or (not q ∈ π(s) and q ∈ π(s')) then
      return false
  return true

when φ = ¬dep(p_1, ..., p_n)
  if S = ∅
    return true
  return false

when φ = ψ ∧ χ
  return (check(K,T,ψ) and check(K,T,χ))

when φ = ψ ∨ χ
  existentially guess two sets of states T_1, T_2 ⊆ S
  if not T_1 ∪ T_2 = T then
    return false
  return (check(K,T_1,ψ) and check(K,T_2,χ))

when φ = ψ ⩒ χ
  return (check(K,T,ψ) or check(K,T,ψ))

when φ = □ψ
  T' := ∅
  foreach s' ∈ S
    foreach s ∈ T
```

```
      if (s, s') ∈ R then
        T' := T' ∪ {s'}
              // T' is the set of all successors of all states in T, i.e. T' = R(T)
  return check(K, T', ψ)

when φ = ◊ψ
  existentially guess a set of states T' ⊆ S
    foreach s ∈ T
      if there is no s' ∈ T' with (s, s') ∈ R then
        return false
              // T' contains at least one successor for every state in T, i.e. TRT'
  return check(K, T', ψ)

when φ = ψ → χ
  universally guess a set of states T' ⊆ T
  if not check(K, ψ, T') or check(K, χ, T')
    return true
  return false
```

□

Coalgebraic Predicate Logic: Equipollence Results and Proof Theory

Tadeusz Litak[1,⋆], Dirk Pattinson[2], and Katsuhiko Sano[3]

[1] FAU Erlangen-Nürnberg
[2] Imperial College London
[3] Japan Advanced Institute of Science and Technology

Abstract. The recently introduced Coalgebraic Predicate Logic (CPL) provides a general first-order syntax together with extra modal-like operators that are interpreted in a coalgebraic setting. The universality of the coalgebraic approach allows us to instantiate the framework to a wide variety of situations, including probabilistic logic, coalition logic or the logic of neighbourhood frames. The last case generalises a logical setup proposed by C.C. Chang in early 1970's. We provide further evidence of the naturality of this framework. We identify syntactically the fragments of CPL corresponding to extended modal formalisms and show that the full CPL is equipollent with coalgebraic hybrid logic with the downarrow binder and the universal modality. Furthermore, we initiate the study of structural proof theory for CPL by providing a sequent calculus and a cut-elimination result.

1 Introduction

Coalgebras over sets provide an universal framework for state-based systems, such as (labelled or unlabelled) transition systems, multigraphs, conditional frames, game frames or (monotone and general) neighbourhood frames. They provide a natural semantics for a wide range of modal logics, ranging from conditional and probabilistic to coalition logic. The development of a full-blown coalgebraic model and correspondence theory is hindered by the lack of a formalism that allows both direct reference to individual states and supports universal quantification and binding: a coalgebraic counterpart of first-order (and higher-order) predicate logic. The framework of coalgebraic predicate logic (CPL) was introduced recently in [9] in a joint paper with Lutz Schröder, where we have provided a complete Hilbert axiomatisation, a modal correspondence theorem and some basic model-theoretic constructions. The present paper is intended as a companion to *op.cit.* presenting more evidence that coalgebraic predicate logic is a natural extension of both (coalgebraic) modal logic and first-order logic.

As explained in *op.cit.*, our approach can be traced back to an unjustly forgotten paper [7] by C. C. Chang. The original motivation was to simplify model theory for what

⋆ The first author gratefully acknowledges the support of the EPSRC grant EP/G041296/1 while employed at the University of Leicester. Thanks are also due to the anonymous referee for careful reading and many insightful comments and to Guram Bezhanishvili for his patience during the editing process.

G. Bezhanishvili et al.(Eds.): TbiLLC 2011, LNCS 7758, pp. 257–276, 2013.

Montague called *pragmatics* and replace Montague's many-sorted setting by one without sorts. In contemporary terminology, Chang's paper deals with model and correspondence theory for *neighbourhood frames*: coalgebras for double contravariant powerset functor (see [8] for a coalgebraic treatment in a two-sorted setting). His constructions and results include both suitable notions of (elementary) submodel/extension, elementary chain of models and ultraproduct and suitable variants of Tarski-Vaught, downward and upward Löwenheim-Skolem and compactness theorems. Curiously, Chang did not prove any completeness results. Apart from these technical developments, advantages of [7] include its lucid, intuitive motivation and examples. But the biggest interest lies in the syntax itself, with only one sort of variables for elements of the state space and no need for explicit quantification over neighbourhood or successors. Apart from a number of papers (e.g., Sgro [16]) on interior operator logic in topology, we are not aware of any work in a similar setup.

Our CPL (Coalgebraic Predicate Logic) is based on a notational variant of Chang's syntax. The interpretation of CPL in coalgebras for arbitrary Set-functors is parametric in the notion of *predicate lifting*; if Λ is a supply of modal operators, then the supply of functors and predicate liftings interpreting modalities in Λ is called Λ-*structure*. In [9], the authors together with Lutz Schröder have proved completeness results for two classes of Λ-structures—the first one (S1SC) generalizing neighbourhood frames, another one (k-bounded) generalizing Kripke structures. Furthermore, it was shown that there are limitations to the possible scope of more general completeness results. We also proved an analogue of the Van Benthem-Rosen theorem, characterizing coalgebraic modal logic (CML) as the behavioural-invariant fragment of CPL. Finally, we provided foundations of model theory for CPL, significantly generalizing the scope of Chang's model-theoretic results discussed above.

This paper is intended as a companion paper to [9]. In the first part, we generalize the results of [4] characterizing the correspondence between predicate logic and extended hybrid formalisms $\mathcal{CHL}_{\Lambda}^{\downarrow,\mathsf{A}}$ and $\mathcal{CHL}_{\Lambda}^{\forall,@}$. We take it as yet another indication that CPL is natural and well-designed both as a generalization of FOL and "the" predicate relative of existing coalgebraic formalisms. Furthermore, due to a somewhat modal character of the CPL syntax, the correspondence is even closer and more natural than for ordinary FOL and additional results on the correspondence between sublanguages of CPL and various extensions of coalgebraic modal/hybrid logic become available. In the second part, we initiate the study of proof theory of CPL. We provide cut-free sequent systems for strongly one-step complete (S1SC) Λ-structures, which generalize the neighbourhood logic (and hence Chang's original setup). Our proof of cut-elimination is entirely syntactic and constructive.

2 Syntax and Semantics

Throughout the paper, we fix a modal similarity type Λ consisting of modal operators with associated arities. We also fix a set Σ of *predicate symbols* (with associated arities) and a set V^i of *individual variables*. The *language* $\mathcal{CPL}_\Lambda$ of coalgebraic predicate logic is given by the grammar

$$\mathcal{CPL}_\Lambda \ni \phi, \psi ::= x = y \mid P(x_1, \dots, x_n) \mid \bot \mid \phi \to \psi \mid \forall x \phi \mid x \heartsuit \lceil x_1 : \phi_1 \rceil \dots \lceil x_n : \phi_n \rceil$$

where $\heartsuit \in \Lambda$ is an n-ary modal operator, $P \in \Sigma$ is an n-ary predicate symbol and $x, y, x_1, \dots, x_n \in \mathsf{V}^i$ are individual variables. For simplicity, we ignore function symbols which can be added in the same way as in [7].

In a formula $x\heartsuit\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil$, $\heartsuit$ is an n-ary modal operator, applied to n arguments $\lceil x_i : \phi_i\rceil$, for $i = 1, \dots, n$. Here, x_i is a *comprehension variable*. Given a first-order structure with carrier set W and variable assignment ϑ, $\lceil x : \phi\rceil$ denotes the set of all those states $w \in W$ such that ϕ holds under the modified assignment $\vartheta[x \mapsto w]$. Our informal reading of $\lceil x : \phi\rceil$ is 'the set of all x such that ϕ'. As a consequence, the n-tuple $\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil$ denotes an n-tuple of predicates on the carrier set, to which we can apply an n-ary modal operator $\heartsuit$ in the usual way. The formula $x\heartsuit\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil$ is then best understood as expressing that the property $\heartsuit\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil$ is true relative to the (interpretation of) x in a first-order structure.

Example 1. We have provided a number of examples of the use of CPL in a variety of situations already in [9] where consider CPL over probabilistic modal logic, over coalition logic and Pressburger modal logic. Here, we content ourselves with the following:

1. As originally noted by by Chang himself, coalgebraic predicate logic is particularly well-suited for reasoning about social situations and relationships between an individual and sets of individuals. Indeed, Chang's examples suggest an *intensional* reading of $\heartsuit$ as 'useful' or 'enjoyable'. Given a unary modality $\Box$ and a binary relation $S(x, y)$ that we read as 'x speaks to y', the formula $x\Box\lceil z : S(z, y)\rceil$ reads as 'x finds it enjoyable to speak to y' where x determines the truth of this sentence by inspecting the the set '$\{z : S(z, y)\}$' of people speaking to y. The fact that whether or not x finds it enjoyable to speak to y may depend *non-monotonically* on the set of people y converses with suggests to interpret $\Box$ as a neighbourhood modality (as we will in fact do in Example 4).

2. Coalgebraic predicate logic can also be used to speak about 'losers', 'jerks' and 'politicians'. In [2], these terms are defined using hybrid logic over Kripke semantics where the underlying binary relation is understood as '*respects*' or '*admires*'. For example, a loser is understood as a person who lacks self-respect. In coalgebraic predicate logic, the fact that x is a loser is expressed by the formula $x\Box\lceil y : \neg(y = x)\rceil$. We read this formula as 'everybody whom x respects has the property of being distinct from x', i.e. x lacks self-respect. Accordingly, our interpretation of $\Box$ (given in Detail in Example 4) in this example will be relational, and coincides with the Kripke-interpretation over relational models. We leave it to the reader to express their own (or [2]'s) notions of 'jerks' and 'politicians' in coalgebraic predicate logic. In Section 3, we will show that coalgebraic hybrid logic is in fact equi-expressive to coalgebraic predicate logic.

3. Coalgebraic predicate logic also extends, for instance, majority logic [11] to a first-order setting. If we take x to be a politician if the majority of people known to them distrusts x, then the fact that x is a politician is expressed by the formula $x\mathbf{M}\lceil y : D(x, y)\rceil$ where $\mathbf{M}$ is the majority operator that we read 'the majority of' (and assume that majorities are taken among people that are known to an individual) and $D(x, y)$ is a

binary relation that expresses that x distrusts y. We will make this semantically precise, in the next example, by interpreting $\mathbf{M}$ as the majority operator of [11].

The semantics of coalgebraic predicate logic is given, as usual, in terms of a *first order structure* and a *variable assignment*. Crucially, the first-order structure must accommodate for the interpretation of the modalities present in the similarity type Λ and must provide a *uniform* interpretation of modalities. We therefore understand first-order structures for Λ as an enrichment of the standard notion of first-order structure with a device to interpret modalities. In our (coalgebraic) context, the interpretation of modal operators is given by Λ-structures.

Definition 2. A Λ-structure is an endofunctor $T : \mathsf{Set} \to \mathsf{Set}$ on the category of sets, together with an assignment of *predicate liftings*, that is, a set-indexed family of maps

$$[\![\heartsuit]\!]_X : (\mathcal{Q}X)^n \to \mathcal{Q}(TX) \qquad (X \in \mathsf{Set})$$

for every n-ary modal operator $\heartsuit \in \Lambda$. Here $\mathcal{Q}$ is the *contravariant* powerset functor, and we require naturality of $[\![\heartsuit]\!]$, that is, $(Tf)^{-1} \circ [\![\heartsuit]\!]_Y = \heartsuit_X \circ (f^{-1})^n$ for every function $f : X \to Y$. We usually denote Λ-structures by the underlying endofunctor T, when the underlying assigned predicate liftings are clear.

In the remainder of this paper, we assume that our chosen set Λ of modal operators comes equipped with a Λ-structure. We now take first-order structures to be T-*coalgebras* that are additionally equipped with an interpretation of the given relation symbols.

Definition 3. A *first-order structure* (for Λ relative to a Λ-structure T) is a triple $\mathfrak{M} = (C, \gamma, \pi)$ where (C, γ) is a T-*coalgebra*, i.e. C is a set and $\gamma : C \to TC$ a (transition) function, and π is an interpretation of the predicate symbols, that is, $\pi(P) \subseteq C^n$ for all n-ary predicate symbols $P \in \Sigma$. We silenty identify T-coalgebras with first-order structures and leave the interpretation of predicate symbols implicit whenever this does not cause confusion.

The semantics of coalgebraic predicate logic is best explained as the algamation of coalgebraic modal logic and (standard) first-order logic. Given a T-coalgebra (C, γ), formula ϕ of coalgebraic modal logic are interpreted as subsets $[\![\phi]\!] \subseteq C$. The crucial clause for modal operators is $[\![\heartsuit(\phi_1, \dots, \phi_n)]\!] = \{c \in C \mid \gamma(c) \in [\![\heartsuit]\!]_C([\![\phi_1]\!], \dots, [\![\phi_n]\!])\}$, discussed in detail e.g. in [13]. Informally speaking, the (coalgebraic) interpretation of $\heartsuit(\phi_1, \dots, \phi_n)$ is the set of individuals $c \in C$ that enjoy property $\heartsuit$ (which depends on $\phi_1, \dots, \phi_n$). In coalgebraic *predicate* logic, this interpretation is relativised to *individuals*: in a first-order structure $\mathfrak{M} = (C, \gamma, \pi)$, the formula $x\heartsuit\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil$ is true under a variable assigment ϑ if the individual $\vartheta(x)$ has property $\heartsuit$ which now depends on the sets of individuals x_i that have property $\phi_i(x_i)$.

Formally, we define *truth* $\mathfrak{M}, \vartheta \models \phi$ of a formula $\phi \in \mathcal{CPL}_\Lambda$ in a first-order structure $\mathfrak{M} = (C, \gamma, \pi)$ relative to a variable assignment $\vartheta : \mathsf{V}^i \to C$ by the standard clauses for propositional and first-order connectives, augmented with

$$\mathfrak{M}, \vartheta \models x\heartsuit\lceil x_1 : \phi_1\rceil \dots \lceil x_n : \phi_n\rceil \iff \gamma(\vartheta(x)) \in [\![\heartsuit]\!]_X([\![\phi_1]\!]^{x_1}_{\mathfrak{M},\vartheta}, \dots, [\![\phi_n]\!]^{x_n}_{\mathfrak{M},\vartheta})$$

where $[\![\phi]\!]^y_{\mathfrak{M},\vartheta} = \{c \in C \mid \mathfrak{M}, \vartheta[y \mapsto c] \models \phi\}$ (we usually drop the subscript $\mathfrak{M}, \vartheta$) and $\vartheta[y \mapsto c]$ is the same variable assignment as ϑ except it maps y to c.

Example 4. We continue Example 1 and describe the structures that give rise to the interpretation put forward above.

1. Chang's original attempt generalises the neighbourhood interpretation of modal logics to the setting of full first-order logic. For a similarity type $\Lambda = \{\Box\}$ containing a single unary operator, neighbourhood semantics is captured coalgebraically by the Λ-structure $\mathcal{N} = \mathcal{Q} \circ \mathcal{Q}$ together with the predicate lifting defined by

$$[\![\Box]\!]_X(A) = \{N \in \mathcal{N}X \mid A \in N\}$$

   which ensures that the standard modal neighbourhood semantics conincides with the coalgebraic semantics of modal formulae. In a first-order setting, this exhibits Chang's original language (and its interpretation) as a special case of coalgebraic predicate logic. In a first-order structure (C, γ), every individual $c \in C$ induces a set $\gamma(c) \subseteq \mathcal{P}(C)$ of neighbourhoods such that – in the spirit of the example – x finds it enjoyable to speak to y if the set $\lceil z : S(z, y)\rceil$ is a (social) neighbourhood of x.
2. In [2], hybrid logic is used to define *losers*, *jerks* and *politicians*, where notions like *respect* or *admiration* are modelled by binary relations between individuals. We replace relations by T-coalgebras for $TC := \mathcal{P}C$ ($\mathcal{P}$ is the *covariant* powerset functor) and interpret the (unary) model operator $\Box$ by $[\![\Box]\!]_C(A) := \{B \in \mathcal{P}C \mid B \subseteq A\}$ which gives the standard semantics. The formula $x\Box\lceil y : \neg(y = x)\rceil$ then expresses that x is a loser, i.e. lacks an arc along the relation expressing self-respect. We leave it to the reader to express the definitions of jerks and politicians given in [2]). Indeed, hybrid logic is translatable to the language discussed here, see below.
3. As a slight variation, we may consider a predicate version of majority logic [11] where we again co-algebraise the relational semantics. We interpret formulae involving an operator $\mathbf{M}$ (read 'the majority ...') over coalgebras of type $(C, \gamma : C \to \mathcal{B}C)$ where $\mathcal{B}C := \{f : C \to \mathbb{N} \mid f(c) \neq 0 \text{ only finitely often}\}$ using

$$[\![\mathbf{M}]\!]_C(A) := \{f \in \mathcal{B}C \mid \sum_{x \in A} f(x) > \sum_{x \notin A} f(x)\}.$$

   This differs from the original semantics of *op.cit.* but induces the same set of true sentences. If we read $\mathbf{M}$ as *the majority of people someone knows* ... and $R(x, y)$ as *likes*, an unpopular person could be characterised by the sentence $x\mathbf{M}\{y : \neg R(y, x)\}$ stipulating that the majority of people x knows don't like x.
4. Frame classes can be combined: instead of using the relation symbol R in the previous example, we could consider coalgebras $(C, \gamma : C \to TC)$ where $TC := \mathcal{B}C \times \mathcal{P}C$ gives a majority structure and a relational structure, and interpret the operators $\mathbf{M}$ and $\Box$ by projecting out the components. Unpopular individuals are then characterised as satisfying $x\mathbf{M}\lceil y : y\Box\lceil z : \neg(z = x)\rceil\rceil$.

3 Equipollence Results

3.1 Coalgebraic Standard Translation for $\mathcal{CML}_\Lambda$

The formulas $\mathcal{CML}_\Lambda(\Sigma)$ of pure (coalgebraic) modal logic in the modal signature Λ over Σ (now all elements of Σ are assumed to be of arity 1) are given by the grammar:

$$\mathcal{CML}_\Lambda \quad \phi, \psi ::= P \mid \bot \mid \phi \to \psi \mid \heartsuit(\phi_1, \ldots, \phi_n),$$

where $P \in \Sigma$.

Satisfaction is defined with respect to $\mathfrak{M} = (C, \gamma, \pi)$ and a specific point $c \in C$ in a standard way, see e.g. [14,15].

Proposition and Definition 5. *Define the* coalgebraic standard translation *as*

$$\begin{aligned} ST_x(P) &:= P(x), \\ ST_x(\heartsuit(\phi_1, \ldots, \phi_n)) &:= x\heartsuit\lceil x : ST_x(\phi_1)\rceil \ldots \lceil x : ST_x(\phi_n)\rceil, \\ ST_x(\bot) &:= \bot, \\ ST_x(\phi \to \psi) &:= ST_x(\phi) \to ST_x(\psi). \end{aligned}$$

Then for any $\phi \in \mathcal{CML}_\Lambda(\Sigma)$ and any $\mathfrak{M} = (C, \gamma, \pi), \vartheta, c$, we have $\mathfrak{M}, c \vDash \phi$ iff $\mathfrak{M}, \vartheta[x \mapsto c] \vDash ST_x(\phi)$.

This definition is more straightforward than the standard translation into FOL of modal logic over ordinary Kripke frames. Moreover, ST_x uses only one variable from V^i, namely x itself. In fact, we can immediately observe that

Proposition 6. *Whenever Σ consists entirely of* unary *predicate symbols, the subset of $\phi \in \mathcal{CPL}_\Lambda(\Sigma)$ obtained as the image of ST_x for a fixed $x \in \mathsf{V}^i$ consists precisely of equality-free and quantifier-free formulae in the variable x.*

Whereas the Van Benthem-Rosen theorem provided in [9] is a semantic characterization of $\mathcal{CML}_\Lambda$ wrt $\mathcal{CPL}_\Lambda$, Proposition 6 above is its syntactic counterpart. In fact, we can combine the two results to obtain.

Corollary 7. *Whenever Σ consists entirely of* unary *predicate symbols (and there are no function symbols), the behaviourally-invariant (over finite structures) formulas of $\mathcal{CPL}_\Lambda$ in one-free variable are up to equivalence (over finite structures) precisely the equality-free and quantifier-free formulas in the single-variable fragment of $\mathcal{CPL}_\Lambda$.*

No such syntactic characterization exists for formulas of ordinary first-order logic invariant under bisimulation. Of course, we can do better thanks to the somewhat more modal character of CPL syntax as compared to ordinary FOL.

3.2 Hybrid Languages

In this section, we generalize the results of [4]. Our ultimate goal is Theorem 13 below which establishes the equivalence of CPL with the hybrid languages $\mathcal{CHL}_\Lambda^{\downarrow,\mathsf{A}}$ and

$\mathcal{CHL}_{\Lambda}^{\forall,@}$. Both correspondences also hold for ordinary predicate logic over relational structures (FOL) and extend to CPL. We take this as yet another indication that CPL is natural and well-designed both as a generalization of FOL and "the" predicate logic cousin of existing coalgebraic formalisms.

This is our main, but not the only motivation. We progress towards this result step-by-step, extending the modal language gradually with new hybrid constructs. In this way, we reveal that a similar correspondence exists between natural fragments of CPL and weaker hybrid languages, most importantly between quantifier-free CPL and $\mathcal{CHL}_{\Lambda}^{\downarrow,@}$—something which has no analogue in the FOL case.

Again, obviously the correspondence between fragments of CPL and extensions of CML is tighter than in the case of FOL and ML only due to the modal flavour of CPL. However, results such as Corollary 10 are useful spadework: any model-theoretic tool to be developed—say, a variant of E-F games—would be adequate for an extended coalgebraic modal formalism (e.g., $\mathcal{CHL}_{\Lambda}^{\downarrow,@}$) **iff** it is adequate for the corresponding fragment of CPL (e.g., the variable-free fragment), so we are free to work with whichever formalism we find more convenient at a given moment. This is closely related to our present research efforts. The straightforward correspondence also provides a good starting point for an extension of research programme sketched in [5]—see Remark 14 at the end of this section.

Given a supply of *world variables* V^w that we are going to keep fixed and implicit, we define the following *coalgebraic hybrid languages*

$\mathcal{CHL}_{\Lambda}^{\downarrow,@}$	$\phi,\psi ::= \; z \mid P \mid \bot \mid \phi \to \psi \mid \heartsuit(\phi_1,\ldots,\phi_n) \mid @_z\phi \mid \downarrow z.\phi$
$\mathcal{CHL}_{\Lambda}^{\downarrow,\mathsf{A}}$	$\phi,\psi ::= \; z \mid P \mid \bot \mid \phi \to \psi \mid \heartsuit(\phi_1,\ldots,\phi_n) \mid \mathsf{A}\phi \mid \downarrow z.\phi$
$\mathcal{CHL}_{\Lambda}^{\forall,@}$	$\phi,\psi ::= \; z \mid P \mid \bot \mid \phi \to \psi \mid \heartsuit(\phi_1,\ldots,\phi_n) \mid @_z\phi \mid \forall z.\phi$

where $z \in \mathsf{V}^w$. We refer the reader to, e.g, [15,4,5] for the semantics. The extension of the standard translation to these formalism is unproblematic in some cases, just like in the case of ordinary hybrid logic over Kripke frames:

$$ST_x(z) := x = z, \quad ST_x(\mathsf{A}\phi) := \forall x.ST_x(\phi), \quad ST_x(\forall z.\phi) := \forall z.ST_x(\phi).$$

One is tempted to put forward also

$$ST_x(@_z\phi) := ST_x(\phi)[z/x], \quad ST_x(\downarrow z.\phi) := ST_x(\phi)[x/z].$$

However, with other clauses remaining the same, this could work only if $[z/x]$ denotes *capture-avoiding* substitution. Sadly, this in turn would entail forsaking the luxury of using just one designated variable for comprehension. Guillame Malod (see [6]) observed that if we restrict the supply of variables, a translation along the above lines—indeed first proposed in the literature (which shows that the present discussion is less trivial than it might seem)—would fail even when embedding the hybrid logic over Kripke frames in the two-variable fragment of FOL. Malod's counterexample used nesting of modalities of level two, but as our translation uses just one designated variable, ST would go wrong already on formulas of depth one. Just consider $ST_x(\downarrow z.\Diamond z)$: we would obtain $x\Diamond\lceil x : x = x\rceil$, which is a formula with a completely different meaning.

There are two ways out. First is to redefine

$$STmod_x(@_z\phi) := \forall x.(x = z \to ST_x(\phi)), \tag{1}$$
$$STmod_x(\downarrow z.\phi) := \forall z.(x = z \to ST_x(\phi)). \tag{2}$$

The second is to keep ST for hybrid formulas as defined above and change the modal clause instead:

$$ST_x(\heartsuit(\phi_1, \dots, \phi_n)) := x\heartsuit\lceil y : ST_y(\phi_1)\rceil \dots \lceil y : ST_y(\phi_n)\rceil, \tag{3}$$

where y is the first (in some fixed enumeration) variable *not used* in $ST_x(\phi_1), \dots, ST_x(\phi_n)$; by *not used* here we mean both free and bound usage. Furthermore, to ensure that the translation works correctly, we have to assume that *neither* x *nor* y appears in V^w. While the requirement to use more bound variables can be cumbersome—particularly for infinite sets of formulas—we prefer this option, as it makes it easier to characterize weaker hybrid languages as suitable syntactic fragments of CPL.

We can now state a generalization of both Proposition 5 and corresponding results from the hybrid logic literature—see, e.g., [4] for references:

Proposition 8. *For any $\phi \in \mathcal{CHL}_\Lambda$ and any $\mathfrak{M} = (C, \gamma, \pi), \vartheta, c$, we have $\mathfrak{M}, \vartheta, c \vDash \phi$ iff $\mathfrak{M}, \vartheta[x \mapsto c] \vDash ST_x(\phi)$.*

As is well-known in the hybrid logic community—see again [4] for references—there is also a translation in the reverse direction for sufficiently expressive hybrid languages. This also generalizes to our setting, see Table 1.

Table 1. Coalgebraic Hybrid Translation from quantifier-free CPL to $\mathcal{CHL}_\Lambda^{\downarrow @}$

$HT(P(x)) := @_x P$	$HT(x = y) := @_x y$	
$HT(\bot) := \bot$	$HT(\phi \to \psi) := HT(\phi) \to HT(\psi)$	
$HT(x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil)$	$:=$	$@_x\heartsuit(\downarrow y_1.HT(\phi_1), \dots, \downarrow y_n.HT(\phi_n))$

Proposition 9. *For any $\phi \in \mathcal{CPL}_\Lambda$ and any $\mathfrak{M} = (C, \gamma, \pi), \vartheta, c$, we have*

$$\mathfrak{M}, \vartheta, c \vDash HT(\phi) \text{ iff } \mathfrak{M}, \vartheta[x \mapsto c] \vDash \phi.$$

Combining Propostions 9 and 8, we get:

Corollary 10. *Whenever Σ consists purely of unary predicates (and no function symbols), $\mathcal{CHL}_\Lambda^{\downarrow @}$ is expressively equivalent to the quantifier-free fragment of $\mathcal{CPL}_\Lambda$, assuming V^i contains V^w plus a disjoint infinite supply of additional individual variables (used for comprehension).*

Remark 11 (Quantifier-free CPL as the bounded fragment of FOL). In the case of ordinary FOL, the fragment equivalent to $\mathcal{CHL}_\Lambda^{\downarrow @}$ is characterized as the *bounded* *fragment*, see, e.g., [1]. In fact, our formula $x\heartsuit\lceil y : \phi\rceil$, despite being quantifier-free on the surface, can be described as a form of bounded quantification. This can be formalized as a result stating that over coalgebras for the covariant powerset functor (Kripke frames), quantifier-free $\mathcal{CPL}_\Lambda$ is equivalent to the bounded-fragment of ordinary FOL, where the role of $\heartsuit$ in $\mathcal{CPL}_\Lambda$ is played by the binary relation symbol R in FOL; details are left to the reader.

Remark 12 (Chang's original syntax). As already mentioned, our syntax is slightly different to the original one proposed by Chang [7]. In that paper, there were no explicit comprehension variables and even in the enriched syntax which allowed constants and function terms, the term on the left-hand side of $\heartsuit$ had to be a variable. This variable was reused then on the right side of $\heartsuit$ as the comprehension variable. In other words, Chang's $x\heartsuit\phi(x)$ was equivalent to ours $x\heartsuit\lceil x : \phi(x)\rceil$. In presence of quantifiers, which can be used to simulate the effect of capture-avoiding substitution as in $STmod$ (this trick in fact stems back to Alfred Tarski), the two languages are obviously equivalent. But when considering fragments, as we do here, the equivalence breaks down; without quantifiers, Chang's syntax does not allow (2) and simple renaming of the comprehension variable on the right-hand side of $\heartsuit$ as in (3) is not possible either.

There are two usual routes in hybrid logic to achieve full first-order expressivity. One is to add universal quantifiers over V^w in presence of the satisfaction operator @. The other is to add the global modality A in presence of the downarrow binder $\downarrow$. The hybrid translation is extended then as follows:

$$HT_{\forall @}(\forall x.\phi) := \forall x.HT(\phi)$$
$$HT_{\mathsf{A}\downarrow}(\forall x.\phi) := \downarrow y.\mathsf{A} \downarrow x.\mathsf{A}(y \to \phi)$$

In $HT_{\mathsf{A}\downarrow}$ we need the proviso that y is not occurring in the whole formula.

Theorem 13. *$\mathcal{CHL}_\Lambda^{\downarrow\mathsf{A}}$, $\mathcal{CHL}_\Lambda^{\forall,@}$ and $\mathcal{CPL}_\Lambda$ are expressively equivalent.*

As we can use $STmod_x$ now and keep reusing x as the comprehension variable, it is enough to assume that $\mathsf{V}^i = \mathsf{V}^w \cup \{x\}$. Since $@_z\phi$ is definable in presence of A (as $\mathsf{A}(z \to \phi)$, $\downarrow$is definable by the universal quantifier over V^w (as $\forall z.(z \to \phi)$) and A is definable by combination of $\forall$ and @ (as $\forall y.@_y\phi$, where y is not used in ϕ), we get in fact seven equivalent languages: $\mathcal{CPL}_\Lambda$, Chang's original language, $\mathcal{CHL}_\Lambda^{\downarrow\mathsf{A}}$, $\mathcal{CHL}_\Lambda^{\forall,@}$, $\mathcal{CHL}_\Lambda^{\downarrow\mathsf{A}}$ with @, $\mathcal{CHL}_\Lambda^{\forall,@}$ with $\downarrow$and the jumbo hybrid language with all connectives introduced above.

Remark 14. The equivalences stated here extend to the case of $\mathcal{CHL}_\Lambda$ and $\mathcal{CPL}_\Lambda$ enriched with quantification over predicates (i.e., second-order languages). It would be interesting to follow more thoroughly the program of *coalgebraic abstract model theory* both above and below $\mathcal{CPL}_\Lambda$ (see Ten Cate's PhD Thesis [5] for spadework in abstract model theory below first-order logic).

4 Proof Theory

4.1 Axiomatisation of Coalgebraic Predicate Logic

This paper's companion [9] already gives a complete Hilbert calculus for coalgebraic predicate logic that we review briefly here before giving an axiomatisation in terms of a cut-free sequent system. The crucial aspect of this system (and also of the sequent system that we will describe) are *one-step rules* that describe the geometry of the Λ-structure under consideration.

The modularity of coalgebraic predicate logic in the precise notion of structure necessitates that we cannot commit to a fixed set of rules. Instead, we import modal deduction rules into a first-order setting. It can be shown [13] that these deduction rules can be restricted to so-called one-step rules that have a very specific format. More precisely:

Definition 15. A ($\mathcal{CML}_\Lambda$) *one-step rule* over a similarity type Λ has the form

$$\frac{\Gamma_1 \Rightarrow \Delta_1 \qquad \cdots \qquad \Gamma_k \Rightarrow \Delta_k}{\heartsuit_1 \boldsymbol{p_1}, \ldots, \heartsuit_n \boldsymbol{p_n} \Rightarrow \heartsuit_{n+1} \boldsymbol{p_{n+1}}, \ldots, \heartsuit_{n+m} \boldsymbol{p_{n+m}}} (R)$$

where $k, n, m \geq 0$, $\heartsuit_1, \ldots, \heartsuit_{n+m} \in \Lambda$, $\boldsymbol{p_i} = (p_i^1, \ldots, p_i^{a(i)})$ are vectors of propositional variables according to the arity $a(i)$ of $\heartsuit_i$ and $\Gamma_1, \ldots, \Gamma_k, \Delta_1, \ldots, \Delta_k \subseteq \{p_i^j \mid 1 \leq i \leq n, 1 \leq j \leq a(i)\}$. We denote the conclusion of (R) by $\Gamma_R \Rightarrow \Delta_R$.

For the Hilbert-style axiomatisation of CPL, we write $\boldsymbol{x}, \boldsymbol{y}, \ldots$ for finite sequeneces of variables and define derivability $\mathcal{HR} \vdash$ as the the least set of formulae that is closed under modus ponens and contains axioms listed in Table 2. We have shown in [9] that

Table 2. Axioms of Coalgebraic Predicate Logic

EG1 all propositional tautologies
EG2 $\forall \boldsymbol{y}.(\forall x.\phi \to \phi)$
EG3 $\forall \boldsymbol{y}.(\forall x.(\phi \to \psi) \to (\forall x.\phi \to \forall x.\psi))$
EG4 $\forall \boldsymbol{y}.(\phi \to \forall x.\phi)$ if x is fresh for ϕ
EG5 $\forall \boldsymbol{y}(x = x)$
EG6.1 $\forall \boldsymbol{y}.(x = z \to P(\boldsymbol{u}, x, \boldsymbol{v}) \to P(\boldsymbol{u}, z, \boldsymbol{v})$ for $P \in \Sigma \cup \{=\}$
EG6.2 $\forall \boldsymbol{y}.(x = z \to x\heartsuit\lceil y_1 : \phi_1 \rceil \ldots \lceil y_n : \phi_n \rceil \to z\heartsuit\lceil y_1 : \phi_1 \rceil \ldots \lceil y_n : \phi_n \rceil)$
CONG $\forall \boldsymbol{y}.(\forall x.(\bigwedge_{i=1}^n (\phi_i \leftrightarrow \psi_i)) \to \forall x.(x\heartsuit\lceil x : \phi_1 \rceil \ldots \lceil x : \phi_n \rceil \leftrightarrow (x\heartsuit\lceil x : \psi_1 \rceil \ldots \lceil x : \psi_n \rceil)))$
ONESTEP(R) $\forall \boldsymbol{y}.\forall z.(\forall x.(P_1 \wedge \cdots \wedge P_k) \to C)$

In ONESTEP, R ranges over the one-step rules in $\mathcal{R}$ of the form above and $C = [\sigma, x, z](\bigwedge \Gamma_R \to \bigvee \Delta_R)$ represents the conclusion of the rule and $P_i = (\bigwedge \Gamma_i \to \bigvee \Delta_i)\sigma$ its premises, where σ sends each p_i to a formula of $\mathcal{CPL}_\Lambda$ and $[\sigma, x, z]$ is the inductive extension of the map sending each $\heartsuit_i \boldsymbol{p_i}$ to $z\heartsuit_i \lceil x : \sigma(p_i^1) \rceil \cdots \lceil x : \sigma(p_i^{a(i)}) \rceil)$.

this calculus is complete for *strongly one-step complete rule sets*. One difference is that—as we are working here with (counterparts of) sequent-style rather than of Hilbert-style one-step rules—our ONESTEP(R) has a slightly more general syntactic shape than the corresponding axiom in [9].

Definition 16. Given any supply of primitive symbols D (which can be any set), define $\mathsf{Prop}(D)$ as the set of boolean combinations of D and $\Lambda(D) = \{\heartsuit(d_1,\ldots,d_n) \mid d_1,\ldots,d_n \in D \text{ and } \heartsuit \in \Lambda \text{ is } n\text{-ary}\}$. For any $C \in \mathsf{Set}$, given a valuation $\tau : D \to \mathcal{P}(C)$, we write $C,\tau \models \alpha$ if $\tau(\alpha) = \top$ for all $\alpha \in \mathsf{Prop}(D)$. We understand $[\![\chi]\!]_{TC,\tau}$, i.e., the interpretation of $\chi \in \mathsf{Prop}(\Lambda(\mathsf{Prop}(D)))$ in the boolean algebra $\mathcal{P}(TC)$ under τ, as the inductive extension of the assignment $[\![\heartsuit(\alpha_1,\ldots,\alpha_n)]\!]_{TX,\tau} = [\![\heartsuit]\!]_C(\tau(\alpha_1),\ldots,\tau(\alpha_n))$. We write $TC,\tau \models \chi$ if $[\![\chi]\!]_{TC,\tau} = TC$, and $t \models_{TC,\tau} \chi$ if $t \in [\![\chi]\!]_{TC,\tau}$. A set $\Xi \subseteq \mathsf{Prop}(\Lambda(\mathsf{Prop}(D)))$ is *one-step satisfiable* with respect to τ if $\bigcap_{\chi\in\Xi}[\![\chi]\!]_{TC,\tau} \neq \emptyset$. If $D \subseteq \mathcal{P}(C)$ and τ is just the inclusion, we will usually drop it from the notation.

Definition 17. Let P be the set of propositional variables and $R = \Gamma_1 \Rightarrow \Delta_1, \ldots, \Gamma_k \Rightarrow \Delta_k \,/\, \Gamma_R \Rightarrow \Delta_R$ a one-step rule. We denote $\bigwedge\{\bigwedge \Gamma_i \to \bigvee \Delta_i \mid 1 \leq i \leq k\}$ ($\in \mathsf{Prop}(\mathsf{P})$) and $\bigwedge \Gamma_R \to \bigvee \Delta_R$ ($\in \mathsf{Prop}(\Lambda(\mathsf{P}))$) by $\mathrm{Prem}(R)$ and $\mathrm{Conseq}(R)$, respectively. $\mathcal{R}$ is *one-step sound* if $TC,\tau \models \mathrm{Conseq}(R)$ whenever $C,\tau \models \mathrm{Prem}(R)$ for a valuation $\tau : \mathsf{P} \to \mathcal{P}(C)$, for all $R \in \mathcal{R}$. Given a set $\mathcal{R}$ of one-step rules and a valuation $\tau : \mathsf{P} \to \mathcal{P}(C)$, a set $\Xi \subseteq \mathsf{Prop}(\Lambda(\mathsf{Prop}(\mathsf{P})))$ is *one-step consistent* (*with respect to* τ) if the set $\Xi \cup \{\mathrm{Conseq}(R)\sigma \mid \sigma : \mathsf{P} \to \mathsf{Prop}(\mathsf{P}); R \in \mathcal{R}; C,\tau \models \mathrm{Prem}(R)\sigma\}$ is propositionally consistent. A rule set $\mathcal{R}$ is *strongly one-step complete* (*S1SC*) for a Λ-structure if for every set C, any $\Xi \subseteq \mathsf{Prop}(\Lambda(\mathsf{Prop}(\mathsf{P})))$ and any $\tau : \mathsf{P} \to \mathcal{P}(C)$, Ξ is one-step satisfiable with respect to τ whenever it is one-step consistent with respect to τ. We say that a set of rules is *finitary S1SC* if the above holds whenever $\tau : \mathsf{P} \to \mathcal{P}_{\mathrm{fin}}(C)$ (but not necessarily for arbitrary τ).

The following was established in the companion paper [9].

Theorem 18. *The Hilbert-calculus is sound and complete, i.e.* $\mathcal{HR} \vdash \phi$ *if and only if* $\mathfrak{M},\vartheta \models \phi$ *for all first-order structures* $\mathfrak{M}$ *and all variable assignments* ϑ *provided* $\mathcal{R}$ *is* strongly one-step complete (*S1SC*) *and* one-step sound.

Strongly one-step complete rule sets are somewhat restrictive, but they do exist for Chang's original logic in terms of neighbourhood semantics and for coalition logic. Here, we are complementing the axiomatisation of coalgebraic predicate logic by a cut-free, complete sequent calculus. Our basis is the system **G3c** of [17] that we extend with modal rules that describe the (fixed) Λ-structure T. We take *sequents* to be pairs (Γ,Δ), written $\Gamma \Rightarrow \Delta$ where $\Gamma,\Delta \subseteq \mathcal{CPL}_\Lambda$ are finite multisets. The sequent calculus for coalgebraic predicate logic contains three types of rules: the standard logical rules for first-order logic, rules for equality and rules for the modal operators. The logical rules are standard as in Table 3. The formula introduced in the conclusion of a logical rule is called the *principal* formula of the rule.

The *equality rules* from Table 3 allow us to replace equals for equals both in predicate symbols and in modal formulae of the kind $x\heartsuit\lceil y_1 : \phi_1\rceil \ldots \lceil y_n : \phi_n\rceil$. Equality rules do not have principal formulae.

To account for the modal operators, we incorporate the one-step rules $\mathcal{R}$ into the sequent system and write ϕ_i^j for $\sigma(p_i^j)$ as in ONESTEP(R). Then, we transform the axiom into its sequent form as follow:

$$\frac{\{(\Gamma_i\sigma)[y/x] \Rightarrow (\Delta_i\sigma)[y/x] \mid 1 \leq i \leq k\} \quad y\,\text{fresh}}{z\heartsuit_1\lceil \boldsymbol{x : \phi_1}\rceil,\ldots,z\heartsuit_n\lceil \boldsymbol{x : \phi_n}\rceil \Rightarrow z\heartsuit_{n+1}\lceil \boldsymbol{x : \phi_{n+1}}\rceil,\ldots,z\heartsuit_{n+m}\lceil \boldsymbol{x : \phi_{n+m}}\rceil}$$

Table 3. Sequent Rules of Coalgebraic Predicate Logic

Logical Rules

$$(\mathsf{Ax})\,\phi,\Gamma\Rightarrow\Delta,\phi\quad(\phi\text{ atomic})\qquad(\mathsf{L}\bot)\,\bot,\Gamma\Rightarrow\Delta$$

$$(\mathsf{L}\wedge)\,\frac{\phi,\psi,\Gamma\Rightarrow\Delta}{\phi\wedge\psi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{R}\wedge)\,\frac{\Gamma\Rightarrow\Delta,\phi\qquad\Gamma\Rightarrow\Delta,\psi}{\Gamma\Rightarrow\Delta,\phi\wedge\psi}$$

$$(\mathsf{L}\vee)\,\frac{\phi,\Gamma\Rightarrow\Delta\qquad\psi,\Gamma\Rightarrow\Delta}{\phi\vee\psi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{R}\vee)\,\frac{\Gamma\Rightarrow\Delta,\phi,\psi}{\Gamma\Rightarrow\Delta,\phi\vee\psi}$$

$$(\mathsf{L}\!\to)\,\frac{\Gamma\Rightarrow\Delta,\phi\qquad\psi,\Gamma\Rightarrow\Delta}{\phi\to\psi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{R}\!\to)\,\frac{\phi,\Gamma\Rightarrow\Delta,\psi}{\Gamma\Rightarrow\Delta,\phi\to\psi}$$

$$(\mathsf{L}\forall)\,\frac{\forall x\phi,\phi[y/x],\Gamma\Rightarrow\Delta}{\forall x\phi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{R}\forall)\,\frac{\Gamma\Rightarrow\Delta,\phi[y/x]\quad y\text{ fresh}}{\Gamma\Rightarrow\Delta,\forall x\phi}$$

$$(\mathsf{L}\exists)\,\frac{\phi[y/x],\Gamma\Rightarrow\Delta\quad y\text{ fresh}}{\exists x\phi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{R}\exists)\,\frac{\Gamma\Rightarrow\Delta,\phi[y/x],\exists x\phi}{\Gamma\Rightarrow\Delta,\exists x\phi},$$

where $\phi[y/x]$ stands for the formula ϕ with all free occurrences of x replaced by y and the assumption of freshness means not occurring in the lower sequent of the rule.

Equality Rules

$$(\mathsf{Ref})\,\frac{x=x,\Gamma\Rightarrow\Delta}{\Gamma\Rightarrow\Delta}\qquad(\mathsf{Repl})\,\frac{\phi[y/z],x=y,\phi[x/z],\Gamma\Rightarrow\Delta}{x=y,\phi[x/z],\Gamma\Rightarrow\Delta}$$

In (Repl), ϕ is resticted to the atomic formulae.

$$(\mathsf{Ren})\,\frac{x=y,x\heartsuit\lceil z:\phi_1\rceil\dots\lceil z:\phi_n\rceil,y\heartsuit\lceil z:\phi_1\rceil\dots\lceil z:\phi_n\rceil,\Gamma\Rightarrow\Delta}{x=y,x\heartsuit\lceil z:\phi_1\rceil\dots\lceil z:\phi_n\rceil,\Gamma\Rightarrow\Delta}$$

Modal Rules $\mathcal{S}(\mathcal{R})$: for every one-step rule of the form $R\in\mathcal{R}$,

$$\mathcal{S}(R)\,\frac{\{(\Gamma_i\sigma)[y/x],\Gamma^+\Rightarrow\Delta^+,(\Delta_i\sigma)[y/x]\mid 1\le i\le k\}\quad y\text{ fresh}}{\Sigma,z\heartsuit_1\lceil\boldsymbol{x:\phi_1}\rceil,\dots,z\heartsuit_n\lceil\boldsymbol{x:\phi_n}\rceil\Rightarrow z\heartsuit_{n+1}\lceil\boldsymbol{x:\phi_{n+1}}\rceil,\dots,z\heartsuit_{n+m}\lceil\boldsymbol{x:\phi_{n+m}}\rceil,\Theta}$$

where σ sends each p_i^j to a formula ϕ_i^j of $\mathcal{CPL}_\Lambda$ and $\lceil\boldsymbol{x:\phi_i}\rceil=\lceil x:\phi_i^1\rceil\dots\lceil x:\phi_i^{a(i)}\rceil$ is a finite sequence of comprehension expressions according to the arity $a(i)$ of $\heartsuit_i$ and $\Gamma^+\Rightarrow\Delta^+$ denotes the lower sequent.

(Admissible) Structural Rules

$$(\mathsf{W}L)\,\frac{\Gamma\Rightarrow\Delta}{\phi,\Gamma\Rightarrow\Delta}\qquad(\mathsf{W}R)\,\frac{\Gamma\Rightarrow\Delta}{\Gamma\Rightarrow\Delta,\phi}\qquad(\mathsf{Con}L)\,\frac{\Gamma,\phi,\phi\Rightarrow\Delta}{\Gamma,\phi\Rightarrow\Delta}\qquad(\mathsf{Con}R)\,\frac{\Gamma\Rightarrow\Delta,\phi,\phi}{\Gamma\Rightarrow\Delta,\phi}$$

$$(\mathsf{Cut})\,\frac{\Gamma\Rightarrow\Delta,\phi\qquad\phi,\Sigma\Rightarrow\Theta}{\Gamma,\Sigma\Rightarrow\Delta,\Theta}$$

Furthermore, we repeat the information of the lower sequent in all the upper sequents to ensure the admissibility of contraction, and add a weakening context Σ, Θ to both premise and conclusion. Finally, we obtain the desired form of $\mathcal{S}(R)$ in Table 3. The formulae $z\heartsuit_i\lceil \boldsymbol{x} : \boldsymbol{\phi_i} \rceil$ are called the *principal formulae* of $\mathcal{S}(R)$.

Example 19. If (K) is the (one-step sound and one-step complete) rule set for the normal modal logic consisting of the rules

$$(K_n)\frac{p \Rightarrow q_1, \ldots, q_n}{\Diamond p \Rightarrow \Diamond q_1, \ldots, \Diamond q_n}$$

for all $n \geq 0$, we obtain the following first-order version

$$\mathcal{S}(K_n)\ \frac{\phi_0[y/z], \Sigma, z\Diamond\lceil x : \phi_0\rceil \Rightarrow z\Diamond\lceil x : \phi_1\rceil, \ldots, z\Diamond\lceil x : \phi_n\rceil, \Theta, \phi_1[y/z], \ldots, \phi_n[y/z]}{\Sigma, z\Diamond\lceil x : \phi_0\rceil \Rightarrow z\Diamond\lceil x : \phi_1\rceil, \ldots, z\Diamond\lceil x : \phi_n\rceil, \Theta}$$

(where y is fresh in the lower sequent) by the previous definition. Modal neighbourhood semantics is axiomatised by the one-step rule

$$(\mathsf{C})\frac{p \Rightarrow q \quad q \Rightarrow p}{\Box p \Rightarrow \Box q}$$

which expresses that $\Box$ is a congruential operator. The first order version of (C) then reads

$$\mathcal{S}(\mathsf{C})\ \frac{\begin{array}{c}\phi_0[y/x], \Sigma, z\Box\lceil x : \phi_0\rceil \Rightarrow z\Box\lceil x : \phi_1\rceil, \Theta, \phi_1[y/x] \\ \phi_1[y/x], \Sigma, z\Box\lceil x : \phi_0\rceil \Rightarrow z\Box\lceil x : \phi_1\rceil, \Theta, \phi_0[y/x]\end{array}}{\Sigma, z\Box\lceil x : \phi_0\rceil \Rightarrow z\Box\lceil x : \phi_1\rceil, \Theta}$$

(where y is fresh in the lower sequent) which we shall later see to provide a complete and cut-free axiomatisation of Chang's original logic.

If $\mathcal{R}$ is a set of one-step rules, we write $\mathcal{SR} \vdash \Gamma \Rightarrow \Delta$ if $\Gamma \Rightarrow \Delta$ can be derived using the logical and equality rules of Table 3, together with the rules $\mathcal{S}(R)$ from Table 3 for every rule $R \in \mathcal{R}$. We write $\mathcal{SR}\mathsf{Cut} \vdash \Gamma \Rightarrow \Delta$ if the *cut rule* (Cut) of Table 3 is used additionally. If $\mathfrak{M} = (C, \gamma, \pi)$ is a first-order structure over a Λ-structure T, we write $\mathfrak{M}, \vartheta \models \Gamma \Rightarrow \Delta$ if $\mathfrak{M}, \vartheta \models \bigwedge \Gamma \to \bigvee \Delta$ and, as usual $\mathfrak{M} \models \Gamma \Rightarrow \Delta$ if $\mathfrak{M}, \vartheta \models \Gamma \Rightarrow \Delta$ for all variable assignments ϑ and finally $T \models \Gamma \Rightarrow \Delta$ if $\mathfrak{M} \models \Gamma \Rightarrow \Delta$ for all first-order structures $\mathfrak{M}$ over T.

We show soundness and completeness of the sequent system $\mathcal{SR}$ by translating into, and from, the Hilbert system $\mathcal{HR}$ which is known to be (semantically) complete. The translation – initially using (Cut) in the sequent system – relies on a few routine facts concerning structural rules that we now summarise (note that the definition of both $\mathcal{SR}$ and $\mathcal{SR}\mathsf{Cut}$ does *not* involve structural rules). The structural rules are standard as in Table 3: we consider weakening both on the left and on the right and the rules of *left and right contraction*. Admissibility of weakening is standard:

Lemma 20. *The rules of* left weakening *and* right weakening *are height-preserving admissible in $\mathcal{SR}$ and $\mathcal{SR}$*Cut.

Proof. By induction on derivations. Note that weakening is built into modal rules $\mathcal{S}(R)$ that are derived from one-step rules in $\mathcal{R}$. The other cases than $\mathcal{S}(R)$ are done, e.g., as in [17, Lemma 3.5.3] and [10, Theorem 3.2.1 and Theorem 4.2.7].

Finally, we note one consequence of the congruence rule before we show that both systems $\mathcal{HR}$ and $\mathcal{SR}\mathsf{Cut}$ have the same deductive power provided that the rules absorb congruence. We introduce the concept of absorption in a slightly more general form which will be used later.

Definition 21. We say that a set S of sequents *covers* a set S' of sequents if each elements of S' is a subset of some element of S'. We write $S \rhd S'$ if S covers S' where we identify sequents with singleton sets. A set $\mathcal{R}$ of rules *absorbs* a rule P/C if there exists a rule $R = Q/D \in \mathcal{R}$ such that $P \rhd Q$ and $D \rhd C$. A rule set *absorbs congruence* if it absorbs the rule

$$(\mathsf{Cong}\heartsuit)\ \frac{p_1 \Rightarrow q_1 \quad \ldots \quad p_n \Rightarrow q_n \qquad q_1 \Rightarrow p_1 \quad \ldots \quad q_n \Rightarrow p_n}{\heartsuit(p_1, \ldots, p_n) \Rightarrow \heartsuit(q_1, \ldots, q_n)}$$

and it *absorbs monotonicity of* $\heartsuit$ *in the* i*-th argument* if the rule

$$(\mathsf{Mon}_i)\ \frac{p_i \Rightarrow q_i}{\heartsuit(p_1, \ldots, p_n) \Rightarrow \heartsuit(p_1, \ldots, p_{i-1}, q_i, p_{i+1}, \ldots p_n)}$$

is absorbed.

Lemma 22. *Suppose that $\mathcal{R}$ absorbs congruence. Then $\mathcal{SR} \vdash \Gamma, \phi \Rightarrow \phi, \Delta$ for all formulas ϕ.*

Proof. Here we assume for simplicity that $\mathcal{R}$ consists of unary modal operators alone. As $\mathcal{R}$ absorbs congruence, the rule $(\mathsf{Cong}\heartsuit)$

$$\frac{\{\Gamma, \phi_i[y/x] \Rightarrow \psi_i[y/x], \Delta \mid 1 \leq i \leq n\} \qquad \{\Gamma, \psi_i[y/x] \Rightarrow \phi_i[y/x], \Delta \mid 1 \leq i \leq n\}}{\Gamma, z\heartsuit\lceil \boldsymbol{x} : \boldsymbol{\phi} \rceil \Rightarrow z\heartsuit\lceil \boldsymbol{x} : \boldsymbol{\psi} \rceil, \Delta}$$

(where y is fresh in the lower sequent and n is the arity of $\heartsuit$) is admissible in $\mathcal{SR}$ (and $\mathcal{SR}\mathsf{Cut}$). This allows us to proceed by induction on the structure of ϕ, where $(\mathsf{Cong}\heartsuit)$ deals with the inductive case where ϕ is of the form $x\heartsuit\lceil y : \phi \rceil$.

One direction of the translation between the two proof systems can now be given as follows:

Theorem 23. *Suppose that $\mathcal{R}$ absorbs congruence and let $\mathcal{HR} \vdash \phi$. Then $\mathcal{SR}\mathsf{Cut} \vdash \Rightarrow \phi$.*

Proof. First, we demonstrate admissibility of modus ponens in $\mathcal{SR}\mathsf{Cut}$ by

$$\begin{array}{ll} (\mathsf{Cut}) & \dfrac{\Rightarrow \phi \to \psi \qquad \phi \to \psi, \phi \Rightarrow \psi}{\phi \Rightarrow \psi} \qquad \Rightarrow \phi \\ (\mathsf{Cut}) & \hline \qquad\qquad \Rightarrow \psi \end{array}$$

where the derivability of $\phi \to \psi, \phi \Rightarrow \psi$ is easily estabilished by Lemma 22. Note that, in our proof of the theorem, we need (Cut) only for this admissibility.

Then, it suffices to show that all the axioms of $\mathcal{HR}$ (recall Table 3) are derivable in $\mathcal{SR}$ (here we do not need the cut rule). Since this is easy to show for non-modal axioms, we focus on (EG6.2), (CONG) and (ONESTEP(R)). First, (EG6.2) is derivable by (Ren) and Lemma 22. Second, the derivability of (CONG) follows from Lemma 22. Finally, let us move to the provability of (ONESTEP(R)). Suppose that $R = \Gamma_1 \Rightarrow \Delta_1, \dots, \Gamma_k \Rightarrow \Delta_k / \Gamma_R \Rightarrow \Delta_R$ is a one-step rule as in Definition 15. We obtain the following derivation where $N = \{1, ..., n\}$ and $M = \{n+1, ..., n+m\}$:

$$\dfrac{\dfrac{\dfrac{\{P_1[y/x] \wedge \cdots \wedge P_n[y/x], (\Gamma_i\sigma)[y/x] \Rightarrow (\Delta_i\sigma)[y/x] \mid 1 \le i \le k\}}{\{\forall x.(P_1 \wedge \cdots \wedge P_n), (\Gamma_i\sigma)[y/x] \Rightarrow (\Delta_i\sigma)[y/x] \mid 1 \le i \le k\}}\,(\mathsf{L}\forall)}{\forall x.(P_1 \wedge \cdots \wedge P_n), \{z\heartsuit_i\lceil \boldsymbol{x} : \boldsymbol{\phi_i}\rceil \mid i \in N\} \Rightarrow \{z\heartsuit_i\lceil \boldsymbol{x} : \boldsymbol{\phi_i}\rceil \mid i \in M\}}\,\mathcal{S}(R), (\mathsf{W}L/R)}{\forall x.(P_1 \wedge \cdots \wedge P_n), \bigwedge\{z\heartsuit_i\lceil \boldsymbol{x} : \boldsymbol{\phi_i}\rceil \mid i \in N\} \Rightarrow \bigvee\{z\heartsuit_i\lceil \boldsymbol{x} : \boldsymbol{\phi_i}\rceil \mid i \in M\}}\,(\mathsf{L}\wedge, \mathsf{R}\vee)$$

which shows derivability of the axiom ONESTEP(R) as the top sequent is readily seen to be provable in $\mathcal{SR}$ (recall that P_i means $(\bigwedge \Gamma_i \to \bigvee \Delta_i)\sigma$).

For the converse direction, absorption of congruence is not required.

Theorem 24. *Suppose that* $\mathcal{SR}\mathsf{Cut} \vdash \Gamma \Rightarrow \Delta$. *Then* $\mathcal{HR} \vdash \bigwedge \Gamma \to \bigvee \Delta$.

Proof. It suffices to show that all the translations of the axioms and rules of $\mathcal{SR}$ are derivable in $\mathcal{HR}$. We can easily handle the cases of the axioms and rules for logical connectives of first-order logic. As for $\heartsuit \in \Lambda$, the provability of the translation of (Ren) and $\mathcal{S}(R)$ follows from (EG6.2) and (ONESTEP(R)), respectively. □

As a corollary, we obtain (for the time being, in a calculus with cut) both soundness and completeness of the sequent calculus.

Corollary 25. *Suppose that* $\mathcal{R}$ *is one-step sound and strongly one-step complete. Then* $\mathcal{SR}\mathsf{Cut} \vdash \Gamma \Rightarrow \Delta$ *iff* $\models \Gamma \Rightarrow \Delta$.

Proof. By Theorems 23 and 24 in conjunction with soundness and completeness of $\mathcal{HR}$ (Theorem 18). The absorption of congruence was shown in Proposition 5.12 of [12].

A paradigm example of a set of rules satisfying the assumptions of Corollary 25 is C and its CPL translation $\mathcal{S}(\mathsf{C})$ from Example 19 above.

As we have remarked above, the assumption of *strongly* one-step complete rule sets is limiting in that there are only few examples. The companion paper [9] gives a complete Hilbert-style axiomatisation also for *bounded* operators. We repeat the definition for convenience:

Definition 26. A modal operator $\heartsuit$ is k-*bounded* in i-th argument for $k \in \mathbb{N}$ and with respect to a Λ-structure T if for every $C \in \mathsf{Set}$ and every $A_1, \dots, A_n \subseteq C$,

$$[\![\heartsuit]\!]_C(A_1, \dots, A_n) = \bigcup_{B \subseteq A_i, \#B \le k} [\![\heartsuit]\!]_C(A_1, \dots, A_{i-1}, B, A_{i+1}, \dots, A_n).$$

Note that this implies in particular that $\heartsuit$ is monotonic in the i-th argument. Examples of bounded modalities include the standard $\Diamond$ of relational modal logic interpreted over Kripke frames, graded modalities over multigraphs and we refer to [15] for more examples. In the Hilbert-calculus, boundedness is reflected syntactically by the axiom

$$\text{BDPL}_{k,i}\forall\overline{y}.(x\heartsuit\lceil y_1:\phi_1\rceil\ldots\lceil y_n:\phi_n\rceil\leftrightarrow\exists z_1\ldots z_k.(x\heartsuit\lceil y_1:\phi_1\rceil\ldots\lceil y_{i-1}:\phi_{i-1}\rceil$$
$$\lceil y_i:y_i=z_1\vee\cdots\vee y_i=z_k\rceil\lceil y_{i+1}:\phi_{i+1}\rceil\ldots\lceil y_n:\phi_n\rceil\wedge\bigwedge_{j\leq k}\phi_i[z_j/y_i]))$$

where each z_i is fresh for all the y_is and ϕ_is. The derivability predicate induced by exteding the Hilbert calculus $\mathcal{HR}$ by the boundedness axiom above gives completeness under weaker conditions.

Definition 27. We write $\mathcal{BHR}\vdash\phi$ if ϕ is derivable in $\mathcal{HR}$ where additionally $(\text{BDPL}_{k,i})$ is used for every operator that is k-bounded in the i-th argument.

Strictly speaking, the derivability predicate $\mathcal{BHR}$ should include information about precisely which operators are assumed to be k-bounded in the i-th argument, but this will always be clear from the context. In the presence of boundedness, completeness of the Hilbert-calculus can be established under weaker conditions, where we again refer to [9] for details:

Theorem 28. *Suppose that $\mathcal{R}$ is one-step sound and* finitary *strongly one-step complete over a Λ-structure T. Let each operator be bounded in every argument. Then $\mathcal{BHR}\vdash\phi$ iff $\mathfrak{M},\vartheta\models\phi$ for every first-order structure $\mathfrak{M}$ and every variable assignment ϑ.*

We can reflect boundedness in the sequent calculus by adding a paste rule, similar in spirit to the paste rule of hybrid logic [3, Section 7] which was generalised to a coalgebraic setting in [15]. In a sequent setting, this rule takes the form

$$(\mathsf{Paste}_i^k)\ \frac{\begin{array}{c}\Gamma,x\heartsuit\lceil x_1:\phi_1\rceil\cdots\lceil x_{i-1}:\phi_{i-1}\rceil\lceil y:\bigvee_{1\leq j\leq k}y=z_j\rceil\lceil x_{i+1}:\phi_{i+1}\rceil\cdots\lceil x_n:\phi_n\rceil,\\ \phi[z_1/y],\ldots,\phi[z_k/y]\Rightarrow\Delta\qquad z_1,\ldots,z_k\text{ fresh}\end{array}}{\Gamma,x\heartsuit\lceil x_1:\phi_1\rceil\cdots\lceil x_{i-1}:\phi_{i-1}\rceil\lceil y:\phi\rceil\lceil x_{i+1}:\phi_{i+1}\rceil\cdots\lceil x_n:\phi_n\rceil\Rightarrow\Delta}\ ,$$

where z_1, ..., z_k are pairwise distinct fresh variables. Additional use of the above paste-rule in the system $\mathcal{SR}$ is denoted by $\mathcal{BSR}$, that is, we write $\mathcal{BSR}\vdash\Gamma\Rightarrow\Delta$ if $\Gamma\Rightarrow\Delta$ is derivable in $\mathcal{SR}$ where (Paste_i^k) may additionally be applied for every modality that is k-bounded in the i-th argument.

In what follows, we assume that $\mathcal{R}$ absorbs the congruence rules for all $\heartsuit\in\Lambda$ and monotonicity of all operators that are k-bounded in the i-th argrument.

Lemma 29. *Suppose that $\mathcal{R}$ absorbs congruence. Then, the replacement axiom $x=y,\phi[x/z]\Rightarrow\phi[y/z]$ is derivable in $\mathcal{SR}$.*

Proof. By induction on ϕ (note that we do not need the cut rule in this proof). It suffices to check the case where ϕ is of the form $v\heartsuit\lceil\boldsymbol{w}:\boldsymbol{\psi}\rceil$, since the other cases than $z\heartsuit\lceil\boldsymbol{x}:\boldsymbol{\psi}\rceil$ are done, e.g., as in [17, Lemma 4.7.2 (i)] and [10, Lemma 6.5.2]. For simplicity, let us consider the case where ϕ is $v\heartsuit\lceil w:\psi\rceil$ (i.e., $\heartsuit$ is unary). In order to show the derivability of $x=y,(v\heartsuit\lceil w:\psi\rceil)[x/z]\Rightarrow(v\heartsuit\lceil w:\psi\rceil)[y/z]$, here we only consider a case where $z\equiv v$ and $z\not\equiv w$ as follows.

$$
\dfrac{\dfrac{x = y, \psi[x/z][w'/w] \Rightarrow \psi[y/z][w'/w] \qquad \dfrac{\dfrac{y = x, \psi[y/z][w'/w] \Rightarrow \psi[x/z][w'/w]}{x = x, x = y, \psi[y/z][w'/w] \Rightarrow \psi[x/z][w'/w]}\,(\mathsf{Repl}), (\mathsf{W}L)}{x = y, \psi[y/z][w'/w] \Rightarrow \psi[x/z][w'/w]}\,(\mathsf{Ref})}{x = y, y\heartsuit\lceil w : \psi[x/z]\rceil \Rightarrow x\heartsuit\lceil w : \psi[y/z]\rceil}\,(\mathsf{Cong}\heartsuit)}{x = y, x\heartsuit\lceil w : \psi[x/z]\rceil \Rightarrow y\heartsuit\lceil w : \psi[y/z]\rceil}\,(\mathsf{W}L), (\mathsf{Ren}) \quad ,
$$

where w' is fresh in the lower sequent of $\mathsf{Cong}(\heartsuit)$ and two top sequents are provable in $\mathcal{SR}$ by induction hypothesis.

Theorem 30. *Suppose that $\mathcal{R}$ absorbs congruence and monotonicity in the i-th argument of every operator that is k-bounded in the i-th argument. Then $\mathcal{BHR} \vdash \phi$ implies that $\mathcal{BSR}\mathsf{Cut} \vdash \Rightarrow \phi$.*

Proof. First of all, if $\mathcal{R}$ absorbs monotonicity in the i-th argument of $\heartsuit \in \Lambda$, the rule

$$
(\mathsf{Mon}_i)\ \dfrac{\Gamma, \phi_i[y/x] \Rightarrow \psi[y/x], \Delta}{\Gamma, z\heartsuit\lceil \boldsymbol{x} : \boldsymbol{\phi}\rceil \Rightarrow z\heartsuit\lceil x : \phi_1\rceil \ldots \lceil x : \phi_{i-1}\rceil\lceil x : \psi\rceil\lceil x : \phi_{i+1}\rceil \ldots \lceil x : \phi_n\rceil, \Delta}
$$

(where y is fresh in the lower sequent) is admissible in $\mathcal{BSR}$ (and $\mathcal{BSR}\mathsf{Cut}$). Almost all the arguments are the same as the proof of Theorem 23, except that we need to show the provablity of (BDLP) by (Paste) (The only place we need the cut rule is the derivability of modus ponens). More precisely, we can show the left-to-right implication of (BDPL) by means of (Paste_i^k) and (Mon_i) gives the reverse direction. For example, when $\heartsuit$ is unary and 1-bounded, the derivability of the right-to-left direction of (BDPL) is demonstrated as follows.

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{w = v, \phi[w/y] \Rightarrow \phi[v/y]}{v = v, v = w, w = v, \phi[w/y] \Rightarrow \phi[v/y]}\,(\mathsf{W}L)}{v = v, v = w, \phi[w/y] \Rightarrow \phi[v/y]}\,(\mathsf{Repl})}{v = w, \phi[w/y] \Rightarrow \phi[v/y]}\,(\mathsf{Ref})}{x\heartsuit\lceil y : y = w\rceil, \phi[w/y] \Rightarrow x\heartsuit\lceil y : \phi\rceil}\,(\mathsf{Mon})}{x\heartsuit\lceil y : y = w\rceil \wedge \phi[w/y]) \Rightarrow x\heartsuit\lceil y : \phi\rceil}\,(\mathsf{L}\wedge)}{\exists z.(x\heartsuit\lceil y : y = z\rceil \wedge \phi[z/y]) \Rightarrow x\heartsuit\lceil y : \phi\rceil}\,(\mathsf{L}\exists) \quad ,
$$

where the top sequent is the replacement axiom, which is derivable by Lemma 29.

The reverse direction of Theorem 30 is established analogously to Theorem 24 and again absorption properties are not needed.

Theorem 31. $\mathcal{BSR}\mathsf{Cut} \vdash \Gamma \Rightarrow \Delta$ *only if* $\mathcal{BHR} \vdash \bigwedge \Gamma \to \bigvee \Delta$.

Proof. The only difference from the proof of Theorem 23 is to need to care about the translation of Paste. However, we can easily establish this by the axiom (BDLP). □

As in the non-bounded case we obtain semantic soundness and completeness, but under weaker coherence conditions.

Corollary 32. *Suppose that $\mathcal{R}$ is one-step sound and strongly finitary one-step complete. Then $\mathcal{BSR}\mathsf{Cut} \vdash \Gamma \Rightarrow \Delta$ iff $\models \Gamma \Rightarrow \Delta$.*

Proof. By Theorems 30 and 31. Note that absorption of congruence and monotonicity follows from (strong, finitary) one-step completeness as in Proposition 5.12 of [12].

A canonical example of a rule set satisfying the assumptions of the above corollary can be obtained by taking K of Example 19 and extending it with (Paste_i^k) for $i = k = n = 1$.

4.2 Elimination of Contraction and Cut

Note that *a priori* we cannot expect that cut elimination holds for cuts between two instances of modal rules: the set $\mathcal{R}$ of one-step rules can possibly consist of a single rule, and a cut between this rule and itself may not be derivable. We therefore need to impose an additional requirement, *cut and contraction closure* to deal with this case.

Definition 33. Let S be a finite set of sequents. The set of sequents that can be derived from premises S using (only) the *contraction rules* is denoted by $\mathsf{Con}(S)$. Similarly, the set of all sequents that can be derived from premises in S using (only) the *cut rule* is denoted by $\mathsf{Cut}(S)$. A rule set $\mathcal{R}$ *absorbs contraction* if, for all rules $R = P/C \in \mathcal{R}$ and all $C' \in \mathsf{Con}(C)$ there exists a rule $R' = Q/D \in \mathcal{R}$ such that $\mathsf{Con}(P) \rhd Q$ and $D \rhd C'$. A rule set $\mathcal{R}$ *absorbs cut*, if for all pairs of rules in $\mathcal{R}$

$$(R_1)\frac{P_1}{\Gamma_1 \Rightarrow \Delta_1, \phi} \qquad (R_2)\frac{P_2}{\phi, \Gamma_2 \Rightarrow \Delta_2}$$

there is a rule $R = P/C \in \mathcal{R}$ such that $\mathsf{Cut}(P_1 \cup P_2) \rhd P$ and $C \rhd \Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2$.

Informally, absorption of cut and contraction allows us to replace an application of cut or contraction to the conclusions of rules in $\mathcal{R}$ by a possibly different rule with possibly weaker premises and stronger conclusion. While these definitions are purely syntactic, a semantic characterisation has been given in [12] in terms of *one-step cut-free completeness*. For many Λ-structures, including those for probabilistic and graded modal logic, the modal logic K, the logic of (monotone) neighbourhood frames, one-step cut-free complete rule sets are known. In particular, these rule sets satisfy absorption of cut, contraction and congruence [12, Section 5].

The absorption requirements directly translate into proof-theoretic properties of the associated sequent calculus for coalgebraic predicate logic that we now collect. Note that weakening is built into one-step rules so that weakening is always admissible without further assumptions.

Lemma 34. *All the logical and equality rules of $\mathcal{SR}$ are height-preserving invertible.*

Lemma 35. *If $\mathcal{SR} \vdash \Gamma \Rightarrow \Delta$ and y is fresh in Γ and Δ, then If $\mathcal{SR} \vdash \Gamma[y/x] \Rightarrow \Delta[y/x]$ with the same height of derivation.*

Proposition 36. *Suppose $\mathcal{R}$ is a set of one-step rules over a similarity type Λ. If $\mathcal{R}$ absorbs contraction, then the rules $(\mathsf{Con}L)$ and $(\mathsf{Con}R)$ are admissible in $\mathcal{SR}$.*

Proof. By induction on the structure of proofs with the help of Lemmas 34 and 35. We have to use absorption of contraction to replace an application of a contraction rule to the conclusion of $\mathcal{S}(R)$ in order to replace the application of $\mathcal{S}(R)$ by (a possibly different) rule $\mathcal{S}(R')$.

We now turn to cut-elimination, where the majority of the cases are straightforward, and in fact identical to the cut-elimination proof in first-order logic.

Theorem 37 (Cut Elimination). *Suppose that $\mathcal{R}$ absorbs cut and contraction. Then the cut rule is admissible in $\mathcal{SR}$.*

Proof. We proceed by double induction on the size of the cut formula and the size of the proof tree. In all cases that do *not* involve the application of a rule $\mathcal{S}(R)$ for some $R \in \mathcal{R}$ it is straightforward to either propagate the cut upwards or to replace the cut by a smaller cut formula using the fact that contraction is admissible. Now fix a one-step rule R and consider the cuts involving $\mathcal{S}(R)$. For cuts between $\mathcal{S}(R)$ and another $\mathcal{S}(R')$, the cut may be eliminated by the fact that $\mathcal{R}$ absorbs cut. For cuts between $\mathcal{S}(R)$ and an equality rule, the cut can be propagated upwards in the proof tree. For a cut between $\mathcal{S}(R)$ and a logical rule, we distinguish two cases depending on whether or not the cut formula A is principal in one of the rules.

In case A not principal in $\mathcal{S}(R)$ the cut may be eliminated by choosing a different weakening context in the application of $\mathcal{S}(R)$. In case A is principal in $\mathcal{S}(R)$ we observe that A is of the form $x\heartsuit\lceil \boldsymbol{y} : \boldsymbol{\phi}\rceil$ and therefore cannot be principal in a logical rule. This allows us to propagate the cut upwards in the proof tree.

As an immediate corollary, we obtain completeness of the cut-free calculus assuming that $\mathcal{R}$ is *strongly* one-step complete:

Corollary 38. *Suppose that $\mathcal{R}$ is one-step sound and strongly one-step complete over a Λ-structure T. Then $\models \Gamma \Rightarrow \Delta$ iff $\mathcal{SR} \vdash \Gamma \Rightarrow \Delta$.*

Proof. This follows from Theorem 37 with the help of Proposition 5.11 and 5.12 of [12], the latter asserting precisely the absorption of cut and congruence.

The situation is more complex in presence of bounded operators where completeness of the Hilbert calculus is only guaranteed in presence of (BDPL), and completeness of the associated sequent calculus relies on (Paste_i^k). The difficulty in a proof of cut-elimnation is a cut-end derivation where a cut is performed on $x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil$ which is introduced by (Paste_i^k) and a (one-step) rule where the same formula is principal. We leave this as an open problem:

Open Problem 39. *Is there a way to modify the rules of $\mathcal{BSR}$ so that completeness with respect to $\mathcal{BHR}$ holds and cut is admissible?*

5 Conclusions

We believe that results obtained here, particularly in Section 3 strengthens the claims first made in [9] concerning naturality of CPL as a (or perhaps "the"?) predicate counterpart of existing coalgebraic formalisms. As concerns sequent systems and cut-elimination results in Section 4, we have fully achieved our goals for those functors and signatures which are "sufficiently neighbourhood-like" (S1SC). We are presently working on the intriguing question whether a constructive proof of cut-elimination can be given on the "Kripke-like" end, i.e., Open Problem 39. We refer the reader to [9] for more on open problems and future work.

References

1. Areces, C., ten Cate, B.: Hybrid logics. In: Blackburn, P., van Benthem, J., Wolter, F. (eds.) Handbook of Modal Logic. Elsevier (2007)
2. Blackburn, P.: Representation, reasoning, and relational structures: a hybrid logic manifesto. Logic Journal of the IGPL 8(3) (2000)
3. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press (2001)
4. Blackburn, P., ten Cate, B.: Pure extensions, proof rules, and hybrid axiomatics. Studia Logica 84(2), 277–322 (2006)
5. ten Cate, B.: Model theory for extended modal languages. PhD thesis, University of Amsterdam, ILLC Dissertation Series DS-2005-01 (2005)
6. ten Cate, B.D., Franceschet, M.: On the Complexity of Hybrid Logics with Binders. In: Ong, L. (ed.) CSL 2005. LNCS, vol. 3634, pp. 339–354. Springer, Heidelberg (2005)
7. Chang, C.: Modal model theory. In: Mathias, A., Rogers, H. (eds.) Cambridge Summer School in Mathematical Logic. Lecture Notes in Mathematics, vol. 337, pp. 599–617. Springer, Heidelberg (1973)
8. Hvid Hansen, H., Kupke, C., Pacuit, E.: Neighbourhood structures: Bisimilarity and basic model theory. Log. Methods Comput. Sci. 5 (2009)
9. Litak, T., Pattinson, D., Sano, K., Schröder, L.: Coalgebraic Predicate Logic. In: Czumaj, A., Mehlhorn, K., Pitts, A., Wattenhofer, R. (eds.) Automata, Languages, and Programming, Part II. LNCS, vol. 7392, pp. 299–311. Springer, Heidelberg (2012)
10. Negri, S., Von Plato, J.: Structural Proof Theory. Cambridge University Press (2001)
11. Pacuit, E., Salame, S.: Majority logic. In: Dubois, D., Welty, C., Williams, M.-A. (eds.) Proc. KR 2004, pp. 598–605. AAAI Press (2004)
12. Pattinson, D., Schröder, L.: Cut elimination in coalgebraic logics. Information and Computation 208, 1447–1468 (2010)
13. Schröder, L.: A finite model construction for coalgebraic modal logic. J. Log. Algebr. Prog. 73, 97–110 (2007)
14. Schröder, L., Pattinson, D.: Coalgebraic Correspondence Theory. In: Ong, L. (ed.) FOSSACS 2010. LNCS, vol. 6014, pp. 328–342. Springer, Heidelberg (2010)
15. Schröder, L., Pattinson, D.: Named models in coalgebraic hybrid logic. In: Marion, J.-Y., Schwentick, T. (eds.) *STACS*. LIPIcs, vol. 5, pp. 645–656. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2010)
16. Sgro, J.: The interior operator logic and product topologies. Trans. AMS 258(1), 99–112 (1980)
17. Troelstra, A., Schwichtenberg, H.: Basic Proof Theory. Cambridge University Press (1996)

Author Index